

Detailed Primary and Secondary Distribution System Modeling and Validation of
Feeders, Loads and Distributed Energy Resources Using Field Measurements

by

Karen Vanessa Montano-Martinez

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2022 by the
Graduate Supervisory Committee:

Vijay Vittal, Chair
Raja Ayyanar
Yang Weng
Anamitra Pal

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

The past few years have witnessed a significant growth of distributed energy resources (DERs) in power systems at the customer level. Such growth challenges the traditional centralized model of conventional synchronous generation, making a transition to a decentralized network with a significant increase of DERs. This decentralized network requires a paradigm change in modeling distribution systems in more detail to maintain the reliability and efficiency while accommodating a high level of DERs. Accurate models of distribution feeders, including the secondary network, loads, and DER components must be developed and validated for system planning and operation and to examine the distribution system performance. In this work, a detailed model of an actual feeder with high penetration of DERs from an electrical utility in Arizona is developed. For the primary circuit, distribution transformers, and cables are modeled. For the secondary circuit, actual conductors to each house, as well as loads and photovoltaic (PV) units at each premise are represented. An automated tool for secondary network topology construction for load feeder topology assignment is developed. The automated tool provides a more accurate feeder topology for power flow calculation purposes. The input data for this tool consists of parcel geographic information system (GIS) delimitation data, and utility secondary feeder topology database. Additionally, a highly automated, novel method to enhance the accuracy of utility distribution feeder models to capture their performance by matching simulation results with corresponding field measurements is presented. The method proposed uses advanced metering infrastructure (AMI) voltage and derived active power measurements at the customer level, data acquisition systems (DAS) measurements at the feeder-head, in conjunction with an AC optimal power flow (ACOPF) to estimate customer

active and reactive power consumption over a time horizon, while accounting for unmetered loads. The method proposed estimates both voltage magnitude and angle for each phase at the unbalanced distribution substation. The accuracy of the method developed by comparing the time-series power flow results obtained from the enhancement algorithm with OpenDSS results and with the field measurements available. The proposed approach seamlessly manages the data available from the optimization procedure through the final model verification.

DEDICATION

I'd like to dedicate this dissertation to my beloved husband, Sergio, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life. This work is also dedicated to my parents, Victor, and Olga, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve. My brothers Brian and Jonathan who believe in me more than anyone else. My friends who were always there even though they don't understand anything I do. My dog Samantha for always being by my side while working regardless of the time. And finally, to the more important person in my life, my daughter Amanda, you are the light in my life, being your mother is the best thing that has ever happened to me.

Thank you, my love for you all can never be quantified. God bless you.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Vijay Vittal for providing me the opportunity to work on this project. His encouragement, guidance and support have played a main role in the completion of this dissertation. I am beyond grateful for his role as my advisor.

I would also like to thank Dr. Raja Ayyanar, Dr. Yang Weng, and Dr. Anamitra Pal for their time and consideration in being a member of my graduate supervisory committee. Their comments and suggestions helped me improve the quality of my work.

I would like to thank the project's postdoc Dr. Shanshan Ma for her time and valuable knowledge on the topic. I would like to thank my college and friend Sushrut Thakar for those extended hours we worked together.

I am thankful to Cynthia Rojas and Dr. Daniel Haughton from the Arizona Public Service company for providing us the data that helped facilitate this work and for meeting with us every time we needed to discuss their data.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES.....	ix
NOMENCLATURE.....	xiii
CHAPTER	
1 INTRODUCTION.....	1
1.1 Background	1
1.2 Motivation and Contributions	2
1.3 Dissertation Organization	5
2 LITERATURE REVIEW.....	6
2.1 Feeder Modeling	6
2.1.1 Feeder Modeling Approaches	8
2.1.2 Gaps in Literature	10
3 UTILITY FEEDER AND DATABASES	12
3.1 Utility Feeder.....	12
3.2 Modeling Data Resources	13
3.2.1 Net-metering Calculation.....	15
3.2.2 Geographic Information System (GIS).....	16
3.3 Process for Constructing the Secondary Network Topology for the Feeder Model	17
3.3.1 DISMOTT: Distribution System Model Transformation Tool.....	17
3.3.2 Misplaced Loads Correction	18

CHAPTER	Page
3.4 High PV Penetration in Distribution Systems	20
4 DISTRIBUTION SYSTEM MODEL ENHANCEMENT ALGORITHM.....	22
4.1 Modeling Data Resources	22
4.2 Distribution System Model Enhancement Algorithm	24
4.2.1 Input Data and Initialization	24
4.2.2 Enhancement Algorithm: ACOPF Based on Current-Voltage (IV) Formulation.....	27
4.2.3 Algorithm Output and Simulation Capabilities	32
4.2.4 OpenDSS Optimal Feeder Model	33
5 MODEL VALIDATION: UTILITY FEEDER RESULTS	36
5.1 Algorithm- OpenDSS Power Flow Comparison	37
5.2 Enhanced Distribution Feeder Model Validation.....	40
5.3 Detailed Feeder Characteristics	45
5.4 Comparison: Enhanced Feeder Model Vs. Using Constant Power Factor for All the Loads.....	49
5.4.1 Active Power Profile: Base-case.....	49
5.4.2 Reactive Power Profile: Base-case	50
5.4.3 Voltage Profile	51
5.4.4 Active and Reactive Profile for Other Days	51
5.4.5 Accuracy Comparison with Enhanced Feeder Model.....	52
6 AUTOMATED SECONDARY NETWORK TOPOLOGY CONSTRUCTION	57
6.1 Automated Secondary Network Topology Construction.....	58

CHAPTER	Page
6.1.1 Input Data.....	59
6.1.2 Algorithm 1: Data Processing.....	60
6.1.3 Algorithm 2: Load Clustering (DBSCAN Clustering Method).....	61
6.1.4 Algorithm 3: Optimization Algorithm for Load Placement.....	64
6.1.5 Output	67
6.2 Results	67
7 DISCUSSION AND CONCLUSIONS	74
7.1 Discussion	74
7.2 Conclusions	75
REFERENCES	77

LIST OF TABLES

Table	Page
1. Comparison of Active and Reactive Power From Power Flow Solution	39

LIST OF FIGURES

Figure	Page
1. Distribution Feeder General Layout, Adapted from[17].	7
2. Arizona Utility Feeder Circuit Diagram with All the Elements.	13
3. Single Line Diagram of a Small Portion of the Cyme Model.....	15
4. Metering Infrastructure at a Typical Premise: Available Measurements are in Blue, Estimated Values are in Green.....	16
5. GIS Model of the Feeder.....	17
6. Flowchart for the CYMEDIST to OpenDSS Conversion Script.	18
7. Distribution Transformer Voltage Profile Vs Premises Connected at the Distribution Transformer Voltage Profiles.	20
8. PV Profile for All the Premises Connected to the Same Distribution Transformer.	21
9. Modeling Data Resources for Distribution System Model Enhancement.	23
10. Flowchart for Distribution System Model Enhancement Algorithm.....	25
11. Source and Feeder-head Representation in the Proposed Framework: Available Measurements are in Blue, Estimated Values are in Green. Three Independent Single-phase Voltage Sources are Implemented in the Unbalanced Source Bus to Model an Unbalanced Circuit.	34
12. Bus Voltage Magnitude Error Per Phase Between Enhancement Algorithm Power Flow and OpenDSS Solution for a Single Snapshot for the Historical Feeder Load Peak Hour for the Modeled Feeder.	40

13. Voltage Errors Between OpenDSS and the Optimization-based Algorithm for Different Hours of a Single Day (Historical Feeder Load Peak Day) for the Modeled Feeder. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 times the Interquartile Range Away from the Box.....	40
14. Active and Reactive Power Feeder-head Comparison Between the Enhanced OpenDSS Model and Das Measurements for the Historical Feeder Load Peak Day.	42
15. Active and Reactive Power Feeder-head Comparison Between the Enhanced OpenDSS Model and Das Measurements for the Maximum Generation Condition Day.....	42
16. Voltage Comparison Results Between OpenDSS and AMI Data for Some Premises along the Feeder for the Historical Feeder Load Peak Day.....	43
17. Voltage Comparison Results Between OpenDSS and AMI Data for Some Premises along the Feeder for the Maximum Generation Condition Day.	43
18. For the Two Scenarios Considered, the RMS Errors Calculated over One Day Between the Field-measured Voltages and the Voltages Obtained from the Enhanced OpenDSS Model for All the Voltage Meters in the System Are Plotted Using a Box and Whisker Plot. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 times the Interquartile Range Away from the Box.....	44

19. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm for the Historical Feeder Load Peak.....	44
20. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm for the Maximum Generation Condition.	45
21. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm During the Night.....	45
22. Distribution Feeder Gross Load, Net Load, and PV Production Behavior.....	46
23. Power Factor for Some Premises Along the Feeder for the Historical Feeder Load Peak Day.	47
24. Customers' Power Factors Obtained from the Enhanced Feeder Model for All the Loads in the Feeder for Both Cases, the Historical Feeder Load Peak Snapshot (Left) and the Maximum Generation Condition Snapshot (Right).....	48
25. Flowchart for Time-series Profile Formulation Based on OpenDSS Model Using Constant Power Factor for All the Loads and Balanced Substation Source.....	49
26. Active and Reactive Power Feeder-head Comparison Between the Feeder Model Using Constant Power Factor for All the Loads in the System and DAS Measurements for the Historical Feeder Load Peak Day.	53
27. Voltage Comparison Results Between the Feeder Model Using Constant Power Factor for All the Loads in the System and AMI Data for Some Premises along the Feeder for the Historical Feeder Load Peak Day.....	54

28. The RMS Errors Calculated over the Historical Feeder Load Peak Day Between the Field-measured Voltages and the Voltages Obtained from the Feeder Model Using Constant Power Factor for All the Loads in the System for All the Voltage Meters are Plotted Using a Box and Whisker Plot. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 Times the Interquartile Range Away from the Box.	55
29. Enhanced Feeder Model Process.	57
30. Secondary Network Topology Construction New Approach.	59
31. Secondary Network Topology Algorithm Detailed.	59
32. Raw Input Data Processing.	60
33. DBSCAN Algorithm.	63
34. Algorithm 1: Input Data.	69
35. Algorithm 1: Raw Data Processing Results of Parcel Centroids and Loads/PVs Nodes.	70
36. Algorithm 2: DBSCAN Preliminary Results.	71
37. Algorithm 2: Nested DBSCANs Results.	72
38. Algorithm 3: Optimization Algorithm for Load Placement Results.	73

NOMENCLATURE

Ω_N	Set of Buses.
Ω_D	Subset of Buses with Load.
Ω_{AMI_V}	Subset of Load Buses with AMI Voltage Information.
Ω_{AMI_D}	Subset of Load Buses with AMI Active Power Information.
Ω_{D1}	Subset of Load Buses without Ami Information ($\Omega_D - \Omega_{AMI_D}$).
Ω_{PV}	Subset of Buses with PV Resources.
Ω_{cap}	Subset of Buses with Capacitor Banks.
Ω_{SB}	Set of Substation Buses.
Ω_L	Set of Distribution Lines.
Ω_H	Set of Feeder-head Buses.
Ω_T	Set of Distribution Transformers.
ψ	Set of Phases, i.e., $\{a, b, c\}$.
$\delta(i)$	Set of Bus Nodes Connected to Bus i .
$B_{c,\phi}$	Capacitance of Capacitor Bank c on Phase ϕ .
d_l^D	Gross Load at Bus l with AMI Active Power Measurements.
P_l^D	AMI Active Power Delivered to Bus l .
P_l^R	AMI Active Power Received at Bus l .
P_l^{PV} / Q_l^{PV}	AMI Active/Reactive Power Produced by the Solar PV at Bus l .
P_m^{Tr}	No-load Loss of Transformer m .
P^H / Q^H	DAS Active/Reactive Power Data at Feeder-head.

$ \hat{V}_{i,\phi} $	AMI Voltage Magnitude Data at Phase ϕ at Bus i .
$R_{i,j}^{\phi,p}$	Resistance of Distribution Line (i,j) Between Phases ϕ and p .
$X_{i,j}^{\phi,p}$	Reactance of Distribution Line (i,j) Between Phases ϕ and p .
$y_{i,j}^{\phi,p}$	Admittance of Distribution Line (i,j) Between Phases ϕ and p .
$PF_l^{max/min}$	Maximum/Minimum Power Factor for Load l .
$ V_{s,\phi} /\angle V_{s,\phi}$	Magnitude/Angle of Voltage at Phase ϕ at Substation s .
$V_{i,\phi}^r/V_{i,\phi}^{im}$	Real/Imaginary Part of Voltage at Phase ϕ at Bus i .
$I_{i,\phi}^{r,inj}/I_{i,\phi}^{im,inj}$	Real/Imaginary Part of Injected Current at Phase ϕ at Bus i .
$I_{i,j,\phi}^r/I_{i,j,\phi}^{im}$	Real/Imaginary Part of Line Flow Current at Phase ϕ in Line (i,j) .
$P_{s,\phi}^G/Q_{s,\phi}^G$	Active/Reactive Power Output at Phase ϕ at Substation s .
$P_{l,\phi}^D/Q_{l,\phi}^D$	Active/Reactive Power at Phase ϕ for Load l .
$Q_{c,\phi}^C$	Reactive Power Output of Capacitor Bank c at Phase ϕ .
$Q_{n,\phi}^{loss}$	Reactive Power Losses for Bus n at Phase ϕ .
Ω_{cl}	Set of Clusters.
Ω_p	Set of Parcels.
Ω_{cl}^p	Subset of Clusters to be Considered for the Load Placement.
Ω_p^{cl}	Subset of Parcels to be Considered for the Load Placement.
Ω_D	Set of Possible Connections Between the Clusters and the Parcels.
$D_{i,j}$	Distance Between a Cluster i and a Parcel j .

$X_{i,j}$

Connection Between a Cluster i and a Parcel j .

CHAPTER 1

INTRODUCTION

1.1 Background

The increased penetration of distributed energy resources (DERs) – which include renewable energy resources, distributed energy storage, and electric vehicles (EVs) – in the electric grid has resulted in unprecedented changes to power system operation, such as bidirectional power flows and increased voltage fluctuations [1]. As the DER penetration level continues to increase, these issues would further impact the planning and operation of power distribution systems, increasing the need to monitor and control these resources and the necessity to model the integration of solar and wind generation into the planning and operation tool for modern electric power systems [2]–[5]. In the United States (U.S.), the average growth for rooftop solar has been 45% per year since 2008 [6]. Among other reasons, the growth of DERs is being driven by policy, with subsidies and incentives for rooftop solar photovoltaics (PV) [3], [7]. Regulatory changes, technical advances, and environmental impacts have led a change in paradigm from the traditional centralized model to a decentralized network with significant increase of DERs, that is a transition from conventional synchronous resources to non-synchronous inverted-based resources [3], [8], [9]. This transition is marked by the intermittent behavior of renewable sources, cost reduction of energy storage technologies and the necessity for a communications infrastructure from the grid transmission level down to the distribution level [10]. As DERs are being connected to the distribution system on a large scale, additional examination and analyses are necessary [7]. High penetration levels of DERs have led to technical

challenges and benefits on the distribution system [11], [12]. On one hand, the rapid growth in distributed solar PVs challenges the ability of the grid to manage voltage and loading in the distribution system [3]. Additionally, as the DER penetration level increases, the frequency of the system, as well as the active and reactive power flow directions and levels, can be significantly affected by the variability of solar irradiance [7], [13]. On the other hand, DERs can be used to alleviate localized overloads of existing distribution substations, where the cost of the substation capacity increase is excessive in comparison to size of the overload [3]. Also, DERs may support the distribution grid via emergency backup, energy savings, and improved power quality.

Due to the recent emphasis on a more accurate representation of the grid, many utilities now have extensive geographic information system (GIS) databases on feeder equipment and conductor segments. Additionally, the utilities are looking to expand the advanced metering infrastructure (AMI) and data acquisition systems (DAS) on the distribution network. Consequently, there have been efforts to model the distribution systems in more detail using various data available from additional field measurements and AMI which measures the utility-to-customer interface. By leveraging these data, a high-fidelity feeder model can be developed to address the needs of the utilities to improve distribution system modeling to effectively plan and operate for future smart distribution systems with DERs.

1.2 Motivation and Contributions

Reliable and accurate modeling of distribution feeders is important to efficiently manage DERs, especially renewable energy sources, with advanced Volt/VAr optimization

and other distribution system automation schemes [14], [15]. This work focuses on the accurate modeling of unbalanced multi-phase distribution feeders based on the available measurement data, such as AMI voltage and derived active power measurements at the customer level and DAS measurements data. In this work, data from an electric utility is used to create a detailed model of an actual feeder in Arizona. For this purpose the open-source software OpenDSS [16] is used due to its modeling and time series analysis capabilities. The secondary network is modeled in detail, representing actual conductors to each house, as well as loads and PV units defined at each premise. To model the secondary network topology accurately, an automated tool for secondary network topology construction for load feeder topology assignment is developed. The automated tool provides a more accurate feeder topology for power flow calculation purposes. Then, this work develops a distribution system model enhancement method to model distribution systems, including the secondary network, to capture the feeder performance by matching feeder model power flow results with corresponding field measurements across the entire feeder. The proposed enhancement method can serve as a benchmark for utilities and academics to improve the model accuracy and reliability of distribution systems for analysis and operation at the utility level with high penetration of DERs. At the same time, the proposed model enhancement method provides a guide for utilities to use automated AMI and DAS measurements to model their distribution feeders with less human intervention.

The effectiveness of the proposed approach is validated on an actual three-phase feeder with high penetration of DERs from an electric utility. The validation is performed in two stages: the first stage compares the power flow solutions obtained using the ACOPF

with an OpenDSS (Open Distribution System Simulator) [16] model with the same network, load, and source data – a very good match between the two solutions at this stage of validation highlights that the ACOPF formulation accurately represents the distribution system power flow. The second stage involves comparing the time-series power flow results from the OpenDSS model enhanced using the proposed method with the DAS and AMI measurements in terms of quantities at the substation and the voltages along the feeder at individual residences – a good match at this stage emphasizes that the distribution feeder model is successfully enhanced and accurately matches the actual feeder.

The key contributions of this work are multifold:

- 1) The proposed approach presents a novel computationally efficient method for estimating customer active and reactive power time-series consumption, including unmetered loads, using AMI voltage and derived power measurements available at some customer locations along the feeder. The reactive power for each premise is calculated without assuming a constant power factor for all the loads in the feeder, providing a more accurate feeder model.
- 2) An unbalanced substation model is proposed and implemented to capture the unbalanced nature of distribution substations in practical utility feeders. For phases at the substation with no available measurements, both the voltage magnitude and angle are estimated.
- 3) The proposed method provides a complete power flow solution, including the secondary circuit representation, using sparse measurements along a feeder, extending the observability and planning capabilities of the feeder under study.

- 4) An automated tool for secondary network topology construction for correct load feeder topology assignation is presented. The automated tool provides a more accurate feeder topology for power flow calculation purposes. This new tool uses commonly available inputs: parcel GIS delimitation data for the location of the feeder (municipal lot survey information), and utility secondary feeder topology database (customers' billing information).

1.3 Dissertation Organization

This report is structured as follows. Chapter 2 provides a literature review on feeder modeling. Chapter 3 describes the modeling data resources available at the utility level and their implementation in the development of a detailed distribution system model. Chapter 4 presents the proposed distribution system model enhancement algorithm, the data implementation to create the input data for the algorithm, the formulation of the proposed ACOPF approach, the output of the enhanced algorithm, and its simulation capabilities. Chapter 5 presents the validation between the enhanced algorithm power flow output and OpenDSS power flow, the validation between the enhanced algorithm power flow output and the field measurements, and some feeder characteristics derived from the enhanced feeder model developed. Chapter 6 presents the automated secondary network topology construction tool developed. Chapter 7 provides a discussion and concludes the report.

