ABSTRACT

CRAWFORD, KEN D. Methods for Topology Monitoring on a Distribution System Using State Estimation. (Under the direction of Dr. Mesut Baran and Dr. Wenyuan Tang.)

With increasing focus moving towards distributed energy resource (DER) deployment to electric power distribution systems and focus on development of smart grids, effective real-time monitoring is crucial to enhancing the reliability and efficiency of distribution system operation and control. This dissertation focuses on real-time monitoring of distribution systems to improve this reliable operation and control. Such operations that are being deployed by utilities include Smart Grid applications like Fault Location, Isolation, and Service Restoration (FLISR) and Integrated Volt-VAR Control (IVVC). These schemes require accurate situational awareness of the system operation point including system voltages. Accurate real-time monitoring can be accomplished through metering connected through a Supervisory Control and Data Acquisition (SCADA) communication infrastructure, however can be quite costly to utilities. Thus, a foundational alternative method, like Distribution System State Estimation (DSSE), is a requirement that allows for the estimation of system voltages from provided SCADA measurements and system modeling. To this end, DSSE has been deployed to determine the operating conditions of distributions feeders. However, distribution systems have unique characteristics that make the application of state estimation more challenging compared to the transmission system. These challenges include the radial topology, three-phase unbalanced system, high resistance to reactance (R/X) ratio, limited and sparse real-time measurements, network model uncertainty, and the "always changing" topology. Therefore, modifications must be made to standard state estimation methods for DSSE. One of the most challenging issues facing accurate real-time monitoring with DSSE is topology errors because DSSE assumes a correct system topology and the presence of topology errors creates significant limitations in voltage estimate accuracy. Therefore the development of methods for topology error monitoring to address these limitations is the focus of this dissertation.

One of the more common and computationally efficient methods of DSSE is the branch current based distribution system state estimation (C-DSSE), which converts actual measurement to its equivalent current measurement form. Commonly occurring topology errors that significantly impact C-DSSE are addressed and the most challenging ones are identified. These most challenging

topology errors are hardly addressed in literature and the performance of C-DSSE and traditional bad data detection (BDD)-based methods in the presence of such challenging topology errors is assessed. Through assessment, key limitations are highlighted and used as motivation for further improvement of topology monitoring methods.

In addressing the limitations on performance of C-DSSE and traditional BDD methods, this research proposes a new approach of modeling measurement uncertainty. This improvement noticeably enhances the overall quality and accuracy of state estimates from C-DSSE and significantly strengthens traditional BDD-based methods capability for topology error monitoring. With this enhanced capability, an effective signature-based analysis is proposed and is unprecedented in existing literature. A final method is proposed taking advantage of this signature-based analysis and is aimed at accurately detecting the presence and type of topology error, localizing the error to a smaller surface area of the system, and finally identifying the source of the detected topology error. This is a significant advancement in using DSSE and BDD based method for topology error monitoring for reliable real-time monitoring of distribution systems.

This challenge is quite common in practice, even across the variety of similar real-time monitoring method like power flow-based state estimation. By employing effective data analytics tools and the proposed signature-based analysis, practical application can be extended to these various real-time monitoring schemes. Extensive analysis is presented utilizing real distribution network data thereby proving the applicability of the analysis in practice.

Another significant area of this thesis is the extension of such signature-based analysis to include alternative sources of topology errors. Potential cyber-attacks against power systems pose a remarkable threat to the system topology and system operation. Recognizing this, a likely attack model threatening system topology is identified and the signature-based analysis is extended to consider such a threat. This effort supports the robustness of the proposed approach and defends its use for topology error monitoring in DSSE.

In conclusion, this thesis provides a detailed evaluation of topology error threats and the impacts and limitations towards DSSE and BDD accuracy are addressed. As a result this thesis introduces innovative approaches to enhancing the capability of methods for topology error monitoring in practical applications, thus offering substantial improvements over traditional techniques and is a significant contribution to the field of distribution system operation and control.