CHAPTER 2

LITERATURE REVIEW

This work presents an AC optimal power flow (ACOPF) formulation to tune a distribution feeder model using field measurements. This chapter presents a literature review relevant to this topic.

2.1 Feeder Modeling

In general, most power systems consist of generation facilities that feed bulk power into the high-voltage bulk transmission network that serves several distribution substations. A typical distribution substation serves from 1 to as many as 10 feeder circuits. A typical feeder circuit may serve numerous loads of all types. Industrial customers based on their demand and requirements may be connected to the distribution feeder primary circuit or directly to the bulk transmission system. On the other hand, commercial and residential customers generally are served by distribution transformers that are connected to a distribution feeder circuit. Components of the feeder may consist of the following [17]:

- Three-phase primary main feeder.
- Three-phase, two-phase, and single-phase laterals.
- Step-type voltage regulators or load tap changing transformer (LTC).
- In-line transformers.
- Shunt capacitor banks.
- Three-phase, two phase, and single-phase loads.
- Distribution transformers (step-down to customer's voltage).

A distribution feeder is fundamentally unbalanced due to the large number of unequal single-phase loads that are served and the distributed inequality in the phases of the feeder.

To perform accurate power flow and short-circuit studies in a distribution system, it is necessary to model the distribution feeder as accurately as possible. This requires the models to be developed in phase reference frame and utilize three-phase models for the major components. Figure 2.1 shows the one-line diagram of a general three-phase feeder and depicts the major components of the distribution system. As observed in Figure 2.1, it is necessary to specify the phasing of the line segments to develop most accurate models.

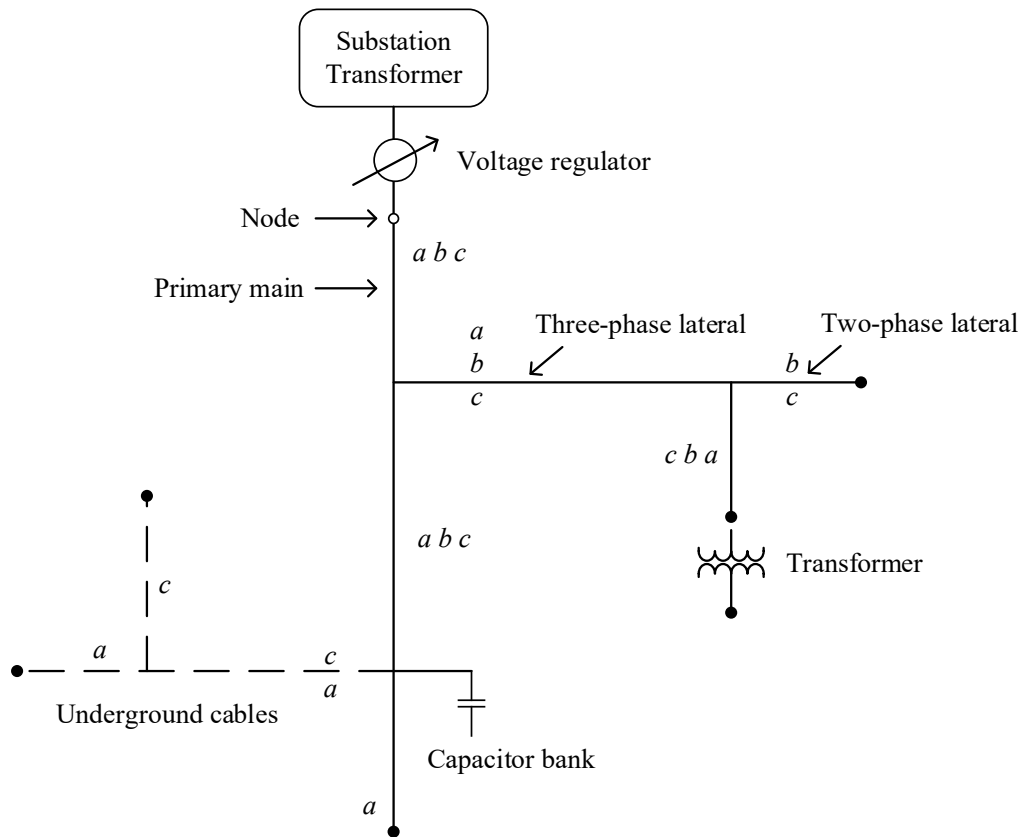


Figure 2.1. Distribution Feeder General Layout, Adapted from[17].

2.1.1 Feeder Modeling Approaches

Conventionally, distribution systems with just one source at the substation have relied on significant model approximations, avoiding detail extension to the secondary circuits [18], [19]. However, a large share of DERs is located at the distribution system secondary at on-site customer locations, creating the need for a paradigm shift to model distribution systems with more accuracy [20]. Historical reasons for resorting to the simplistic models are lack of accurate measurements and data, as well as a more challenging modeling effort than that for transmission systems owing to the three-phase unbalance [21]. Distribution systems are especially complex to analyze due to the nature of the unbalanced impedances and loads [13]. Several authors have proposed methods to create approximate models of the secondary circuit [22]–[25]. Nevertheless, an inaccurate model of the secondary network can misrepresent the effects of DERs, resulting in different voltages and incorrect power calculation across the secondary network when conducting power flow analysis [26]. All components of the electric grid must be represented by mathematical models that provide simulations that accurately represent system performance. It is important to separate the modeling of DERs from the load, even when both exist “behind the meter”[9]. Reliable and accurate distribution system modeling, including the secondary network and the various components such as load and DER, are essential for distribution system operational analysis while accommodating a high level of DERs. Issues created by increased PV penetration requires more accurate modeling detail to effectively assess the issues, and simulate mitigating solutions [27]. As recognized by [28], a need for modeling the secondary network and the actual loads in the system for smart grid applications is continuously increasing. The authors in [26] show that an

inaccurate model of the secondary network or an aggregate model of the load can misrepresent the effects of the secondary network and result in a different voltages across the secondary network and incorrect power measurements. Hence, modeling the secondary network in detail is essential to assess a distribution network with high penetration of DERs, and identify overvoltage due to the high penetration of solar photovoltaic (PV) [29].

By leveraging field measurements, a high-fidelity feeder model can be developed. For instance, the authors in [22] provide a test case based on a real feeder in Iowa which models the primary distribution system in detail. A different project in [23] details feeder modeling in Flagstaff, AZ using geographic information systems (GISs) and AMI data available. The authors in this project have used GIS and AMI data to create an accurate model for the time series analysis of the system. Nevertheless, the feeder models utilized are limited to the representation of the primary distribution infrastructure. Similarly, the authors in [30] use meter readings to model the load at a primary node.

Lately, some researchers have tried to include a representation of the secondary circuit in the distribution system model. The authors of [31] model a low voltage network where they aggregated meter readings at each pole. However, the results are not compared against the field measurements, and the model only pertains to a single snapshot in time. On the other hand, the authors of [32] model the primary circuit in detail and also use the AMI data to create a time series model. While they do not model the secondary in detail using the actual secondary network data, they do create an approximate representation of the secondary. This increase in distribution system observability due to AMI and other emerging sensors has raised the interest in new methods to model the distribution network accurately [33]. A distribution system parameter estimation (DSPE) method using optimal

linear regression model and AMI data is proposed in [14]. However, the authors validated their method using a small 66-node three-phase test circuit and a single-phase secondary circuit. Additionally, the authors assumed the availability of reactive power measurements or power factors. In reality, this assumption is not always true; therefore, this method portrays an approximate representation of the secondary. Similarly, the authors in [34], [35] use GIS and AMI data to model single-phase loads and high penetration of DERs. However, the authors only modeled loads where measurements exist and assumed a constant power factor for all loads.

A method to estimate the impedance of secondary branches using AMI measurements of voltage and active and reactive power is proposed in [36]. The authors presented an optimization algorithm based on gradient search to calculate the voltage of the upstream node from a measured load. However, this approach requires complete observability of all the loads in the feeder to create an accurate model and may be inaccurate for the feeders with unmetered loads. Another drawback of [14], [33]–[36] is that the authors assumed the source at the feeder-head to be balanced, which does not reflect the unbalanced nature of distribution substations in practical utility feeders.

2.1.2 Gaps in Literature

For the distribution system model enhancement, this work proposes an approach based on nonlinear AC optimal power flow (ACOPF) for unbalanced multi-phase distribution systems to increase the system observability and estimate the unmetered loads using available field measurements. The proposed method is based on the rectangular current-voltage (IV) power flow formulation and an extension of the ACOPF model

presented in [37]. It is important to note that goal of this report is different from that of [37]; consequently, the proposed models are different. The approach presented in this work models the unbalanced multi-phase distribution system in detail and accurately since it considers the impact of distribution lines' mutual impedance and shunt elements on the voltage profile and power loss. Furthermore, compared with the unmetered load estimation function in some commercial software that equally distribute unmetered loads, the proposed ACOPF estimates the unmetered loads by minimizing the norm of the difference between bus voltage magnitudes and the corresponding AMI voltage measurement data. The method developed in this work seamlessly transfers the available data through the optimization-based method to the final model verification, with limited human intervention.

CHAPTER 3

UTILITY FEEDER AND DATABASES

A real-life feeder from an electrical utility is selected for this work. In this chapter, the feeder characteristics and databases provided by the utility are presented. Additionally, the secondary network topology is estimated and constructed. For the primary circuit, distribution transformers, cables, and large loads are modeled. For the secondary circuit, actual conductors to each house, as well as loads and solar photovoltaic (PV) units at each premise are represented.

3.1 Utility Feeder

This section describes an actual 12.47 kV, nine km-long utility feeder in Arizona that serves residential customers. Figure 3.1 shows the circuit diagram of the feeder with all its elements. The peak net load on the feeder was 7.35 MW on 07/15/2019. The feeder has one of the highest PV penetrations among the utility's operational feeders, with 3.8 MW of residential rooftop PV installed. Hence, a penetration level of more than 200% compared to the feeder total gross load (3.8 MW total solar PV generation/1.6 MW total gross load) is observed during peak solar PV production hours. The OpenDSS model for this feeder has an unbalanced 69/12.47 kV source representing the substation, 7864 buses, 1790 primary sections, 5782 secondary sections, 371 distribution transformers, four capacitor banks of 1.2 MVAR rating each, 1737 loads, and 766 PV units. There are 784 PV (production) and 1652 load (billing) meters installed on the system, and 1194 meters of these also report voltage magnitudes.

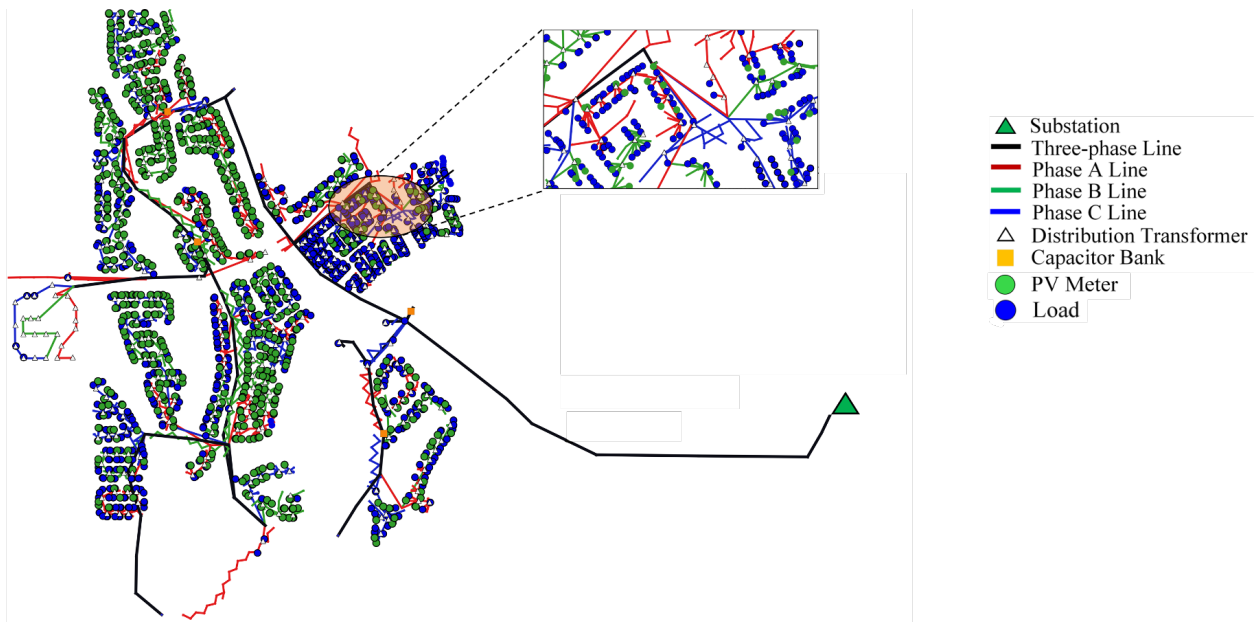


Figure 3.1. Arizona Utility Feeder Circuit Diagram with All the Elements.

3.2 Modeling Data Resources

Various data have been acquired from the utility regarding this feeder to undertake the modeling process:

1. *CYME model of the feeder:* The CYME [38] model provided by the utility contains all the sections, distribution transformers, loads, capacitors and switching elements in the system including the sections and loads on the secondary. The PV generation is provided as aggregated values at the distribution transformers. In addition, the secondary-side locations of installed PV are provided in the model. Figure 3.2 shows a small portion of the single line diagram from this CYME model, illustrative of the detailed modeling of the secondary side involved.
2. *GIS data for all the sections in the system:* The GIS database contains conductor details and the latitude and the longitude of both endpoints of all the circuit sections, including the secondary network. Additionally, equipment ratings of transformers and capacitors

and locations of the system elements such as loads, and PV units are available. These data are used to create the distribution system topology, including load allocation [39], [40].

3. *DAS database*: The feeder-head data included hourly measurements of the feeder-head active and reactive power, the magnitude of individual phase currents and the magnitude of phase A voltage for one year.
4. *AMI meter measurements for 15 days*: The AMI database contains measurements of energy (*kWh*) from 784 PV (production) meters and 1652 load (billing) meters. These measurements are given as aggregated values every 15 minutes or each hour, depending on the meter. The average active power (*kW*) (henceforth termed “active power” in this document) is derived for each meter by aggregating the measurements of energy consumption for one hour. The AMI database also includes voltage magnitude measurements from 1194 load meters given each 15 minutes. These considerations are taken after discussions with the utility, who provided the measurements and information about how they manage their data. From AMI meters, the measurements of voltage magnitude and active power in hourly intervals are acquired for a given day in the summer of 2019 and for two weeks in the spring of 2019.
5. *Location data for all the meters*: The physical locations for all the meters are also obtained from the utility. The locations of some loads or PV meters are obtained by using this physical location information to find the latitude and the longitude.

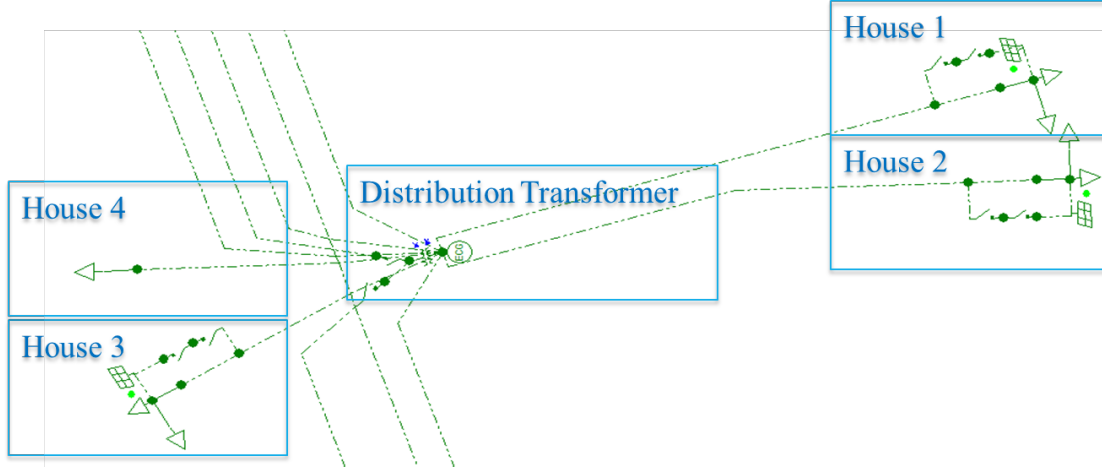


Figure 3.2. Single Line Diagram of a Small Portion of the Cyme Model.

3.2.1 Net-metering Calculation

There are two classes of loads: metered loads (Ω_{AMI_D}) and unmetered loads (Ω_{D1}). For the metered loads, the active power definition is based on the AMI measurements of the derived active power available at that premise. Figure 3.3 shows the metering infrastructure installed at a typical metered premise, where the household has separate meters for PV production and billing. The load, active power definition for metered loads is represented by the gross load ($d_l^D, l \in \Omega_{AMI_D}$), which is the total active power demand at a household. The production meter measurements are used to derive the active power produced by the PVs ($P_l^{PV}, l \in \Omega_{PV}$) and to define the PV generation. The billing meter is a bi-directional meter whose measurements are used to derive the power delivered by the utility to the customer (P_l^D) and the power received by the utility from the customer (P_l^R). Hence, (3.1) is used to estimate the gross load of a metered premise.

$$d_{l,\phi}^D = P_{l,\phi}^D - P_{l,\phi}^R + P_{l,\phi}^{PV}, \forall l \in \Omega_{AMI_D}, \phi \in \psi(l) \quad (3.1)$$

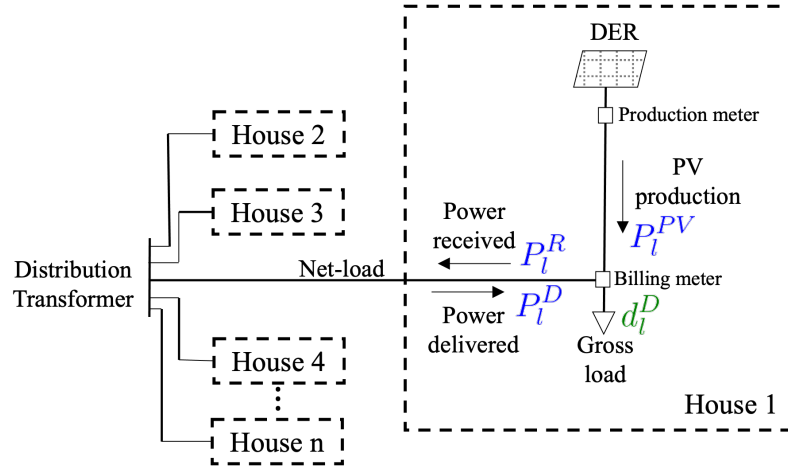


Figure 3.3. Metering Infrastructure at a Typical Premise: Available Measurements are in Blue, Estimated Values are in Green.

On the other hand, the power definition for unmetered loads ($P_l^D, l \in \Omega_{D1}$) is set as a variable to be estimated by the enhanced method.

3.2.2 Geographic Information System (GIS)

The GIS data of the sections are given in terms of the latitude and the longitude of both endpoints of all the sections, including the secondary sections. By combining this data with the CYME model, locations of all the other elements such as transformers, loads, capacitors, and PV units are inferred. This data is then imported to QGIS [41], which is an open source geographical information software. Further, to validate the circuit, aerial imagery obtained from Google Maps is added as an additional layer in QGIS. Figure 3.4 shows a small part of this GIS model, it is apparent that the secondary locations of the loads and PV do correspond to individual houses. Further, PV panels can be observed on the houses with a PV unit in the electrical model.

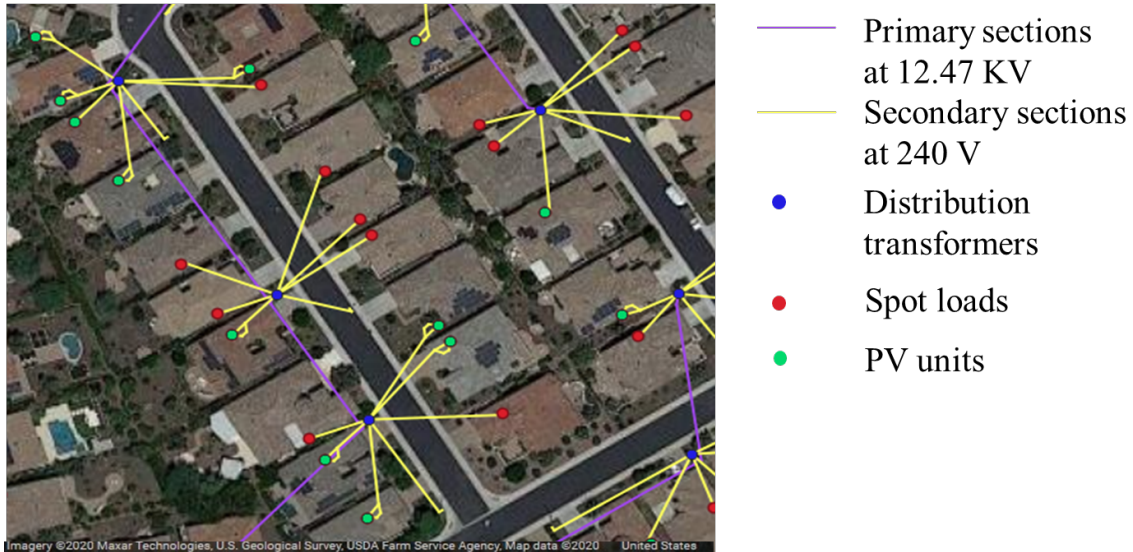


Figure 3.4. GIS Model of the Feeder.

3.3 Process for Constructing the Secondary Network Topology for the Feeder Model

The following steps are followed while constructing secondary network topology:

1. A preliminary OpenDSS circuit model in steady state is constructed by converting the CYME model available from the database. A new tool is developed for this conversion. This tool is presented in Section 3.3.1.
2. Loads from the CYME database are included for the locations where the load meter measurements are not available, but where there is a load defined in CYME database. The GIS model of the circuit is used to identify these locations.
3. A misplaced loads correction process is carried out.

3.3.1 DISMOTT: Distribution System Model Transformation Tool

An interface is developed to convert the feeder from the CYME data format to provide input to OpenDSS and successfully solve a three-phase power flow. No ready-made converter supports recent versions of CYME (8.x). Therefore, a converter that retains

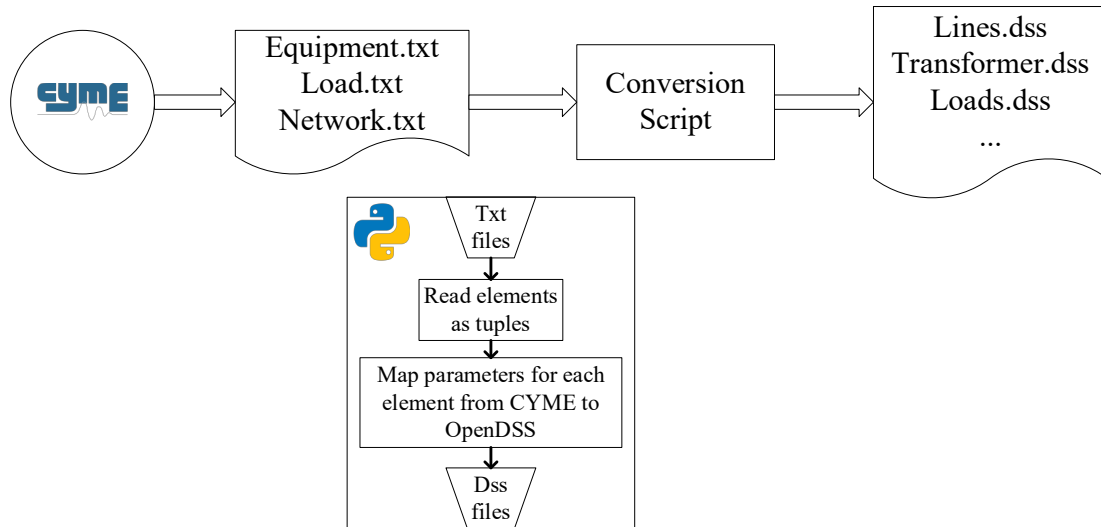


Figure 3.5. Flowchart for the CYMEDIST to OpenDSS Conversion Script.

all the equipment information while obtaining the same power flow results is necessary to accurately model the feeders. A Python script is developed for converting data from CYMEDIST to OpenDSS. A version of this tool was released as an open-source tool [40]. The operation of this script is explained as a flowchart in Figure 3.5.

The key challenge in the conversion process is that not all component models in OpenDSS and CYME are identical. For instance, cables and transformers are modeled differently. Hence, an accurate mapping of parameters when the CYMDIST and OpenDSS models are different is required.

3.3.2 Misplaced Loads Correction

After placing the loads using the CYME and GIS databases, the following procedure is carried out to move the loads that were misplaced in the utility databases:

1. The AMI database contains power readings per meter, linking the meters to their corresponding premises. No misplaced loads are identified in this database.

2. On the other hand, the CYME database contains group of loads connected at the same location. This group of loads generates an overload of more than 100% at the distribution transformer that they are connected to. Then, these loads which have a repeated location were identified as misplaced loads.
3. Once the misplaced loads were identified, the concentration of misplaced loads in each phase of the feeder according to their misplaced location was calculated.
4. A power flow was performed without the misplaced loads recognized in the second step. This power flow is used to calculate the phase current values at the head of the feeder and is compared with the values of current measured at the head of the feeder.
5. A ratio of how much load is needed to be moved to each phase to match the current measurement at the head of the feeder is calculated. This proportion is applied to the total load from step 2.
6. Those transformers loaded with less than 20% are now considered as locations to move the misplaced loads from step 2. It was found that 45 transformers were loaded at less than 20% of their rating. Transformers that did not have any load were also considered.
7. To move the misplaced loads, the GIS location of the transformers from step 6 is used to plot the loads currently connected to those distribution transformers and identify which premises do not currently have any load. The idea is to examine where they are connected and to what phase and select the nearest distribution transformer with low load to move the misplaced load.
8. 60 premises were identified without any load connected to them. These premises are used to locate the misplaced loads.

3.4 High PV Penetration in Distribution Systems

To highlight the importance of modeling the secondary side of a distribution network with high penetration of PV, a time-series voltage analysis is carried out for a single distribution transformer. Figure 3.6 shows the 24-hour voltage profile for a single distribution transformer as well as the voltage profile of the premises that are connected to that distribution transformer.

As observed in Figure 3.6, if the distribution transformer voltage is selected to model the secondary network using an aggregate model of the load, it would result in the misrepresentation of the secondary elements' effect, showing different voltages across the elements connected to the secondary side of the distribution transformer. In this case, if only the transformer is modeled, the overvoltage occurring at a premise will not be captured in the time series power flow. Therefore, modeling the secondary side of the distribution system in detail, as performed in this work, is essential to assess a distribution network

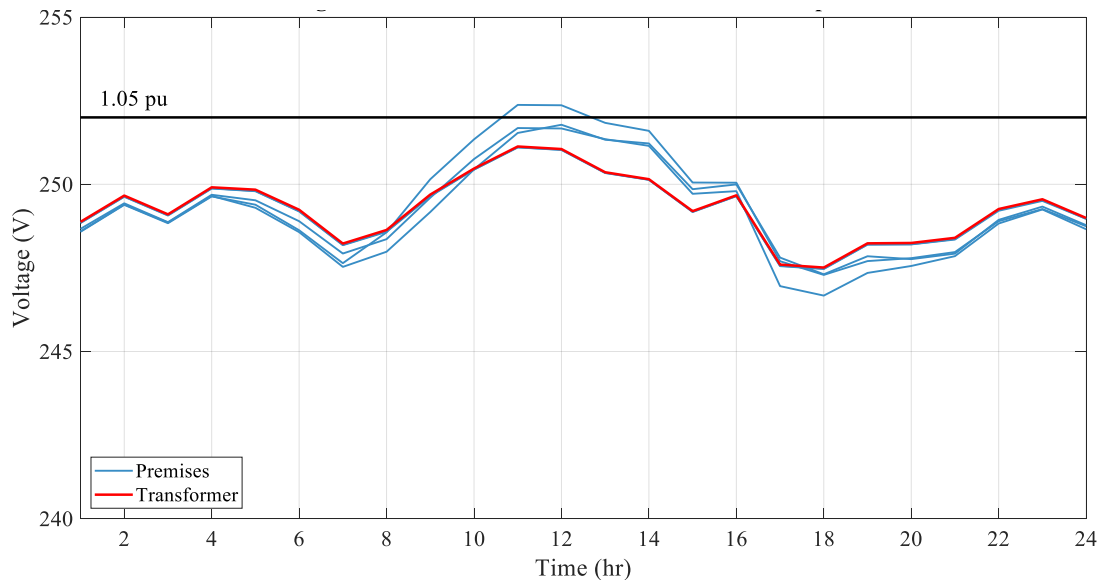


Figure 3.6. Distribution Transformer Voltage Profile Vs Premises Connected at the Distribution Transformer Voltage Profiles.

with high PV penetration, which as in this case, tends to have over voltages on the secondary side of the system.

Figure 3.7 shows the active power produced by all PV units connected to two different transformers. It is seen that they can differ in shape due to PV rating as well as to external factors such as clouds, module orientation, among others. Hence, these PV units need to be modeled individually to capture this variation.

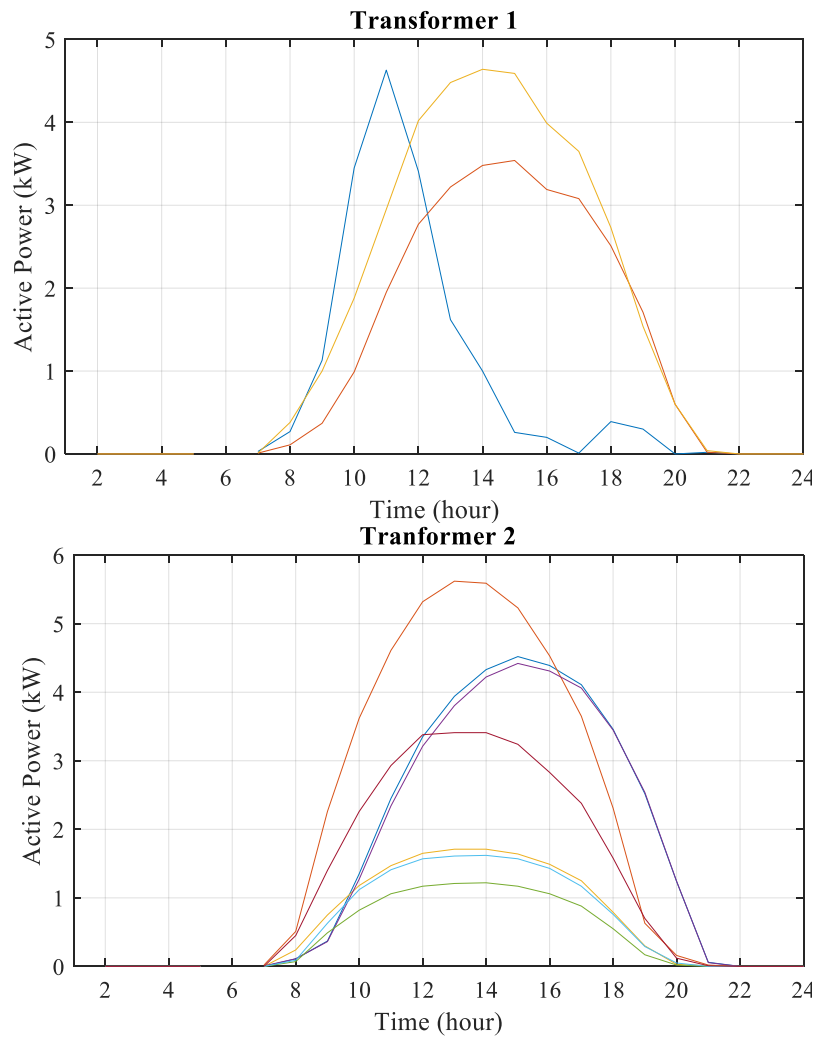


Figure 3.7. PV Profile for All the Premises Connected to the Same Distribution Transformer.

CHAPTER 4

DISTRIBUTION SYSTEM MODEL ENHANCEMENT ALGORITHM

This chapter presents a highly automated, novel method to enhance the accuracy of utility distribution feeder models to capture their performance by matching simulation results with corresponding field measurements [42].

4.1 Modeling Data Resources

The objective of the proposed model enhancement method is to obtain an accurate and detailed distribution system model to capture the performance of the feeder by matching simulation results with corresponding field measurements.

Figure 4.1 illustrates the modeling data resources for the proposed distribution system model enhancement method.

The proposed method relies on the following assumptions regarding the available data:

1. The secondary network topology is assumed to be known. If the topology is unknown, the approach presented in Chapter 3 [39] can be used to estimate the system topology.
2. The following feeder-head measurements are assumed to be available and accurate: feeder-head total three-phase active power (P^H) and reactive power (Q^H) and feeder-head voltage magnitude ($|\hat{V}_{i,\phi}|, i \in \Omega_H$) for at least one phase, ϕ .
3. The gross load of the metered loads ($d_l^D, l \in \Omega_{AMI_D}$) is assumed to be available or derivable from the AMI measurements available in the system.

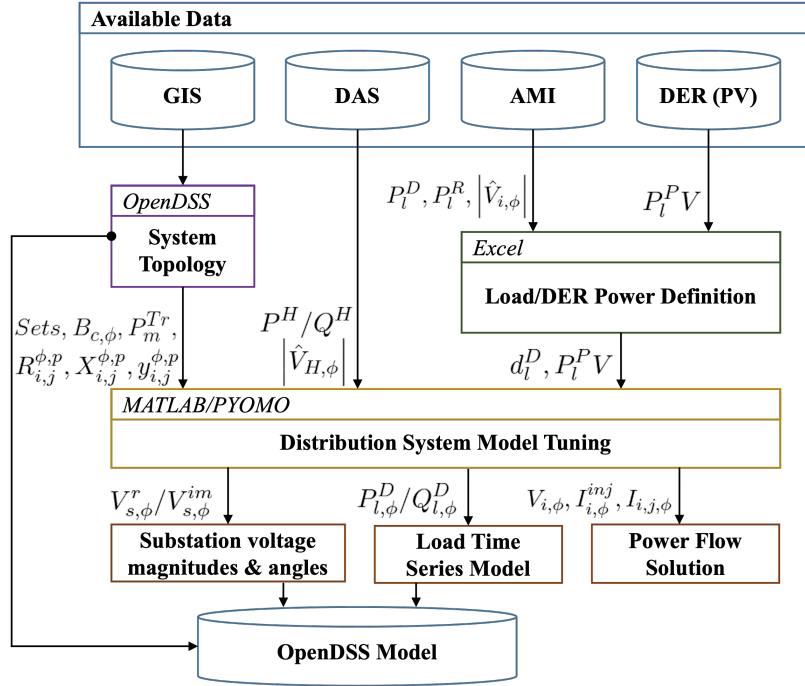


Figure 4.1. Modeling Data Resources for Distribution System Model Enhancement.

4. The active/reactive powers of the DERs ($P_l^{PV}/Q_l^{PV}, l \in \Omega_{AMI}$) are assumed to be available from the AMI measurements available in the system. In the system considered, the reactive power of the DERs ($Q_l^{PV}, l \in \Omega_{AMI}$) is assumed to be zero - at present, all DERs in the system operate at unity power factor and do not participate in any reactive power support of the system.
5. Load voltage measurements ($|\hat{V}_{i,\phi}|, \forall i \in \Omega_{AMI_V}$) are assumed to be available for some of the loads. However, it is not necessary to have these voltage measurements for the same loads as in Assumption 3.
6. The field measurements available are assumed to be accurate.

The system topology, feeder-head measurements, load voltage measurements available, DERs power definition, and the initial power definition of the loads are processed in MATLAB to create the input data to the optimization-based method. The ACOPF based on an IV formulation for distribution system model enhancement is programmed using Pyomo, which is a Python-based, open-source optimization modeling language [43], [44]. IPOPT is used as the nonlinear solver for the proposed ACOPF approach [45]. The output from the distribution system model enhancement includes time-series substation voltage magnitudes and angles, load active and reactive power demand, and power flow solution. The power flow from the resulting enhanced distribution system model is compared with the power flow from OpenDSS to validate the accuracy of the proposed optimization-based method.

4.2 Distribution System Model Enhancement Algorithm

This section presents the formulation of the proposed method, the data implementation to create the input data for the algorithm, the formulation of the proposed ACOPF approach, and the algorithm output and simulation capabilities. A flowchart explaining the optimization technique proposed is summarized in Figure 4.2.

4.2.1 Input Data and Initialization

The input data contains the feeder topology information and AMI/DAS measurement data. The feeder topology data includes the information to create the subset of buses with load, PV, capacitor and transformers, and impedance information of each distribution line. The AMI measurement data includes the energy measurements (*kWh*)

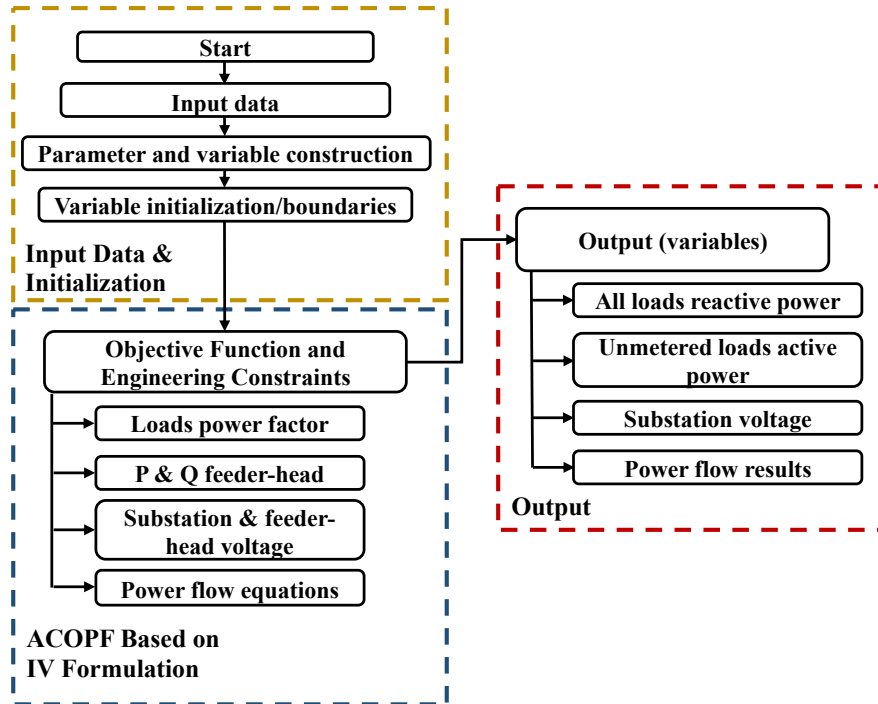


Figure 4.2. Flowchart for Distribution System Model Enhancement Algorithm.

used to derive the hourly active power (kW) for the metered load. The AMI database also includes voltage magnitude measurements for some loads in the feeder. The DAS measurement data include total active/reactive power at the feeder-head and the voltage magnitude of one phase. The detailed procedure for creating the input data is described in Algorithm 1.

Most nonlinear solvers only find local optimal solutions for nonconvex problems [46]. Therefore, a good initialization is essential to find a solution that meets the problem requirements. After the input data is created, the parameters and variables are defined. The parameters are listed under the assumptions in Section 4.1. The variables in the enhancement algorithm include bus voltages, unmetered load active power definitions, load reactive power definitions, bus injection currents, line flow currents, reactive power production of capacitors, and per phase substation active/reactive powers and voltages. The

Algorithm 1: Input data creation

Input: Feeder topology and AMI/DAS measurements.

Output: Distribution system model enhancement algorithm input data.

Read feeder topology and line impedances $(R_{i,j}^{\phi,p}, X_{i,j}^{\phi,p}, y_{i,j}^{\phi,p})$.

Create topology sets $(\Omega_N, \Omega_D, \Omega_{PV}, \Omega_{cap}, \Omega_{SB}, \Omega_L, \Omega_H, \Omega_T)$.

Read equipment information $(B_{c,\phi}, P_m^{Tr})$.

Read DAS measurements $(P^H, Q^H, |\hat{V}_{H,\phi}|)$.

Read AMI measurements of energy (kWh) and voltage magnitude $(|\hat{V}_{i,\phi}|)$.

Create sets of buses with measurements $(\Omega_{AMV}, \Omega_{AMID})$.

Write line/transformer data $(R_{i,j}^{\phi,p}, X_{i,j}^{\phi,p}, y_{i,j}^{\phi,p})$ for the algorithm.

FOR hour h in $1 \dots 24$ **DO**

Derive hourly active power measurements at each meter by aggregating the measurements of energy consumption $(P_l^D, P_l^R, P_l^{PV}, \forall l \in \Omega_{AMID})$.

Derive the gross load $(d_l^D, l \in \Omega_{AMID})$ using $()$.

Read the PV active power $(P_l^{PV}, l \in \Omega_{PV})$.

Write gross load and PV active power data.

Write equipment data $(B_{c,\phi}, P_m^{Tr})$ for the algorithm.

Write voltage measurement data $(|\hat{V}_{i,\phi}|)$ for the algorithm.

END FOR

bus voltage magnitudes, bus injection currents, and line flow currents are initialized using a flat start: balanced voltages of magnitude 1.0 p.u. and line currents of magnitude 0 p.u. for all the buses and lines. The active powers for unmetered loads may be initialized based on the type of load (residential/commercial/industrial) and the feeder location or based on the active power measurements from other metered loads available. However, this initialization can be adjusted without loss of generality, knowing whether the load is residential, commercial, or industrial. The reactive power for all the loads is initialized using a constant power factor of 0.9 lagging. The reactive power for all the capacitor banks is initialized using a terminal voltage of 1 p.u. and the nominal capacitance using (4.8). The substation per phase active/reactive powers are initialized using one-third of the total three-

phase measurement at the feeder-head. The substation voltage is initialized at the same voltage as the feeder-head measurement.

4.2.2 Enhancement Algorithm: ACOPF Based on Current-Voltage (IV) Formulation

This section formulates the optimization-based problem proposed for the distribution system model enhancement method as an ACOPF formulation based on a three-phase IV model for unbalanced distribution networks with mutual impedance, which is more appropriate for distribution networks [37], [47]. The IV formulation solves a linear system of equations without decomposition, unnecessary constraints, or omissions, and it may be computationally easier to solve than the traditional quadratic power flow formulations [46]. The ACOPF presented in this work co-optimizes active and reactive power along the distribution feeder. The nonlinear formulation is carried out in rectangular coordinates. The proposed model enhancement algorithm reads the input data and initial values of the variables and transfers the information through the engineering constraints while optimizing the system performance according to the objective function, shown in Figure 4.2.

The objective of this formulation is to minimize the norm of the difference between the solved bus voltage magnitudes and the corresponding AMI voltage measurement data. Since the rectangular representation of currents and voltages is considered, the square of the voltage magnitude is used, which is equal to the summation of the squares of the real part and imaginary parts of the voltage. The objective function formulation is shown in (4.1).

$$\min \sum_{i \in \Omega_{AMI_V}} \sum_{\phi \in \psi} \left(V_{i,\phi}^r{}^2 + V_{i,\phi}^{im2} - |\hat{V}_{i,\phi}|^2 \right)^2 \quad (4.1)$$

Let i and j be the indices of the sending and receiving buses of a line (i, j) . For line flow equations, the mathematical relation between the voltage difference between the two buses (i, j) for each phase (ϕ) of a line and the current flow for each phase of a line in an unbalanced three-phase distribution system are expressed in (4.2)-(4.3).

$$V_{i,\phi}^r - V_{j,\phi}^r = \sum_{p \in \psi} R_{i,j}^{\phi,p} \left(I_{i,j,p}^r + \frac{1}{2} \sum_{k \in \psi} y_{i,j}^{p,k} V_{i,k}^{im} \right) - \sum_{p \in \psi} X_{i,j}^{\phi,p} \left(I_{i,j,p}^{im} - \frac{1}{2} \sum_{k \in \psi} y_{i,j}^{p,k} V_{i,k}^r \right),$$

$$\forall (i, j) \in \Omega_L, \phi \in \psi \quad (4.2)$$

$$V_{i,\phi}^{im} - V_{j,\phi}^{im} = \sum_{p \in \psi} R_{i,j}^{\phi,p} \left(I_{i,j,p}^{im} - \frac{1}{2} \sum_{k \in \psi} y_{i,j}^{p,k} V_{i,k}^r \right) + \sum_{p \in \psi} X_{i,j}^{\phi,p} \left(I_{i,j,p}^r + \frac{1}{2} \sum_{k \in \psi} y_{i,j}^{p,k} V_{i,k}^{im} \right),$$

$$\forall (i, j) \in \Omega_L, \phi \in \psi \quad (4.3)$$

The real and imaginary parts of the current injection constraint are defined using (4.4) and (4.5), respectively.

$$I_{i,\phi}^{r,inj} = \sum_{j \in \delta(i)} I_{i,j,\phi}^r, \forall i \in \Omega_N, \phi \in \psi \quad (4.4)$$

$$I_{i,\phi}^{im,inj} = \sum_{j \in \delta(i)} I_{i,j,\phi}^{im}, \forall i \in \Omega_N, \phi \in \psi \quad (4.5)$$

The active power balance constraint is defined in (4.6). The reactive power balance constraint is defined in (4.7), considering capacitor output $(Q_{c,\phi})$.

$$\begin{aligned}
V_{i,\phi}^r I_{i,\phi}^{r,inj} + V_{i,\phi}^{im} I_{i,\phi}^{im,inj} &= \sum_{\substack{\forall s \in \Omega_{SB}, \\ s=i}} P_{s,\phi}^G - \sum_{\substack{\forall m \in \Omega_T \\ m=i}} P_{m,\phi}^{Tr} - \sum_{\substack{\forall l \in \Omega_{D1} \\ l=i}} P_{l,\phi}^D + \sum_{\substack{\forall l \in \Omega_{PV} \\ l=i}} P_{l,\phi}^{PV} \\
&- \sum_{\substack{\forall l \in \Omega_{AMI_D} \\ l=i}} d_{l,\phi}^D, \forall i \in \Omega_N, \phi \in \psi
\end{aligned} \tag{4.6}$$

$$\begin{aligned}
V_{i,\phi}^{im} I_{i,\phi}^{r,inj} - V_{i,\phi}^r I_{i,\phi}^{im,inj} &= \sum_{\substack{\forall s \in \Omega_{SB}, \\ s=i}} Q_{s,\phi}^G + \sum_{\substack{\forall c \in \Omega_{cap} \\ c=i}} Q_{c,\phi}^C \\
&- \sum_{\substack{\forall l \in \Omega_D \\ l=i}} Q_{l,\phi}^D + \sum_{\substack{\forall l \in \Omega_{PV} \\ l=i}} Q_{l,\phi}^{PV}, \forall i \in \Omega_N, \phi \in \psi
\end{aligned} \tag{4.7}$$

where the reactive power output of a connected capacitor is modeled using a constant capacitance model. Therefore, the reactive power is expressed as follows:

$$Q_{c,\phi}^C = B_{c,\phi} \left(V_{c,\phi}^r{}^2 + V_{c,\phi}^{im}{}^2 \right), \forall c \in \Omega_{cap}, \phi \in \psi \tag{4.8}$$

The substation model shown in Figure 4.3 is proposed and implemented to capture the unbalanced nature of distribution substations in practical utility feeders (Section 4.2.4). The voltage magnitude limits at the feeder-head are expressed in (4.9).

$$V_{i,\phi}^{min}{}^2 \leq V_{i,\phi}^r{}^2 + V_{i,\phi}^{im}{}^2 \leq V_{i,\phi}^{max}{}^2, \forall i \in \Omega_H, a \neq \phi \in \psi \tag{4.9}$$

The upper and lower bounds in (4.9) are established according to the feeder-head available measurements of voltage ($|\hat{V}_{i,\phi}|, i \in \Omega_H$). The feeder modeled in this work has available hourly measurements of voltage magnitude at the feeder-head for phase a

$(|\hat{V}_{H,a}|)$, which is considered a parameter for this phase by the model enhancement algorithm. On the other hand, the voltage magnitude at the feeder-head of phases b and c are calculated individually by the algorithm proposed. Hence, the voltage magnitudes of phases b and c are allowed a deviation of 2% from the measured phase a voltage via (4.9) – the limits $V_{i,\phi}^{min}$ and $V_{i,\phi}^{max}$ should be set based on the available measurements for the feeder modeled. For a distribution feeder with separate voltage magnitude measurements available for all three phases ϕ , all three feeder-head voltages $(|\hat{V}_{H,\phi}|, \phi \in \psi)$ would be considered as parameters.

The voltage magnitude limits for the source behind the equivalent impedance are expressed in (4.10). The voltage magnitudes of the three phases for the source are calculated individually by the algorithm proposed with a maximum deviation of 2.5% from $|\hat{V}_{H,a}|$. The limits in (4.10) are determined based on the order of magnitude of the source impedance and to facilitate higher feasibility for the optimization algorithm.

$$V_{s,\phi}^{min} \leq |V_{s,\phi}| \leq V_{s,\phi}^{max}, \forall s \in \Omega_{SB}, \phi \in \psi \quad (4.10)$$

The real and imaginary part of the voltage at the substation are given by (4.11)-(4.12), respectively.

$$V_{s,\phi}^r = |V_{s,\phi}| \cos(\angle V_{s,\phi}), \forall s \in \Omega_{SB}, \phi \in \psi \quad (4.11)$$

$$V_{s,\phi}^{im} = |V_{s,\phi}| \sin(\angle V_{s,\phi}), \forall s \in \Omega_{SB}, \phi \in \psi \quad (4.12)$$

where the voltage angles are limited using (4.13). The limits in (4.13) are set to allow the consideration of an unbalanced source while still maintaining a roughly 120° angle difference between any two phases.

$$\angle V_{s,\phi}^{min} \leq \angle V_{s,\phi} \leq \angle V_{s,\phi}^{max}, \forall s \in \Omega_{SB}, \phi \in \psi \quad (4.13)$$

The reactive power of each load is limited by its maximum and minimum power factor using (4.14)-(4.15).

$$d_{l,\phi}^D \sqrt{\left(\frac{1}{(PF_l^{max})^2} - 1\right)} \leq Q_{l,\phi}^D \leq d_{l,\phi}^D \sqrt{\left(\frac{1}{(PF_l^{min})^2} - 1\right)}, \quad \forall l \in \Omega_{AMI_D}, \phi \in \psi \quad (4.14)$$

$$P_{l,\phi}^D \sqrt{\left(\frac{1}{(PF_l^{max})^2} - 1\right)} \leq Q_{l,\phi}^D \leq P_{l,\phi}^D \sqrt{\left(\frac{1}{(PF_l^{min})^2} - 1\right)}, \quad \forall l \in \Omega_{D1}, \phi \in \psi \quad (4.15)$$

where PF_l^{max} and PF_l^{min} are the limits on the power factors of the loads. These limits should be selected to be suitable for the modeled feeder. Since DAS measurements of the total three-phase active and reactive power are available at the feeder-head, the summation of power injections in the different phases at the feeder-head is assumed to be equal to the measured value. Since DAS measurement of the total three-phase active and

reactive power are available at the feeder-head, the summation of power injections at the different phases at the feeder-head is assumed to be equal to the measured value.

$$P^H = \sum_{\phi \in \psi} (V_{h,\phi}^r I_{h,\phi}^{r,inj} + V_{h,\phi}^{im} I_{h,\phi}^{im,inj}), \forall h \in \Omega_H \quad (4.16)$$

$$Q^H = \sum_{\phi \in \psi} (V_{h,\phi}^{im} I_{h,\phi}^{r,inj} - V_{h,\phi}^r I_{h,\phi}^{im,inj}), \forall h \in \Omega_H \quad (4.17)$$

4.2.3 Algorithm Output and Simulation Capabilities

The algorithm developed solves a three-phase distribution system power flow problem. The power flow can be solved in standard single snapshot mode and daily variable time-interval mode. The time interval can be any time period. The feeder model developed in this work is solved for each day using 24-hourly steps. When the power flow is solved, the losses, voltages, flows, and other information are available for the entire system and each element.

For each instant in time, the algorithm automatically exports the power flow solution of the system, as well as the active and reactive power definition of the loads ($P_{l,\phi}^D/Q_{l,\phi}^D, \forall l \in \Omega_D$), distributed generators output ($P_l^{PV}, \forall l \in \Omega_{PV}$), and source voltage magnitudes and angles ($|V_{s,\phi}|/\angle V_{s,\phi}, \phi \in \psi$), as shown in Figure 4.2. These data are then used to complete the OpenDSS model, as explained in the next section.

4.2.4 OpenDSS Optimal Feeder Model

An OpenDSS time-series feeder model is constructed based on the optimization algorithm results as the output of the feeder model enhancement method. This enhanced OpenDSS model is then used for any further studies involving the distribution feeder modeled. For this time series model, the voltage source and each load and PV generator follow hourly profiles obtained from the enhancement algorithm and transferred to OpenDSS through shape files. The profiles are created for active and reactive power for each load/PV generator and the voltage magnitude of each of the three phases at the unbalanced source. The loads are modeled using constant P and constant Q to preserve the power flow obtained from the enhancement algorithm. The solar PV units are modeled as electronically coupled generators using the current-limited constant kW OpenDSS model and based on the AMI measurements.

OpenDSS models a three-phase voltage source as a balanced voltage source behind a Thévenin equivalent impedance. However, balanced sources at the distribution network are not accurate enough to represent the unbalanced voltages common in practical utility feeders [48]. Therefore, to model the unbalanced distribution system more accurately, the optimization-based method proposed implements three independent single-phase voltage sources to model an unbalanced substation source, as shown in Figure 4.3. Since the feeder-head measurement of voltage is available only for the magnitude of one phase ($|\hat{V}_{H,a}|$) for the modeled feeder, the feeder-head voltage for the remaining phases and the source voltages are obtained by the ACOPF formulation proposed.

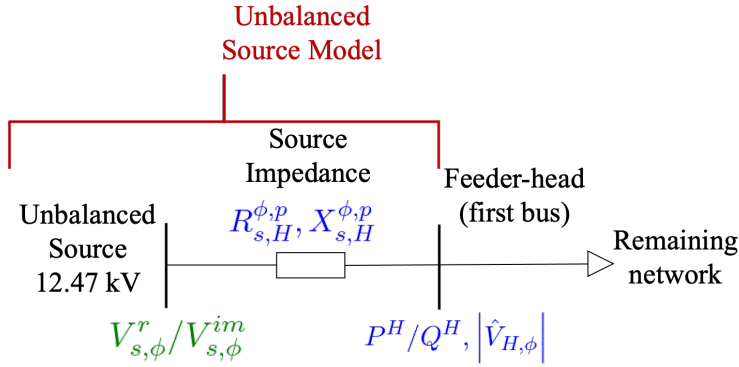


Figure 4.3. Source and Feeder-head Representation in the Proposed Framework: Available Measurements are in Blue, Estimated Values are in Green. Three Independent Single-phase Voltage Sources are Implemented in the Unbalanced Source Bus to Model an Unbalanced Circuit.

The voltage source and feeder-head representation are shown in Figure 4.3 and used in the algorithm corresponds to the Thévenin equivalent impedance representing the substation transformer and the sub-transmission/transmission system. In the enhancement algorithm, this source impedance is represented as a line with no capacitance ($R_{s,H}^{\phi,p}, X_{s,H}^{\phi,p}, \phi, p \in \psi$). As a result, three shapefiles corresponding to the voltage magnitude for each of the three phases at the substation are created for the OpenDSS model. On the other hand, while the enhancement algorithm provides the voltage angle for each phase at each step of the time series simulation, OpenDSS does not allow voltage source angle variation through shapefiles. Therefore, unbalanced angles are defined at each of the three single-phase sources at the substation bus and hence are kept constant while solving a time series power flow problem in OpenDSS.

Monitors are then set at the feeder-head bus and every load bus in the system to capture the results of voltages and powers at those points. Then, the results of the monitors

are compared against the AMI measurements at the loads and DAS measurements at the feeder-head to validate the system. The model validation is discussed in Chapter 5.

CHAPTER 5

MODEL VALIDATION: UTILITY FEEDER RESULTS

This chapter presents the validation between the enhancement algorithm power flow output and OpenDSS power flow and the validation between the enhancement algorithm power flow output and the field measurements [42].

Since the modeled feeder is a residential feeder in Arizona, and since a typical household peak load in Arizona lies between 4-7 kW due to the need for air conditioning, a value of 5 kW is chosen to initialize the unmetered loads' active power. For loads connected to a distribution transformer with other measured loads, similar values as those measured loads are used to initialize their active power instead. For the feeder modeled, the limits of the power factors of the loads (PF_l^{max} and PF_l^{min}) are selected to include the power factor at the feeder-head - obtained from the measurements of active and reactive power at the feeder-head for a time-step. Most solar PV units installed on this feeder operate with unity power factor in practice. Therefore, the reactive power exchanged by the solar PV units ($Q_{l,\phi}^{PV}$) with the rest of the distribution system is set to zero in (4.7).

The optimization algorithm presented is implemented using Pyomo (version 5.7) and solved using IPOPT solver (version 3.11.1). The algorithm is implemented on a computer with a 4-core 1.8 GHz Intel Core i7-8550U CPU and 16 GB of RAM. On this machine, optimizing the feeder model for one snapshot (once loaded) requires approximately 32 seconds and involves 134,623 variables and 133,264 equality and inequality constraints.

5.1 Algorithm- OpenDSS Power Flow Comparison

This section presents the validation between the enhancement algorithm power flow output and OpenDSS power flow. To perform the validation, the actual historical feeder peak load snapshot on 07/15/2019 (high load and relatively low PV) is chosen for the analysis. The resultant load active and reactive powers from the enhancement algorithm, the per-unit voltages at the three phases of the source, and the input data such as the network topology and parameters, capacitors, and solar PV generation are implemented in an OpenDSS model. Then, the power flow obtained from the optimization algorithm and the power flow solution obtained from OpenDSS for the same operating point are compared to validate the representation of the power flow equations in the optimization algorithm. A good match between the two would imply that the optimization algorithm accurately represents the power flow equations.

For this comparison, the consumed/produced active and reactive power for each element in the feeder are calculated from the algorithm solution and compared against the corresponding power element losses exported from OpenDSS. The consumed/produced active and reactive power for each element in the feeder are calculated according to the type of the element.

For all the lines (Ω_L) in the feeder, where i and j are the sending and receiving bus indices, the active and reactive power consumption (inductive lines) and reactive power production (capacitive underground lines) are calculated as the power difference between the sending and receiving ends.

Similarly, for all the transformers (Ω_T) in the feeder, the active and reactive power consumption is calculated as the power difference between the sending and receiving ends

plus the no-load loss ($P_m^{Tr}, \forall m \in \Omega_T$), which represents a resistive branch in parallel with the magnetizing inductance. For all the capacitors on (Ω_{cap}), (4.8) is used to calculate the reactive power injections.

The algorithm uses (4.6)-(4.7) to calculate the load active power ($d_l^D, \forall l \in \Omega_{AMI_D}; P_l^D, \forall l \in \Omega_{D1}$) and reactive power ($Q_l^D, \forall l \in \Omega_D$). The active and reactive power production from the solar PV units ($P_l^{PV}, Q_l^{PV}, \forall l \in \Omega_{PV}$) are input from the AMI data and are held to be the same between the optimization algorithm and the OpenDSS model.

The source active and reactive powers are calculated by using the receiving end (feeder-head end) of the line that connects the substation with the feeder-head, that is,

$$P_{s,\phi}^G = V_{H,\phi}^r I_{s,H,\phi}^r + V_{H,\phi}^{im} I_{s,H,\phi}^{im}, \phi \in \psi \quad (5.1)$$

$$Q_{s,\phi}^G = V_{H,\phi}^{im} I_{s,H,\phi}^r - V_{H,\phi}^r I_{s,H,\phi}^{im}, \phi \in \psi \quad (5.2)$$

The comparison between the consumed/produced active and reactive power for each type of element in the optimization algorithm power flow solution and the corresponding power element losses exported from OpenDSS model during a single snapshot (historical feeder load peak) is shown in Table 5-1. For instance, the row corresponding to “Lines” lists the total power consumed by all the lines in the system for both the optimization algorithm and the OpenDSS model power flow, as well as the percentage difference between both. The corresponding comparisons match under 0.17% error (except for the transformer active power comparison, which has an error of 2.95%,

equivalent to 3 kW – still relatively low considering that this 3 kW difference is the combined losses for all the distribution transformers).

TABLE 5-1. Comparison of Active and Reactive Power from Power Flow Solution.

Component	Proposed Algorithm		OpenDSS		% Error	
	P (MW)	Q (MVar)	P (MW)	Q (MVar)	P	Q
Source	-5.978	-0.522	-5.979	-0.523	0.01	0.13
Lines	0.083	-0.423	0.083	-0.423	0.17	0.06
Capacitors	0.000	-1.261	0.000	-1.260	0.00	0.01
Loads	7.725	2.166	7.728	2.167	0.03	0.02
Transformers	0.073	0.040	0.070	0.040	2.95	0.01
PVs	-1.903	0.000	-1.903	0.000	0.00	0.00

For the same snapshot (historical feeder load peak), Figure 5.1 shows a per phase comparison of the voltage magnitude of all the buses (Ω_N) between the enhancement algorithm and OpenDSS solution. Figure 5.2 shows the voltage errors between OpenDSS and the enhancement algorithm for time-series analysis (using the same load demands, solar PV generation, and source voltage) for different hours of a single day (historical feeder load peak day). The low values of errors show that the enhancement algorithm models the power flow constraints correctly and that the results obtained from the enhancement algorithm match closely with the electrical model and assumptions employed in a state-of-the-art distribution system power flow solver such as OpenDSS. Therefore, the OpenDSS time-series feeder model constructed based on the optimization algorithm results reflects the enhanced feeder model accurately and can be used for further studies involving the distribution feeder modeled.

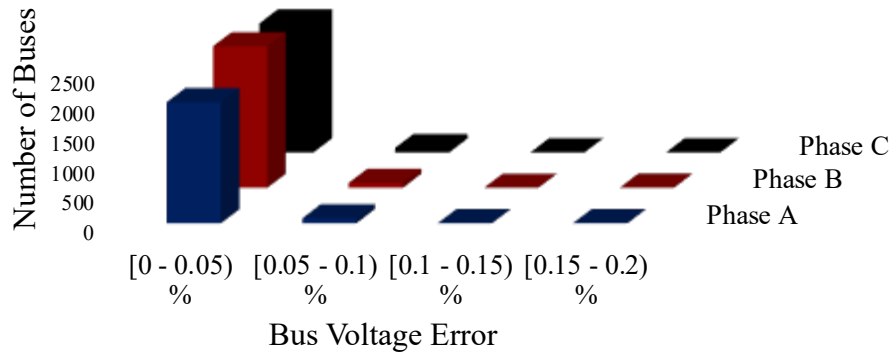


Figure 5.1. Bus Voltage Magnitude Error Per Phase Between Enhancement Algorithm Power Flow and OpenDSS Solution for a Single Snapshot for the Historical Feeder Load Peak Hour for the Modeled Feeder.

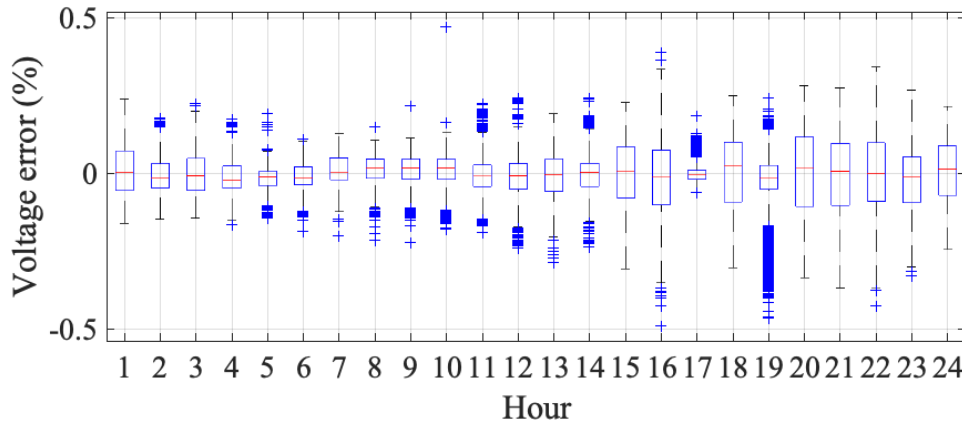


Figure 5.2. Voltage Errors Between OpenDSS and the Optimization-based Algorithm for Different Hours of a Single Day (Historical Feeder Load Peak Day) for the Modeled Feeder. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 times the Interquartile Range Away from the Box.

5.2 Enhanced Distribution Feeder Model Validation

This section presents the validation between the enhancement algorithm power flow output and the field measurements. Two days corresponding to the actual historical feeder load peak on 07/15/2019 (high load and relatively low PV) and the maximum

generation condition on 03/15/2019 (high PV) were chosen for the analysis to validate the feeder enhancement algorithm power flow.

Using the OpenDSS time-series feeder model constructed based on the optimization algorithm results from the previous sections, a time series power flow yielded a good match with the measured values. The OpenDSS feeder-head active and reactive powers are compared with the corresponding feeder-head measurements for both days in Figure 5.3 and Figure 5.4. For the historical feeder load peak day, the feeder-head active and reactive powers have root mean square (RMS) errors over a day of 0.104% and 0.234%, respectively. For the day with the maximum generation, the feeder-head active and reactive powers have RMS errors over a day of 1.897% and 0.0493%, respectively. The small RMS errors imply that the powers at the feeder-head from the enhanced OpenDSS time-series feeder model power flow accurately represents the field-measured values. Note that the reactive power along the feeder is completely calculated by the optimization-based technique by minimizing the objective function (4.1) and under the constraints of Section 4.2.2 as there are no reactive power measurements available at any point along the feeder aside from the feeder-head values. The active power is varied for a subset of the loads (Ω_{D1}) by the enhancement algorithm; therefore, a small error at the feeder-head active power indicates the model successfully being tuned to represent the measurements and, by proxy, the actual feeder status.

As further validation, the voltages at the premises along the feeder where AMI measurements are available are compared against these measurements. The comparison between the AMI measurements and model voltages for three representative meters at different locations along the feeder for the historical feeder peak load day and maximum

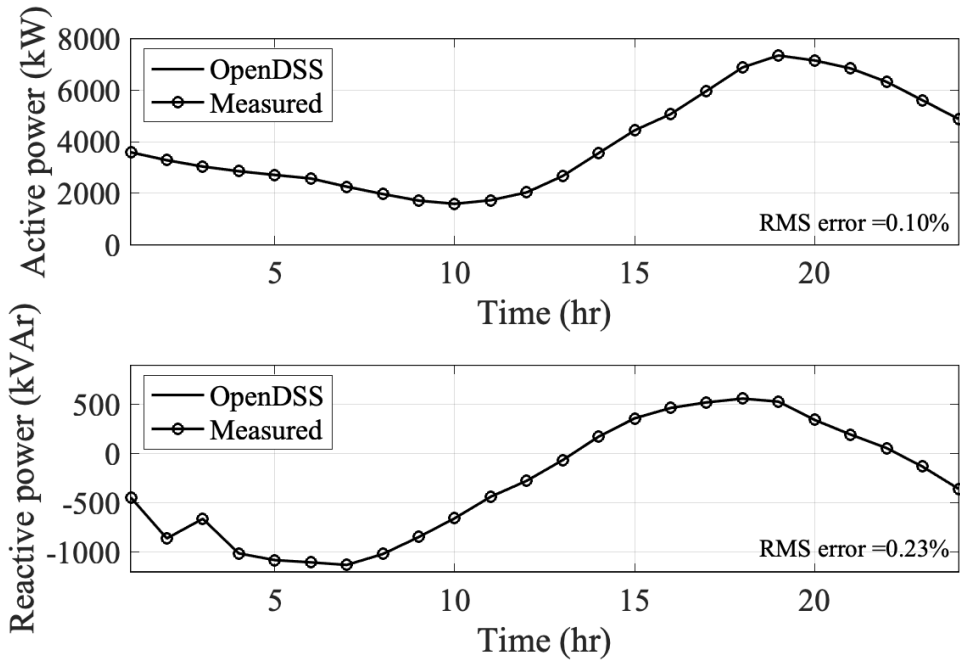


Figure 5.3. Active and Reactive Power Feeder-head Comparison Between the Enhanced OpenDSS Model and Das Measurements for the Historical Feeder Load Peak Day.

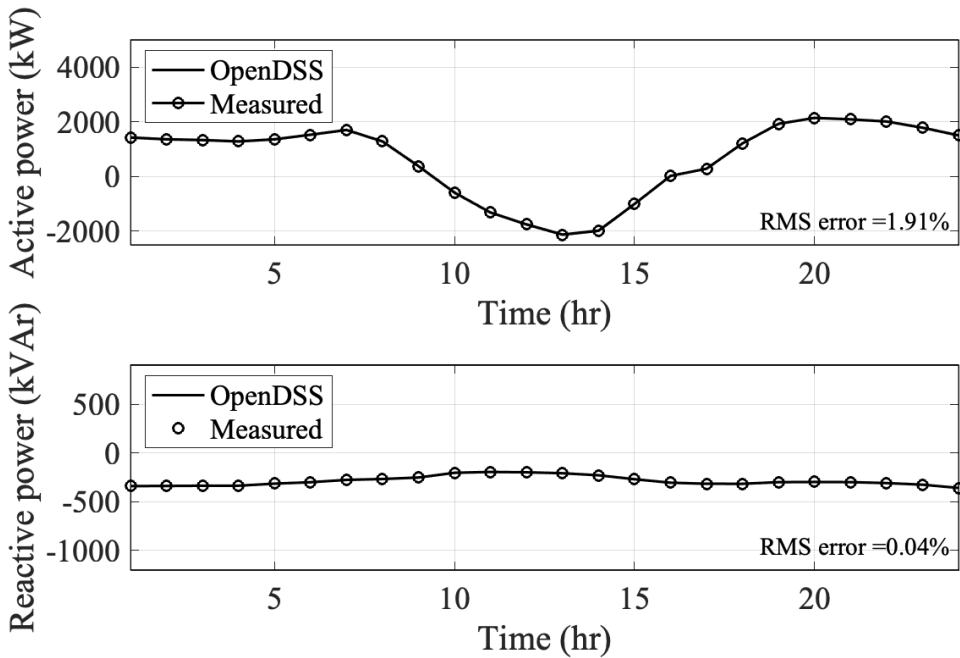


Figure 5.4. Active and Reactive Power Feeder-head Comparison Between the Enhanced OpenDSS Model and Das Measurements for the Maximum Generation Condition Day.

generation condition day are shown in Figure 5.5 and Figure 5.6, respectively. The distances from the substation of the locations corresponding to the plot are shown above each plot. The RMS error over a day is also calculated for all the meters along the feeder. Figure 5.7 shows the RMS error calculated over a day for both days in a box-and-whisker plot. As observed, the average RMS error along the feeder length is around 0.4 %, for the historical feeder load peak day and 0.2% for the maximum generation condition day, which shows that the proposed method achieves a very good match when compared against field measurements and that the enhanced feeder model accurately represents the actual field measurements.

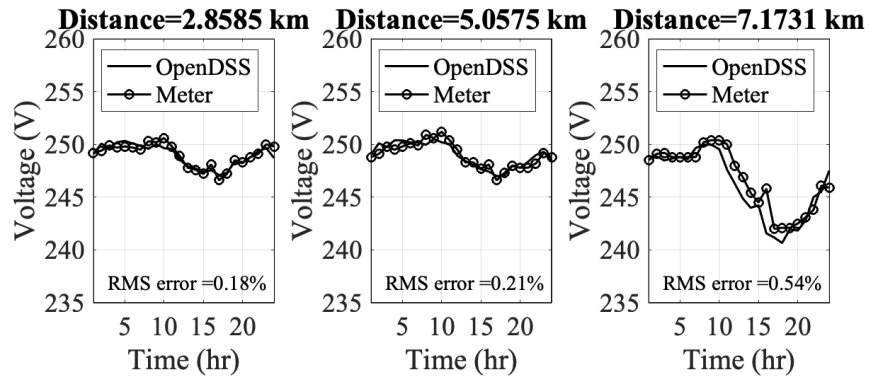


Figure 5.5. Voltage Comparison Results Between OpenDSS and AMI Data for Some Premises along the Feeder for the Historical Feeder Load Peak Day.

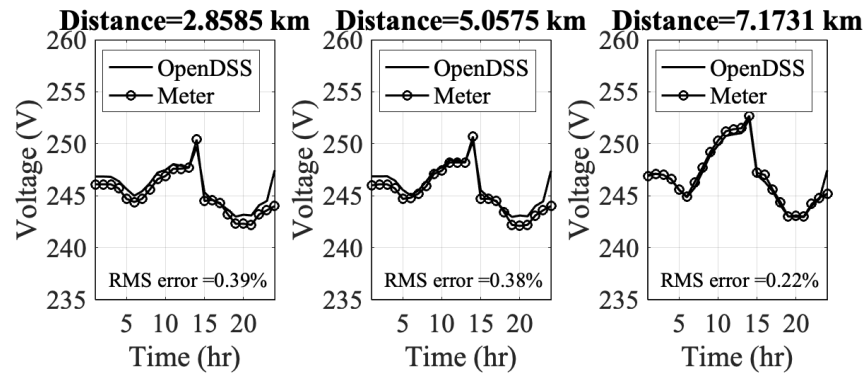


Figure 5.6. Voltage Comparison Results Between OpenDSS and AMI Data for Some Premises along the Feeder for the Maximum Generation Condition Day.

Figure 5.8, Figure 5.9, and Figure 5.10 show the AMI measurements received from the utility compared against the corresponding voltage profile obtained from the enhancement algorithm for three different snapshot conditions - historical feeder load peak, the maximum generation condition, and during the night (no PV production). The method proposed provides a complete power flow solution for both the primary and the secondary circuits by using sparse measurements at the secondary level along the feeder, extending the observability and planning capabilities of the feeder under study.

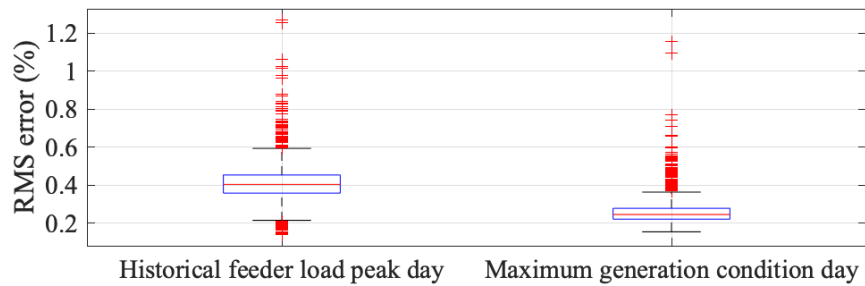


Figure 5.7. For the Two Scenarios Considered, the RMS Errors Calculated over One Day Between the Field-measured Voltages and the Voltages Obtained from the Enhanced OpenDSS Model for All the Voltage Meters in the System Are Plotted Using a Box and Whisker Plot. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 times the Interquartile Range Away from the Box.

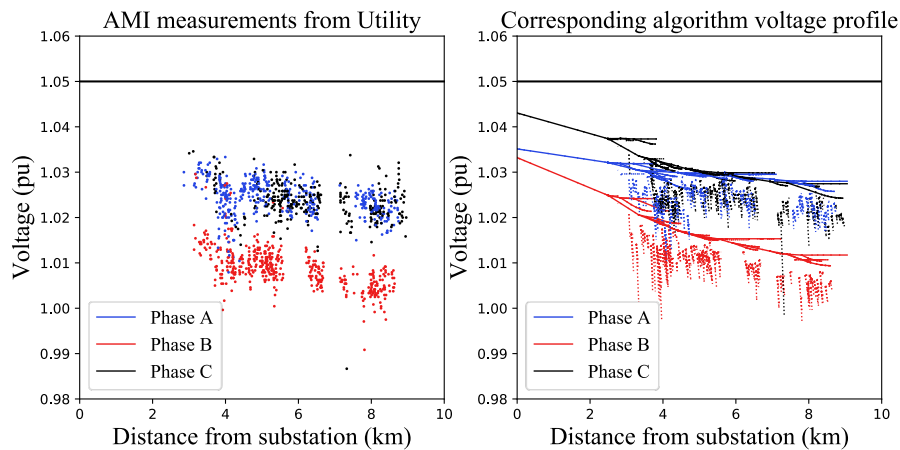


Figure 5.8. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm for the Historical Feeder Load Peak.

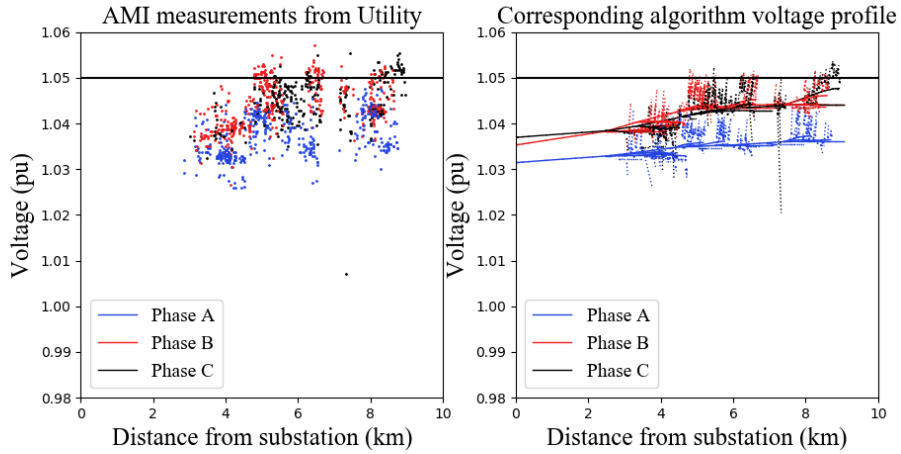


Figure 5.9. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm for the Maximum Generation Condition.

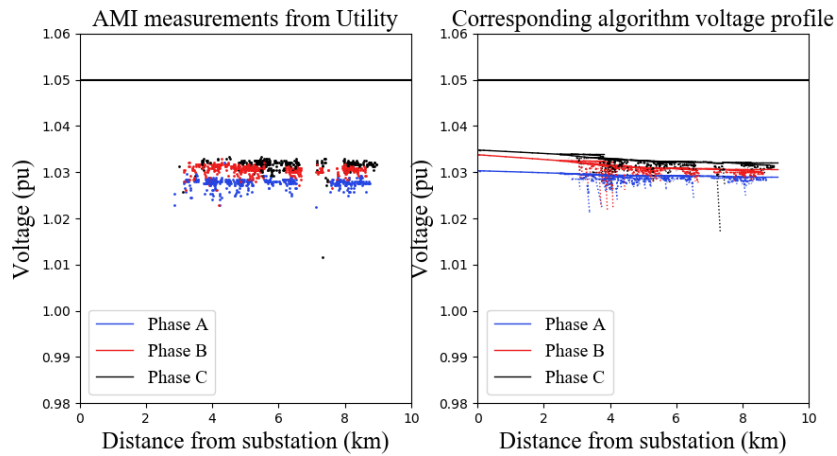


Figure 5.10. AMI Measurements from the Utility Compared Against the Corresponding Voltage Profile Obtained from the Enhancement Algorithm During the Night.

5.3 Detailed Feeder Characteristics

This section presents some feeder characteristics derived from the enhanced feeder model developed. Figure 5.11 shows the gross load, net-load, and PV production for the historical feeder load peak day and maximum generation condition day. For the maximum generation condition day, it is seen that the PV production significantly impacts the net

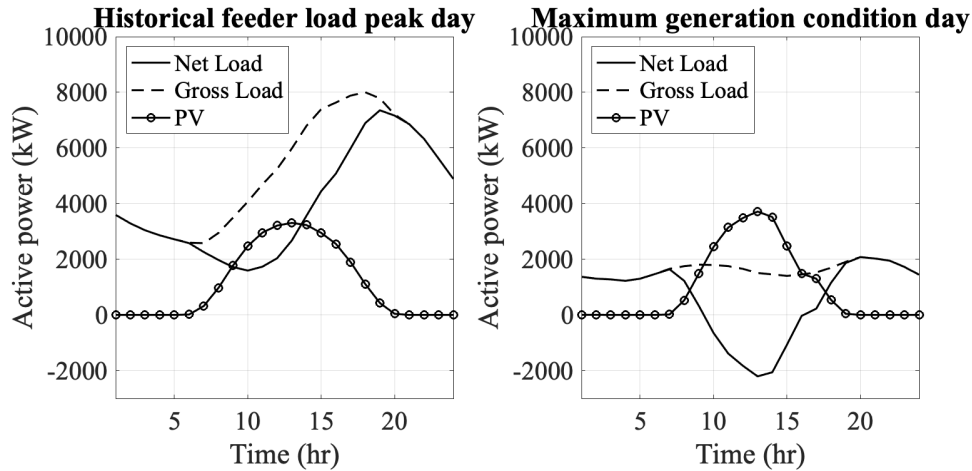


Figure 5.11. Distribution Feeder Gross Load, Net Load, and PV Production Behavior.

load of the system. In this case, due to the significant penetration of solar PV, the net load is negative between 10 AM-3 PM. Hence there is a large reverse active power flow from this feeder during this time. Due to this behavior, the feeder experiences large overvoltages during this time. The maximum solar PV generation occurs around 1 PM. The voltage profile for this time is shown in Figure 5.9, showing a trend of increasing magnitude moving away from the substation due to the reverse active power flow as well as the capacitive nature of the distribution cables. Note that the unbalance between the phases in the feeder is successfully captured by the optimization-based technique proposed.

As shown in Figure 5.11, the historical feeder load peak day has a high load, which is why there is no reverse flow at the feeder-head – there is still a significant reduction in the net load due to solar PV production. In Figure 5.8, the voltage profile of the feeder corresponding to an evening condition (high load) for the historical feeder load peak day shows a decreasing trend as we go away from the substation, which is traditionally expected for distribution systems. Figure 5.10 shows a power flow snapshot at night (2 AM) on the maximum generation day. In this case, because of the absence of solar PV

generation and the presence of light load, the voltage profile does not experience a large change as compared with the other two cases.

The contrast in the feeder voltage profiles between the snapshots presented in Figure 5.8, Figure 5.9, and Figure 5.10 as well as the gross and net loads presented in Figure 5.11 highlight the fact that a distribution system with a high solar PV penetration can exhibit a wide range of behaviors, hence accurately modeling the distribution system is important for any studies involving it.

Figure 5.12 shows the power factor for some premises along the feeder for the historical feeder load peak day. The enhancement algorithm estimates the power factor for each load independently. Figure 5.13 shows the customers' power factor obtained from the enhanced feeder model for all the loads in the feeder for both cases, the historical feeder load peak snapshot (a) and the maximum generation condition snapshot (b). The difference in the two cases is reflected in the different power factor levels – the power factors for the load peak are lower than the maximum generation condition, due to the high load demand.

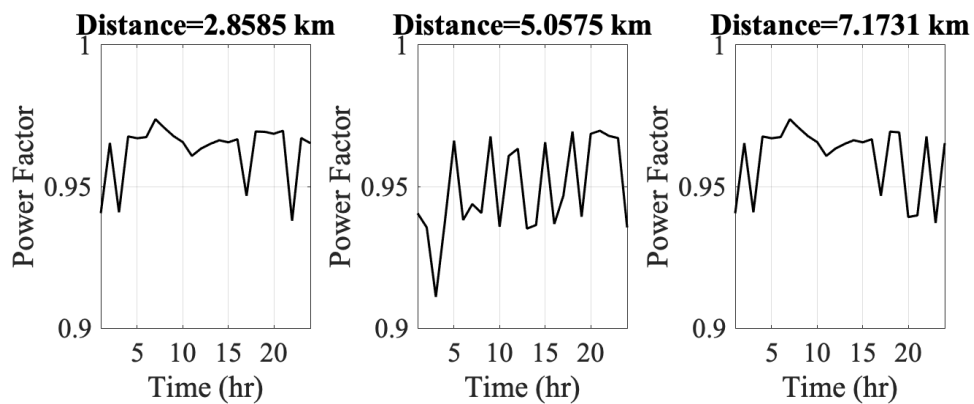


Figure 5.12. Power Factor for Some Premises Along the Feeder for the Historical Feeder Load Peak Day.

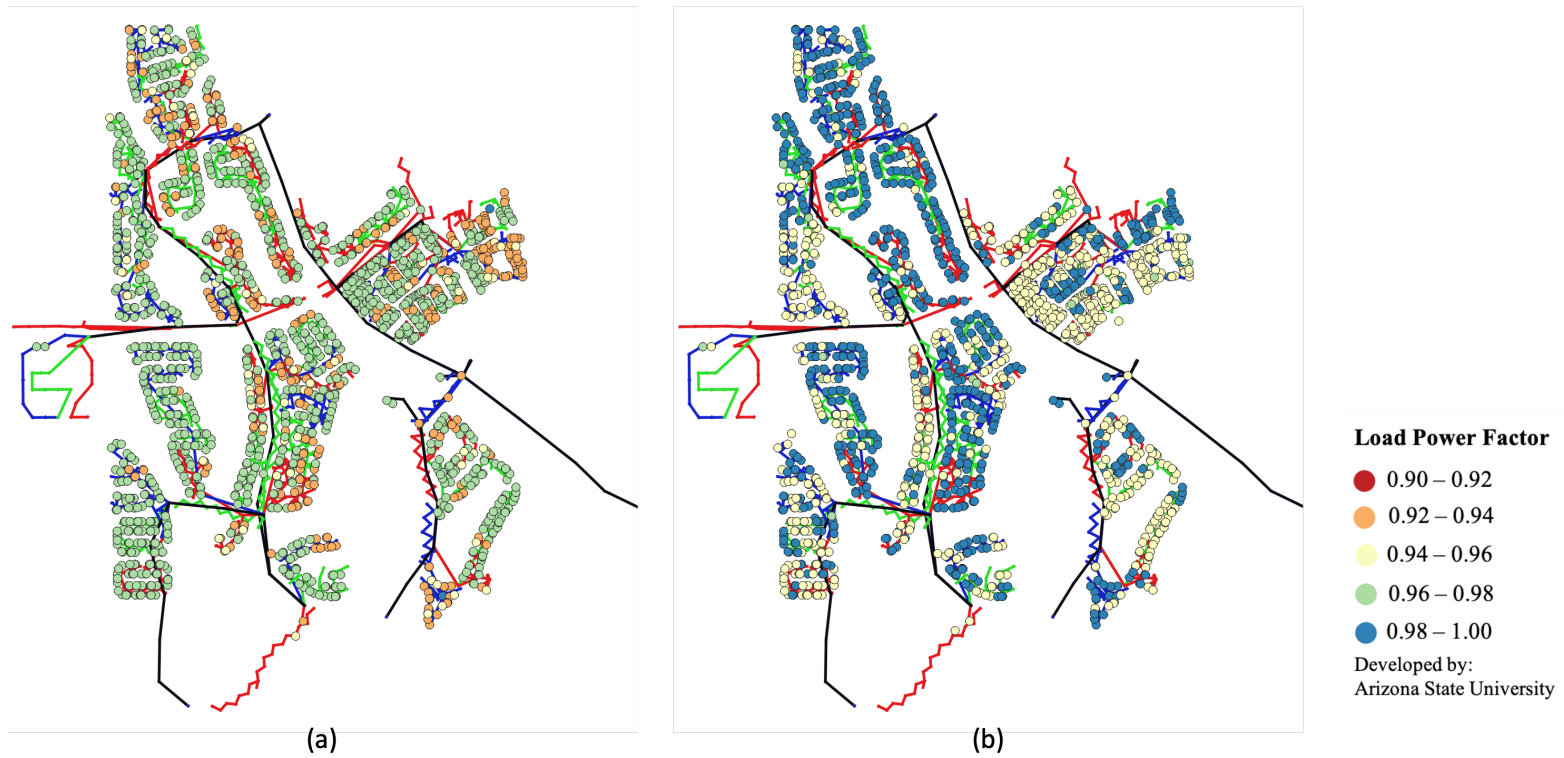


Figure 5.13. Customers' Power Factors Obtained from the Enhanced Feeder Model for All the Loads in the Feeder for Both Cases, the Historical Feeder Load Peak Snapshot (Left) and the Maximum Generation Condition Snapshot (Right).

5.4 Comparison: Enhanced Feeder Model Vs. Using Constant Power Factor for All the Loads

As a preliminary stage of the proposed method, a time series model is created by constructing the active/reactive power profiles for loads and PVs using constant power factor for all the loads in the system, as well as a balanced substation source for the feeder-head voltage profile. A flowchart explaining the creation of the power profiles is summarized in Figure 5.14. To accurately capture system behavior, the power profiles are obtained based on the information available for each of the loads/PVs. The procedure followed to calculate the profiles for all the cases present in the system is detailed below.

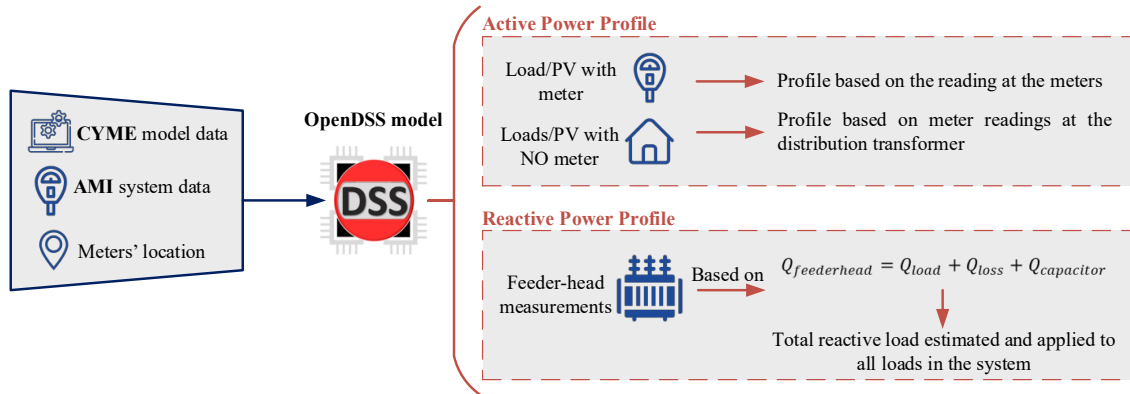


Figure 5.14. Flowchart for Time-series Profile Formulation Based on OpenDSS Model Using Constant Power Factor for All the Loads and Balanced Substation Source.

5.4.1 Active Power Profile: Base-case

To construct the base case, one hour is selected as the kW definition of the loads, then load multipliers ($dM_{l,\phi}$) are calculated based on the gross load definition for each hour using (5.3).

$$dM_{l,\phi}|_h = \frac{d_{l,\phi}^D|_h}{d_{l,\phi}^D|_{base-case}}, h \in 1, \dots, 24, \phi \in \psi(l) \quad (5.3)$$

1. *Loads with meter*: For these loads, the active power profile is based on the meters' readings. To achieve this objective, gross load calculation is carried out for each hour for all the loads in the system that have meters. Load multipliers then are calculated for the remaining hours.
2. *Loads with NO meter*: For loads with no meter, an active power profile based on the distribution transformers power calculation is used. For these cases, the total gross load is calculated at distribution transformer level for every hour using the loads that have meters connected to the distribution transformer. Subsequently, load multipliers are calculated based on the load definition at the distribution transformer level for the remaining hours. Then, the loads without meters are configured to use their specific distribution transformer profile.
3. *PV with meter*: The electrical utility did not provide the rating of the PV units, therefore the production meters value at the same time as the one used for load profiles are used for kW definitions. Then, PVs multipliers are calculated for the remaining hours.

5.4.2 Reactive Power Profile: Base-case

For reactive power, no measurements are available along the feeder at the meters. Hence, the reactive load in the system is estimated using (5.4) with the measurements from the head-of-the-feeder taken at the substation.

$$Q^H = \sum_{\forall l \in \Omega_D} Q_{l,\phi}^D + \sum_{\forall c \in \Omega_{cap}} Q_{c,\phi}^C + \sum_{\forall n \in \Omega_N} Q_{n,\phi}^{loss} \quad (5.4)$$

To calculate the reactive power of the loads ($Q_{l,\phi}^D$, $l \in \Omega_D$) using (5.4), a time series power flow is run with an approximate reactive load assuming a power factor of 0.9 lagging for all the loads. From this power flow analysis, the reactive power of the active capacitor banks ($Q_{c,\phi}^C$, $\forall c \in \Omega_{cap}$) and the reactive power losses of the feeder ($Q_{n,\phi}^{loss}$, $\forall n \in \Omega_N$) are estimated and hence a better approximation for the reactive power of the loads is computed. Iterating in this manner, the reactive power profile is calculated using $Q_{l,\phi}^D$ from (5.4) and is applied to all the loads in the system according to the total connected load at the distribution transformer each load is connected to.

5.4.3 Voltage Profile

The voltage source is located at the distribution substation (69/12.47 kV). The voltage profile is selected from the head-of-the-feeder measurements, assuming a three-phase balanced source. This assumption is made due to only having the voltage measurement on one phase from the utility. For OpenDSS, this source is modeled behind a Thévenin equivalent impedance representing the upstream network at the head-of-the-feeder, therefore the head-of-the-feeder voltage in OpenDSS is not necessarily balanced.

5.4.4 Active and Reactive Profile for Other Days

Using the base case scenario power definition, active and reactive power profiles are created for other days for which the electrical utility provided AMI data. For these days,

only AMI data and feeder-head measurements are available, i.e., there is no CYME model data (network topology information). All loads/PVs power definitions are retained from the base-case scenario.

The first step in the creation of the power profiles for other days is to compare the meters available for the new day with the meters in the base-case. For those loads/PVs that have meters on both days, the active power profiles are created based on the multiplier calculation between the load/PV definition and the meter measurements using (5.3). For the remaining loads/PVs that do not have meters in one or both days, the active profile is calculated based on the distribution transformers power calculation, similar to the base-case. The reactive power profiles are calculated similar to the base-case.

5.4.5 Accuracy Comparison with Enhanced Feeder Model

This section describes the time-series results using constant power factor for all the loads in the system and its comparison against the results of using the enhanced feeder model proposed in this dissertation. The comparison is carried out for the actual historical feeder peak day on 07/15/2019 (high load and relatively low PV). Applying the load and PV profiles accordingly created using this method and using the enhanced distribution feeder model, a comparison against the corresponding field measurements is performed. Figure 5.15 shows the active and reactive power feeder-head comparison between the feeder model using constant power factor for all the loads in the system and DAS measurements for the historical feeder load peak day.

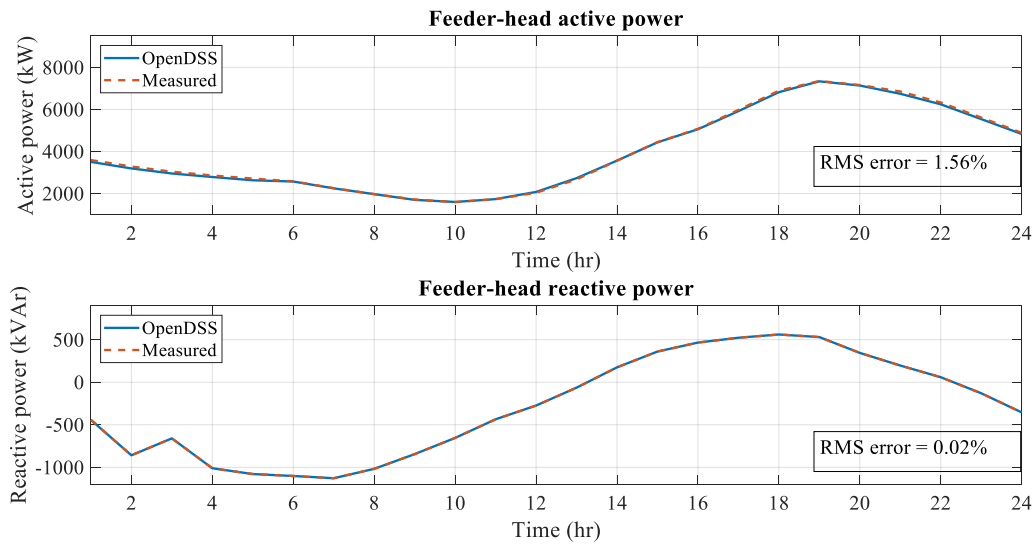


Figure 5.15. Active and Reactive Power Feeder-head Comparison Between the Feeder Model Using Constant Power Factor for All the Loads in the System and DAS Measurements for the Historical Feeder Load Peak Day.

Using constant power factor for all the loads in the system and a balanced substation source the feeder-head active and reactive powers have an RMS error over a day of 1.56% and 0.02% respectively. Note that the feeder-head reactive power measurements are used to generate the reactive power profile in this method, hence the error is expected to be quite low provided the status of the capacitor banks are being correctly modeled. On the other hand, as shown in Figure 5.3, using the enhanced feeder model, for the historical feeder load peak day, the feeder-head active and reactive powers have RMS errors over a day of 0.104% and 0.234%, respectively. Note that the reactive power along the feeder is completely calculated by the optimization-based technique as there are no reactive power measurements available at any point along the feeder aside from the feeder-head values. The result of this comparison shows that for the feeder-head both methods provide a similar

solution, however, as mentioned before, it is important to note how each method calculates these values.

The Voltage comparison results between the feeder model using constant power factor for all the loads in the system and AMI data for some premises along the feeder for the historical feeder load peak day is shown in Figure 5.16. Using constant power factor for all the loads in the system and a balanced substation source, the voltages near the beginning, middle and end of the feeder have RMS errors over a day of 0.67%, 1.04%, and 0.61%, respectively. On the other hand, using the enhanced feeder model as shown in Figure 5.5, the RMS errors over a day for the same premises are 0.18%, 0.21%, and 0.54%, respectively. Note that using the enhanced feeder model, not only does the RMS errors decrease but the wave forms match closely between the model results and the field measurements, making this method more accurate.

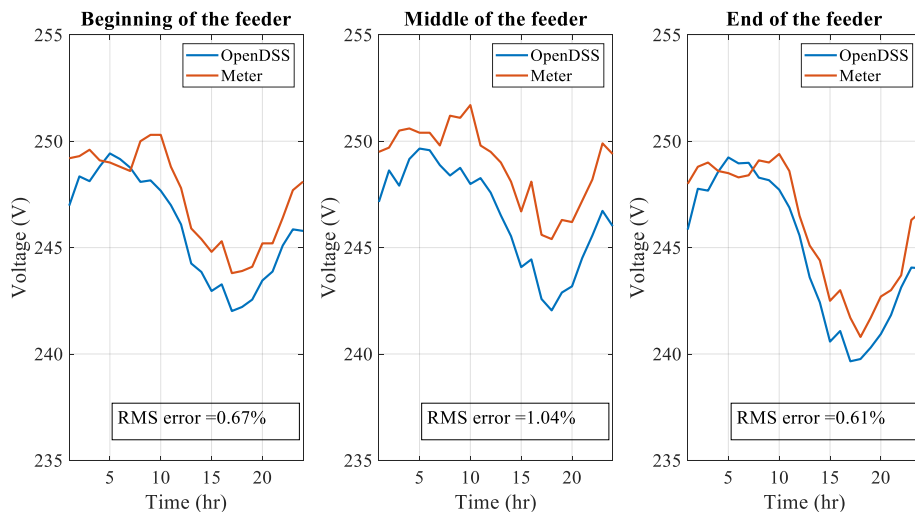


Figure 5.16. Voltage Comparison Results Between the Feeder Model Using Constant Power Factor for All the Loads in the System and AMI Data for Some Premises along the Feeder for the Historical Feeder Load Peak Day.

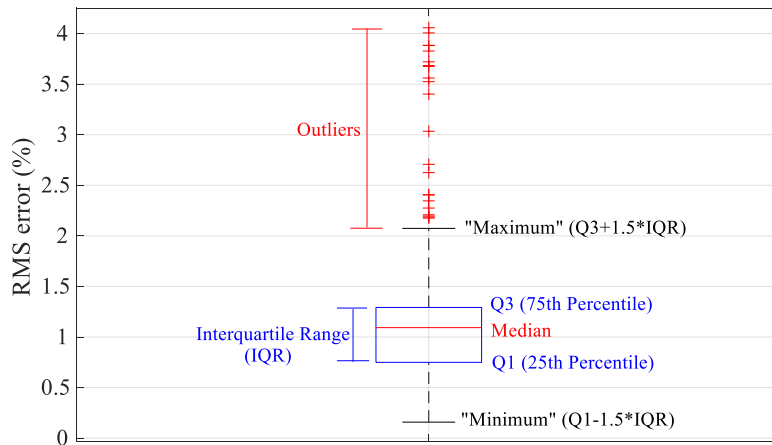


Figure 5.17. The RMS Errors Calculated over the Historical Feeder Load Peak Day Between the Field-measured Voltages and the Voltages Obtained from the Feeder Model Using Constant Power Factor for All the Loads in the System for All the Voltage Meters are Plotted Using a Box and Whisker Plot. The Red Line in the Middle in Each Case Indicates the Median Error, the Box Indicates the Interquartile Range, and the Whiskers Are Defined as 1.5 Times the Interquartile Range Away from the Box.

The RMS error over a day is also calculated for all the voltage measurements available along the feeder. Figure 5.17 shows the RMS errors calculated over the historical feeder load peak day between the field-measured voltages and the voltages obtained from the feeder model using constant power factor for all the loads in the system. Using constant power factor for all the loads in the system and a balanced substation source, the average RMS error along the feeder length is around 1.2%. On the other hand, as shown in Figure 5.7, using the enhanced feeder model, the average RMS error along the feeder length is around 0.4%.

As observed in Figure 5.15, Figure 5.16, and Figure 5.17, using the enhanced feeder model a better match is achieved when compared against field measurements. It is also observed that the enhanced feeder model accurately represents the actual field

measurements with very low RMS errors. Even though using the same power factor for all loads method gives relatively small voltage errors, this is not a correct representation of the voltages' waveform along the feeder when compared against field measurements as observed in Figure 5.16, which may result in wrong analyses of the behavior of the feeder under study. This comparison highlights the accuracy of the method proposed in this dissertation.

CHAPTER 6

AUTOMATED SECONDARY NETWORK TOPOLOGY CONSTRUCTION

As a final component of this research work, some improvements to the previous methods used were identified. For the dissertation, as an additional objective an automated tool for secondary network topology construction for correct load feeder topology assignment is developed. The automated tool would guarantee that all the premises in the feeder have at least one load defined for the power flow calculation. This is necessary to create the load input for the distribution system model enhancement algorithm already developed in Pyomo. Figure 6.1 shows how the current approach and the new approach are related to the work developed in this dissertation.

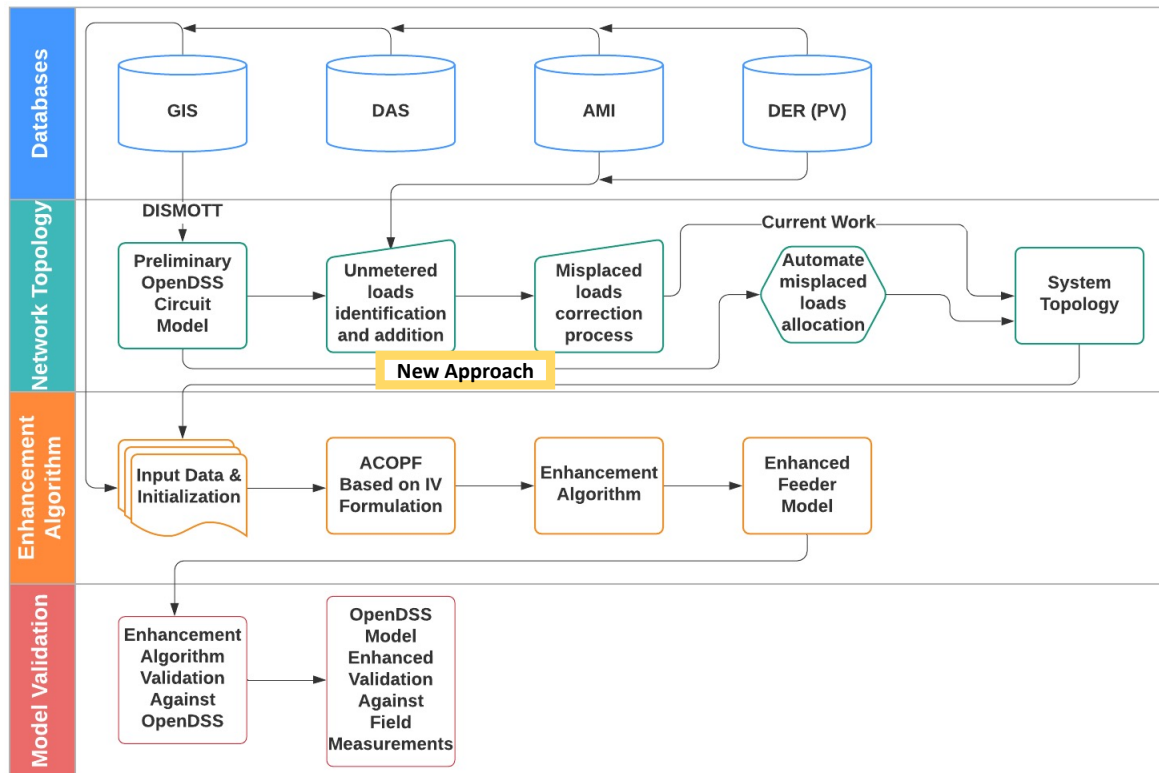


Figure 6.1. Enhanced Feeder Model Process.

As shown in Figure 6.1, the current process for the secondary network topology construction is carried out manually using the data received from the utility, following the procedure described in Section 3.3. An automated secondary network topology is proposed and implemented to provide a more accurate system topology input for the enhancement algorithm. This approach presents an automated tool for secondary network topology construction for correct load feeder topology assignment. The automated tool provides a more accurate feeder topology for power flow calculation purposes. This approach consists of three main stages: input data stage, data processing stage, and the output stage. To simplify the complexity of this new tool, only two commonly available inputs are being used as input data: parcel GIS delimitation data for the location of the feeder (obtained from a municipal lot survey information), and utility secondary feeder topology database (customers' billing information). This approach reads the raw data, then uses a DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [49], [50] clustering algorithm method to cluster the meters that correspond to the same customer to provide a single set of coordinates per customer. The procedure then uses an optimization algorithm developed in Pyomo [43], [44] to assign the loads to their corresponding customer parcel by minimizing the sum of the distances between the loads and the premises. The output of this tool is a (".csv") file with the meter locations in coordinates, and their physical address.

6.1 Automated Secondary Network Topology Construction

Figure 6.2 shows the new approach for the secondary network topology construction process. This new approach consists of three main stages: input data stage,

data processing stage carried out in three different algorithms developed in Python, and the output stage. These stages are detailed in Figure 6.3.

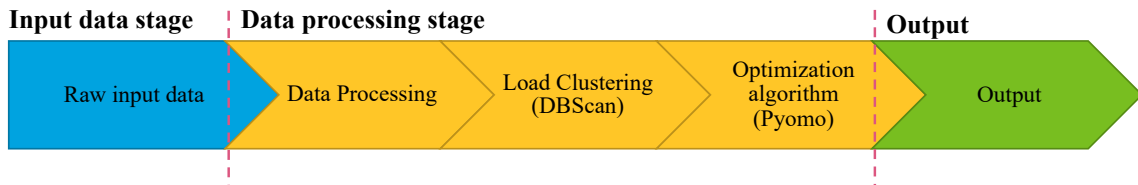


Figure 6.2. Secondary Network Topology Construction New Approach.

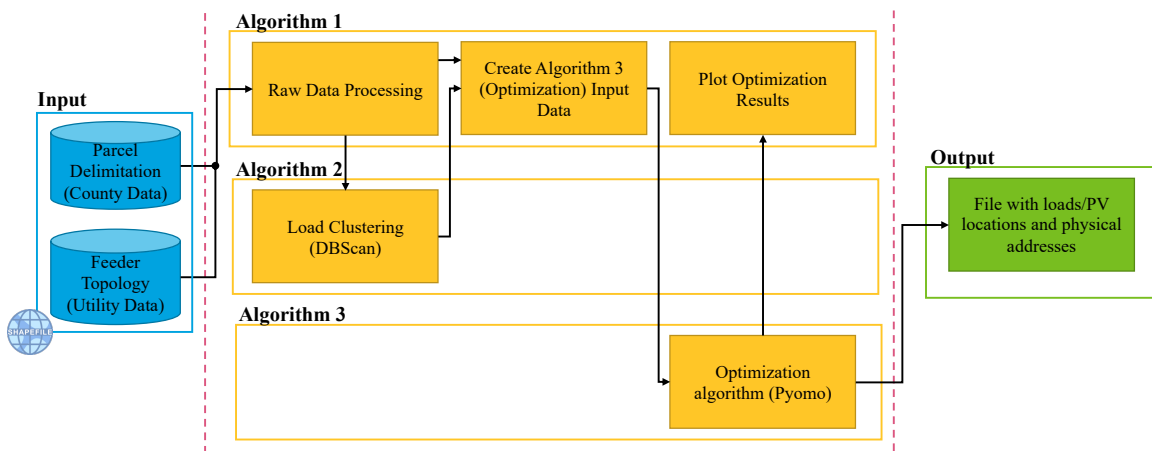


Figure 6.3. Secondary Network Topology Algorithm Detailed.

6.1.1 Input Data

To create the automated tool for load allocation, the input data should be read as raw data, with limited human intervention. To simplify the complexity of this new tool, only two commonly available inputs are being used as input data:

1. Parcel GIS delimitation data for the location of the feeder. This data is usually available for entire counties around the country, and it is accurate since it is used as geographically-reference data for tax purposes. Additionally, these databases are public, and it includes parcel attributes such as address and size.

- Utility secondary feeder topology database. Usually, utilities have databases for their customers' location for billing purposes. These databases are frequently inaccurate in terms of GIS reference (coordinates), which are used as input for the tool proposed in this work.

The input data can be available in any GIS format as long as it could be translated to a shapefile format (filename extension “.shp”, “.shx”, “.dbf”), which is a vector data format for GIS software. This format describes vector features such as points, lines, and polygons, which represents the components of the database and usually contains other attributes [51].

6.1.2 Algorithm 1: Data Processing

This algorithm consists of three main sections as shown in Figure 6.3.

- Raw Data Processing*: this section of algorithm 1 processes the raw input data to create the input for the other two algorithms. This process is shown in Figure 6.4.

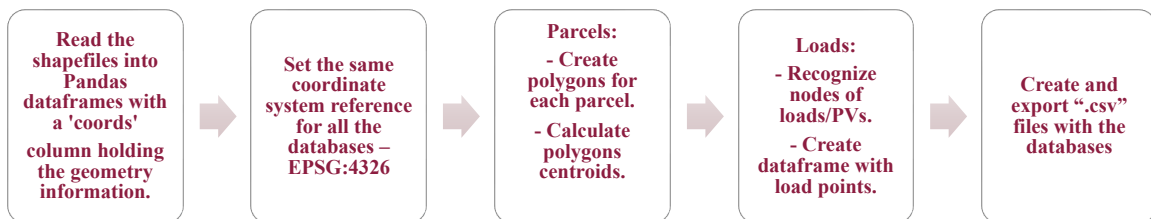


Figure 6.4. Raw Input Data Processing.

- Create Input Data for Algorithm 3 (Optimization)*: this section of algorithm 1 reads the data frames created in the raw data processing stage, as well as the load clustering algorithm output to create the input data for the optimization algorithm in (“.dat”)

format. This optimization input data consists of parcel centroids with their respective latitude and longitude information, and cluster of loads with their respective latitude and longitude data.

3. *Plot Optimization Results*: this section of algorithm 1 uses the parcel polygons created in the raw data processing stage, reads the output of the optimization algorithm automatically exported in (“.csv”) format, and plots them in a map to show the optimization results, that is, the placement of the loads in the feeder.

6.1.3 Algorithm 2: Load Clustering (DBSCAN Clustering Method)

This algorithm reads the (“.csv”) file of the loads’ latitude and longitude exported from the raw data processing stage of Algorithm 1. Since the data received from the utility corresponds to the billing information of the customers, the loads and PVs are defined in terms of meters. That is, if a customer has a PV system, two independent meters are created for this customer: one for load consumption and one for PV production. Therefore, an algorithm to cluster the meters that corresponds to the same customer should be developed to provide a single set of coordinates per customer (cluster coordinates) to be assigned to a parcel in the optimization algorithm. For the purposes of this work, this algorithm is developed using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm method [49], [50].

6.1.3.1 DBSCAN - Density-Based Spatial Clustering of Applications with Noise

The DBSCAN algorithm is a density-based clustering algorithm that discover clusters of arbitrary shape – spherical, drawn-out, linear, and other similar shapes. This is

especially useful over other clustering algorithms such as k -means which assumes that the clusters are convex shaped. To find the clusters the DBSCAN starts with an arbitrary point p and clusters all points that are density reachable from p using the parameters specified. This algorithm is efficient for large spatial databases and only requires two input parameters, $min_samples$ and eps , which define the density of the clusters. A point p is considered inside a cluster when there exist min_sample number of other points within a distance of eps , which are defined inside the same cluster as the point p [49], [50]. The parameter $min_samples$ controls the algorithm tolerance to noise by specifying the minimum number of points inside a cluster. The parameter eps controls the local neighborhood of the points and is a distance function. If this parameter is chosen too small, most data will not be clustered. If it is chosen too large, close clusters would merge into one cluster [50].

Due to the algorithm efficiency on large databases, its minimal requirements of knowledge to determine the input parameters (appropriate values are unknown in advance when dealing with large databases), and its ability to discover clusters with arbitrary shape, it is selected to cluster the customer's meters to provide a single set of coordinates per household.

6.1.3.2 Problem Formulation

The formulation for the DBSCAN clustering algorithm is as follows:

1. *Objective:* Create clusters of meters that corresponds to the same customer and calculate the cluster's centroid.

2. *Constraints:* The constraints are specific to the feeder under study and are needed to define the parameters of the algorithm. As shown in Figure 3.3, the loads and PVs are defined in terms of meters in the billing information database provided by the utility. In this case, each customer has a billing meter and there are customers with none, one or two individual PV systems. Therefore, the parameter `min_samples` is set as 1 and the maximum number of meters per customer (cluster size) is 3. Additionally, the maximum Euclidian distance between points inside a cluster is needed to set the parameter `eps`. This distance is obtained by measuring the distance between meters of the same customer.
3. *Unknown data:* Total number of clusters (customers/loads).
4. *Output:* Clusters centroid's latitude and longitude.
5. *Process:* The algorithm reads the loads' coordinates obtained from Algorithm 1 and performs a nested DBSCAN as shown in Figure 6.5 and then calculates the clusters' centroid.

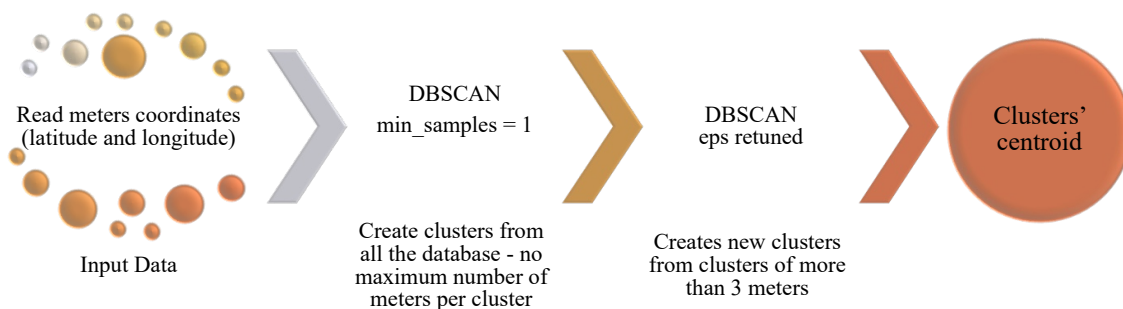


Figure 6.5. DBSCAN Algorithm.

The first DBSCAN creates clusters from the utility database using `min_samples = 1` (at least one meter per cluster). Since the DBSCAN algorithm does not allow the setting

of the maximum number of members of a cluster as a parameter, a second DBSCAN is needed for all the clusters with more than 3 elements (maximum 3 meters per customer). The second DBSCAN algorithm is performed for each cluster with more than 3 elements and for some clusters with less or equal to 3 elements where the parameter eps needed to be retuned. Each one of the second DBSCANs performed creates new clusters from the grouped meters (e.g., one cluster of 4 elements would result in two clusters, one of 3 elements and one of 1 element, or two clusters with 2 elements each). Finally, the algorithm calculates the centroid of the clusters as the geometrical center point which depends on the number of elements in the cluster. A “csv” file with the clusters, the elements of the cluster, and the coordinates of the centroid is then exported. This file is then used to create the input data for the optimization algorithm.

6.1.4 Algorithm 3: Optimization Algorithm for Load Placement

This algorithm uses the (“.dat”) file created using Algorithm 1. This file contains the parcel centroids’ latitude and longitude, and the cluster centroids’ latitude and longitude. Additionally, the user should specify as an input parameter the maximum distance (D_{max}) that a cluster/load can be from the parcels to be considered a customer load and not a streetlight or other type of load. The optimization algorithm is developed using Pyomo [43], [44] and it can be solved using Gurobi [52], CPLEX [53], or any other solver that supports convex problems.

At first, the optimization algorithm calculates the distances D between the clusters and the parcels using the coordinates as,

$$D_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}, \forall i \in \Omega_{cl}, \forall j \in \Omega_p \quad (6.1)$$

Where Ω_{cl} are the clusters, and Ω_p are the parcels from the parcel delimitation.

Using D_{max} , the input data is filtered using the following rules:

1. If a load is at a farther distance than D_{max} from all the parcels' centroid, then this load is considered a streetlight. That is,

$$\Omega_{cl}^p = \Omega_{cl} - \Omega_L \quad (6.2)$$

Where Ω_L are the streetlights, and Ω_{cl}^p are the clusters to be considered for the load placement.

2. If a parcel is at a farther distance than D_{max} from all the clusters' centroids, then this parcel is not part of the considered feeder. That is,

$$\Omega_p^{cl} = \Omega_p - \Omega_{p_{out}} \quad (6.3)$$

Where $\Omega_{p_{out}}$ are the premises not connected the feeder under consideration, and Ω_p^{cl} are the parcels to be considered for the load placement.

3. If a distance $D_{i,j}$ is lower than D_{max} for $i \in \Omega_{cl}^p$, for $j \in \Omega_p^{cl}$, then the pair $[i, j] \in \Omega_D$, where Ω_D are the possible connections between a cluster and a parcel.

Using these rules, the objective function is defined to minimize the sum of the distances between the clusters (Ω_{cl}^p) and the premises (Ω_p^{cl}), and to maximize the number of clusters ($i \in \Omega_{cl}^p$) assigned to the parcels ($j \in \Omega_p^{cl}$) for $[i, j] \in \Omega_D$. Therefore, the objective function is formulated as shown in (6.4),

$$\min \left(\sum_{i \in \Omega_{cl}^p} \sum_{j \in \Omega_p^{cl}} (X_{i,j} D_{i,j}) + \text{len}(\Omega_{cl}^p) - \sum_{i \in \Omega_{cl}^p} \sum_{j \in \Omega_p^{cl}} (X_{i,j}) \right), \forall [i, j] \in \Omega_D \quad (6.4)$$

Where $X_{i,j}$ is a binary variable that is 1 if a cluster $i \in \Omega_{cl}^p$ is assigned to a parcel $j \in \Omega_p^{cl}$.

The variable $X_{i,j}$ has the following constraints:

1. Only one parcel ($j \in \Omega_p^{cl}$) can be assigned to a selected cluster ($i \in \Omega_{cl}^p$) ($X_{i,j} = 1$) or none for a cluster that is far away ($X_{i,j} = 0$).

$$\sum_{j \in \Omega_p^{cl}} X_{i,j} \leq 1, \forall [i, j] \in \Omega_D \quad (6.5)$$

2. Only one cluster ($i \in \Omega_{cl}^p$) can be assigned to a selected parcel ($j \in \Omega_p^{cl}$) ($X_{i,j} = 1$) and zero loads for empty parcels ($X_{i,j} = 0$).

$$\sum_{i \in \Omega_{cl}^p} X_{i,j} \leq 1, \forall [i, j] \in \Omega_D \quad (6.6)$$

6.1.5 Output

As shown in Figure 6.3, the output of this tool is a (“`.csv`”) file with the meters’ (loads and PVs) location in coordinates, that is, the coordinates of the centroid of the parcel they belong to, and their physical address, which is obtained from the parcel GIS delimitation data. This data is then used to construct the secondary network topology which is used as an input for the enhancement algorithm as shown in Figure 6.1.

6.2 Results

A part of the feeder is shown as an example to demonstrate step by step how the secondary network topology construction process described in Figure 6.2 is performed. Figure 6.6 shows the raw input data after being read, processed to be set in the same coordinate system reference, and plotted by the Algorithm 1. This figure shows the parcel GIS delimitation data for the location of the feeder obtained from the county, and the utility secondary feeder topology database. Figure 6.7 shows the raw data processing result of Algorithm 1, that is, the parcel centroids, and the loads/PV nodes after being automatically recognized by the algorithm.

Figure 6.8 and Figure 6.9 show the results of the nested DBSCANs. Figure 6.8 shows the results of the first DBSCAN where some clusters are merged into one cluster (yellow circles). Then, Figure 6.9 shows how the second DBSCANs take care of these clusters and divide them into two or more clusters.

Finally, Figure 6.10 shows the optimization algorithm for load placement results. The yellow arrows connect the clusters’ centroids to the parcels according to the Algorithm 3 results.

The proposed method provides a more accurate secondary network topology after place the loads/PVs to a more precise location. Further improvements can be made to the algorithm parameters to tackle any special cases and to further improve accuracy.



Figure 6.6. Algorithm 1: Input Data.



Figure 6.7. Algorithm 1: Raw Data Processing Results of Parcel Centroids and Loads/PVs Nodes.



Figure 6.8. Algorithm 2: DBSCAN Preliminary Results.



Figure 6.9. Algorithm 2: Nested DBSCANs Results.



Figure 6.10. Algorithm 3: Optimization Algorithm for Load Placement Results.

CHAPTER 7

DISCUSSION AND CONCLUSIONS

7.1 Discussion

Using the method proposed in this work, accurate and detailed primary and secondary distribution system models can be created for distribution feeder analyses using certain field measurements. An accurate and detailed feeder model is essential to capture the behavior of distribution feeders for various studies such as snapshot/dynamic hosting capacity, snapshot/dynamic impact analysis, distribution system integration costs of renewables analysis, as well as to capture voltage and thermal violations accurately and to efficiently manage renewable energy sources with advanced Volt/VAr optimization and other distribution system automation schemes.

The proposed method assumes that the network topology, line impedances, and field measurements, i.e., AMI/DAS/SCADA measurements, are known and accurate. Using these measurements, the optimization-based method estimates the reactive power demand for all the loads in the feeder – no reactive power measurements are necessary along the feeder but can be added to the model as parameters if available, as well as the active power demands for the unmetered loads.

In the absence of load reactive power measurements, the method uses other available measurements such as active power and voltage measurements to calculate the active and reactive powers, i.e., power factor, for the loads. Since the reactive power/power factor is estimated based on the available field measurements of active power and voltage,

the resulting model is more accurate than a model constructed with the assumption of a single power factor for all the loads. However, a possible source of inaccuracies is any potential inconsistency arising due to different possible methods of measuring various quantities (e.g., instantaneous value versus moving average).

As discussed on Chapter 5, the proposed method takes ~32 seconds to enhance the feeder model for one snapshot for the selected distribution feeder on a computer with a 4-core 1.8 GHz Intel Core i7-8550U CPU and 16 GB of RAM. Hence, this method is suitable for planning studies where a detailed model of the feeder is desired. The enhanced feeder model can be updated for operational analyses conducted once every minute/few minutes. It is noteworthy to highlight the AMI measurements may update only a few times an hour - or even a few times a day. Therefore, the proposed method is suitable to continuously calculate a detailed and accurate feeder model for each system status.

7.2 Conclusions

To accurately represent the distribution systems in studies, constructing a detailed model which corresponds to the actual feeder(s) is essential. This work describes a novel procedure to enhance the model of a real distribution system feeder using AMI and DAS data. This novel enhancement algorithm formulates an ACOPF based on IV formulation for matching the voltages at various nodes to AMI measurements, ensuring that the tuned model closely reflects the real-life status of the feeder for a snapshot. The model is validated using time-series analysis in OpenDSS. Simulation results for voltages have an average RMS error along the feeder under 0.5%, and all RMS errors are under 1.4% compared to the field measurements providing confidence in the developed method. In the

future, this modeling procedure can be used to create comprehensive databases for further analysis and study of the distribution networks when no AMI data is available to construct the profiles and hence guide future extensions to the distribution systems.

Additionally, this work proposes an automated tool for secondary network topology construction for load feeder topology assignation. The automated tool provides a more accurate feeder topology for power flow calculation purposes. This new tool uses commonly available inputs: parcel GIS delimitation data for the location of the feeder (municipal lot survey information), and utility secondary feeder topology database (customers' billing information) making it easy to replicate. Further improvements can be made to the algorithm to tackle any special cases and to improve accuracy.

REFERENCES

- [1] B. Kroposki *et al.*, “Autonomous Energy Grids: Controlling the Future Grid With Large Amounts of Distributed Energy Resources,” *IEEE Power Energy Mag.*, vol. 18, no. 6, pp. 37–46, Nov. 2020, doi: 10.1109/MPE.2020.3014540.
- [2] A. Bloom *et al.*, “It’s Indisputable: Five Facts about Planning and Operating Modern Power Systems,” *IEEE Power Energy Mag.*, vol. 15, no. 6, pp. 22–30, 2017, doi: 10.1109/MPE.2017.2729079.
- [3] D. Manz, R. Walling, N. Miller, B. Larose, R. D’Aquila, and B. Daryanian, “The grid of the future: Ten trends that will shape the grid over the next decade,” *IEEE Power Energy Mag.*, vol. 12, no. 3, pp. 26–36, 2014, doi: 10.1109/MPE.2014.2301516.
- [4] M. Braun *et al.*, “The Future of Power System Restoration: Using Distributed Energy Resources as a Force to Get Back Online,” *IEEE Power Energy Mag.*, vol. 16, no. 6, pp. 30–41, 2018, doi: 10.1109/MPE.2018.2864227.
- [5] R. Quint *et al.*, “Transformation of the grid: The impact of distributed energy resources on bulk power systems,” *IEEE Power Energy Mag.*, vol. 17, no. 6, pp. 35–45, 2019, doi: 10.1109/MPE.2019.2933071.
- [6] J. McCalley *et al.*, “Wide-Area Planning of Electric Infrastructure,” *IEEE Power Energy*, no. December, pp. 83–93, 2017.
- [7] G. J. Cho, C. H. Kim, Y. S. Oh, M. S. Kim, and J. S. Kim, “Planning for the Future: Optimization-Based Distribution Planning Strategies for Integrating Distributed Energy Resources,” *IEEE Power Energy Mag.*, vol. 16, no. 6, pp. 77–87, 2018, doi: 10.1109/MPE.2018.2864228.
- [8] C. Vartanian, R. Bauer, L. Casey, C. Loutan, D. Narang, and V. Patel, “Ensuring System Reliability: Distributed Energy Resources and Bulk Power System Considerations,” *IEEE Power Energy Mag.*, vol. 16, no. 6, pp. 52–63, 2018, doi: 10.1109/MPE.2018.2863059.
- [9] NERC, “Distributed Energy Resources: Connection Modeling and Reliability Considerations,” 2016. Accessed: Sep. 10, 2019. [Online]. Available: [https://www.nerc.com/comm/Other/essntlrbltysrvdstskfrDL/DERTF Draft Report - Connection Modeling and Reliability Considerations.pdf#search=Distributed energy resources%2C connection%2C modeling%2C and reliability considerations](https://www.nerc.com/comm/Other/essntlrbltysrvdstskfrDL/DERTF%20Draft%20Report%20-%20Connection%20Modeling%20and%20Reliability%20Considerations.pdf#search=Distributed%20energy%20resources%2C%20connection%2C%20modeling%2C%20and%20reliability%20considerations).
- [10] F. Ipakchi, Ali ; Albuyeh, “Grid of the Future,” *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, 2009, doi: 10.1109/MPE.2008.931384.

- [11] D. Lew *et al.*, “The Power of Small: The Effects of Distributed Energy Resources on System Reliability,” *IEEE Power Energy Mag.*, vol. 15, no. 6, pp. 50–60, 2017, doi: 10.1109/MPE.2017.2729104.
- [12] E. Ela *et al.*, “Electricity Markets and Renewables,” *IEEE Power Energy Mag.*, vol. 15, no. december, pp. 70–82, 2017, doi: 10.1109/MPE.2017.2730827.
- [13] G. J. Shirek and B. A. Lassiter, “Photovoltaic power generation: Modeling solar plants’ load levels and their effects on the distribution system,” *IEEE Ind. Appl. Mag.*, vol. 19, no. 4, pp. 63–72, 2013, doi: 10.1109/MIAS.2012.2216000.
- [14] J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, “Distribution System Model Calibration With Big Data From AMI and PV Inverters,” *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2497–2506, Sep. 2016, doi: 10.1109/TSG.2016.2531994.
- [15] Y. F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, “State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid,” *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 33–43, 2012, doi: 10.1109/MSP.2012.2187037.
- [16] EPRI, “OpenDSS - EPRI Distribution System Simulator.” Electrical Power Research Institute (EPRI), 2019, [Online]. Available: <https://smartgrid.epri.com/SimulationTool.aspx>.
- [17] T. Gonen, *Electric Power Distribution Engineering*, Third Edit. CRC Press, 2014.
- [18] T. Gonen, *Electric Power Distribution Engineering*, vol. 3, no. 2. CRC Press, 2015.
- [19] W. H. Kersting, *Distribution System Modeling and Analysis*, Fourth Ed. CRC Press, 2017.
- [20] J. Guerrero *et al.*, “Distributed Generation: Toward a New Energy Paradigm,” *IEEE Ind. Electron. Mag.*, vol. 4, no. 1, pp. 52–64, Mar. 2010, doi: 10.1109/MIE.2010.935862.
- [21] M. Baran, “Challenges in distribution system analysis,” *1999 IEEE Power Eng. Soc. Summer Meet. PES 1999 - Conf. Proc.*, vol. 2, p. 627, 1999, doi: 10.1109/PESS.1999.787387.
- [22] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, “A Time-Series Distribution Test System Based on Real Utility Data,” *51st North Am. Power Symp. NAPS 2019*, 2019, doi: 10.1109/NAPS46351.2019.8999982.

- [23] Y. Tang and R. Ayyanar, "Modeling and validation of a distribution system with high PV penetration using zone division method," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, pp. 0–4, 2014, doi: 10.1109/tdc.2014.6863357.
- [24] T. J. Morrell, V. Venkataramanan, A. Srivastava, A. Bose, and C.-C. Liu, "Modeling of Electric Distribution Feeder Using Smart Meter Data," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, Apr. 2018, vol. 2018-April, pp. 1–9, doi: 10.1109/TDC.2018.8440540.
- [25] J. Giraldez, P. Gotseff, A. Nagarajan, R. Ueda, J. Shindo, and S. Suryanarayanan, "Distribution Feeder Modeling for Time-Series Simulation of Voltage Management Strategies," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, Apr., pp. 1–5, doi: 10.1109/TDC.2018.8440379.
- [26] J. A. Taylor, T. A. Short, and B. Bushey, "Efficiency impacts of distribution secondaries," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, pp. 1–6, 2012, doi: 10.1109/TDC.2012.6281661.
- [27] J. R. Agüero and S. J. Steffel, "Integration challenges of photovoltaic distributed generation on power distribution systems," *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–6, 2011, doi: 10.1109/PES.2011.6039097.
- [28] G. J. Shirek, B. A. Lassiter, W. C. Carr, and W. H. Kersting, "Modeling secondary services in engineering and mapping," *IEEE Trans. Ind. Appl.*, vol. 48, no. 1, pp. 254–262, 2012, doi: 10.1109/TIA.2011.2175874.
- [29] M. J. Reno and R. J. Broderick, "Statistical analysis of feeder and locational PV hosting capacity for 216 feeders," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2016-Novem, 2016, doi: 10.1109/PESGM.2016.7741269.
- [30] T. J. Morrell, V. Venkataramanan, A. Srivastava, A. Bose, and C. C. Liu, "Modeling of Electric Distribution Feeder Using Smart Meter Data," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, vol. 2018-April, 2018, doi: 10.1109/TDC.2018.8440540.
- [31] P. Balakrishna, K. Rajagopal, and K. S. Swarup, "AMI/GIS based distribution system load flow for extended situational awareness," *2014 18th Natl. Power Syst. Conf. NPSC 2014*, pp. 1–6, 2015, doi: 10.1109/NPSC.2014.7103790.
- [32] J. Giraldez, P. Gotseff, A. Nagarajan, R. Ueda, J. Shindo, and S. Suryanarayanan, "Distribution Feeder Modeling for Time-Series Simulation of Voltage Management Strategies," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, vol. 2018-April, pp. 2–6, 2018, doi: 10.1109/TDC.2018.8440379.

- [33] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, and R. G. Harley, "Leveraging AMI Data for Distribution System Model Calibration and Situational Awareness," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2050–2059, Jul. 2015, doi: 10.1109/TSG.2014.2385636.
- [34] Y. Tang, X. Mao, and R. Ayyanar, "Distribution system modeling using CYMDIST for study of high penetration of distributed solar photovoltaics," *2012 North Am. Power Symp. NAPS 2012*, 2012, doi: 10.1109/NAPS.2012.6336408.
- [35] Y. Tang and R. Ayyanar, "Methodology of Automated Protection Analysis for Large Distribution Feeders With High Penetration of Photovoltaic Systems," *IEEE Power Energy Technol. Syst. J.*, vol. 4, no. 1, pp. 1–9, Mar. 2017, doi: 10.1109/JPETS.2016.2578942.
- [36] B. Wang and W. Luan, "Generate distribution circuit model using AMI data," *China Int. Conf. Electr. Distrib. CICED*, vol. 2014-Decem, no. Ciced, pp. 1251–1255, 2014, doi: 10.1109/CICED.2014.6991907.
- [37] Z. Soltani, S. Ma, M. Khorsand, and V. Vittal, "Simultaneous Robust State Estimation, Topology Error Processing, and Outage Detection for Unbalanced Distribution Systems," May 2021, [Online]. Available: <http://arxiv.org/abs/2105.10111>.
- [38] EATON, "CYME power engineering software for a smarter grid," *IEEE power and energy for electric power professionals*, vol. 15, no. 6. 2017, [Online]. Available: <https://www.cyme.com/software/>.
- [39] K. Montano-martinez, S. Thakar, V. Vittal, R. Ayyanar, and C. Rojas, "Detailed Primary and Secondary Distribution System Feeder Modeling Based on AMI Data," *Proceedings NAPS 2020 - 52th North American Power Symposium*. Tempe, AZ, USA, 2021.
- [40] K. Montano and S. Thakar, "DISMOTT: DISTRIBUTION SYSTEM MODEL TRANSFORMATION TOOL," 2020. <https://github.com/thakars/DISMOTT>.
- [41] QGIS Development Team, "QGIS Geographic Information System." Open Source Geospatial Foundation, 2020.
- [42] K. Montano-Martinez *et al.*, "Detailed Primary and Secondary Distribution System Model Enhancement Using AMI Data," *IEEE Open Access J. Power Energy*, vol. 9, pp. 2–15, 2022, doi: 10.1109/OAJPE.2021.3125900.
- [43] W. E. Hart *et al.*, *Pyomo — Optimization Modeling in Python*, vol. 67. Cham: Springer International Publishing, 2017.

- [44] W. E. Hart, J.-P. Watson, and D. L. Woodruff, “Pyomo: modeling and solving mathematical programs in Python,” *Math. Program. Comput.*, vol. 3, no. 3, pp. 219–260, Sep. 2011, doi: 10.1007/s12532-011-0026-8.
- [45] A. Wächter and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” *Math. Program.*, vol. 106, no. 1, pp. 25–57, Mar. 2006, doi: 10.1007/s10107-004-0559-y.
- [46] R. P. O’Neill, A. Castillo, and M. B. Cain, “The IV formulation and linear approximations of the AC optimal power flow problem (OPF Paper 2),” *FERC Staff Tech. Pap.*, no. December, pp. 1–18, 2012, [Online]. Available: <http://www.ferc.gov/industries/electric/indus-act/market-planning/opf-papers/acopf-2-iv-linearization.pdf>.
- [47] D. K. Molzahn *et al.*, “A Survey of Distributed Optimization and Control Algorithms for Electric Power Systems,” *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017, doi: 10.1109/TSG.2017.2720471.
- [48] T. A. Short, “Advanced metering for phase identification, transformer identification, and secondary modeling,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 651–658, 2013, doi: 10.1109/TSG.2012.2219081.
- [49] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, and others, “A density-based algorithm for discovering clusters in large spatial databases with noise.” in *kdd*, 1996, vol. 96, no. 34, pp. 226–231.
- [50] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in {P}ython,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [51] A. Esri and W. Paper, “ESRI & StatSci statistical software integrated with ARC/INFO GIS,” *Comput. Stat. Data Anal.*, vol. 16, no. 3, pp. 370–371, 1993, doi: 10.1016/0167-9473(93)90138-j.
- [52] Gurobi Optimization, LLC, “Gurobi Optimizer Reference Manual.” 2022, [Online]. Available: <https://www.gurobi.com>.
- [53] I. B. M. I. Cplex, “V12. 1: User’s Manual for CPLEX,” *Int. Bus. Mach. Corp.*, vol. 46, no. 53, p. 157, 2009.