Methods for Topology Monitoring on a Distribution System
Using State Estimation

by
Ken D. Crawford

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina

2024

APPROVED BY:

Ning Lu	Brad Reaves
Mesut Baran Co-chair of Advisory Committee	Wenyuan Tang Co-chair of Advisory Committee

DEDICATION

To my wonderful wife and better half, for your unconditional love and support that has always been my greatest comfort and my greatest motivation through this journey. You are my rock and this accomplishment is not just mine, but also yours.

To our wonderful sons, Koen, Kerrigan, and Kallahan, for the light you have brought to our lives. I strive everyday to be my best self for you.

To my lovely mom, for your infinite love and sacrifices to provide me the best opportunities and constant reassurement to achieve my dreams. I attribute my strength and perseverance to you.

And to my siblings and built-in best friends, Julia, Hailey, and Chase, for your enduring friendship and endless support. I know in times of struggle, you three will always be by my side.

BIOGRAPHY

Ken Crawford was born in Flemington, New Jersey. He received the Bachelor of Science in Engineering degree from the University of South Carolina in Columbia, SC in 2019 majoring in electrical engineering with a minor in computer science. Later in 2019, he began his graduate studies to pursue a Ph.D. in electrical engineering from North Carolina State University in the Future Renewable Electric Energy Deliver and Management (FREEDM) Systems Center, under the guidance of Dr. Mesut Baran. Along the way in the fall of 2020, he received his Master of Science degree also in electrical engineering. He will receive his Ph.D. degree from NC State University in fall of 2024. His research interests include real-time monitoring of distribution systems and cyber-security of power systems.

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank my advisor and mentor, Dr. Mesut Baran, for his support and guidance throughout my research tenure. I am very fortunate to have an advisor and mentor with such care for his students and such knowledge in our field. I am beyond grateful for the skills and expertise that he has instilled in me that is necessary for effective and impactful research. Whatever I am today both professionally and technically, I owe in a very large part to his unwavering support and mentorship.

I would also like to express my gratitude to my committee co-chair, Dr. Wenyuan Tang, committee members, Dr. Ning Lu and Dr. Brad Reaves, for serving on my doctoral dissertation committee. Their insights and expertise significantly improved the quality of this thesis.

In a period of time that was so unsure and required isolation from my peers, I am beyond grateful for my friends and colleagues at FREEDM Systems center, Dr. Gokhan Cakir, Dr. Keith DSouza, Dr. Valliappan Muthukaruppan, Dr. Victor Paduani, and Priya Raghuraman. I value their support and friendship in this challenging time more than words can express.

Additionally, I am grateful for the staff members of FREEDM Systems Center, Karen Autry, Ken Dulaney, Rebecca McLennan, and Hulgize Kassa. Though my time in FREEDM was limited, their constant source of enthusiasm, positivity, and support never failed to always brightened our day and made FREEDM feel like home.

I would also like to express my gratitude to UNC Charlotte faculty, Dr. Valentina Cecchi and Dr. Badrul Chowdhury for their guidance and expertise on project that we collaborated on as well as with our Duke Energy sponsors and collaborators, Cara DeCoste Chacko, Steve Whisenant, Mike Simms, and Jordan Tynan, for their guidance and support.

Finally, I would like to express my deepest gratitude to my my wife, our sons, my mom, my siblings, and all of my relatives and friends for their encouragement, laughter, support, and love. It is commonly said that no person can thrive as an island, that its the village they surround themselves with that support this, and I believe I have the greatest village possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

The need for real-time monitoring methods has become more crucial over recent years particularly because of the growing impact of distributed energy resources (DERs) and increased focus on microgrids. The stochasticity and uncertainty surrounding DER generation creates the need for a utility's centralized control center to monitor its many widespread distribution systems that it provides electricity to more carefully, as an accurate awareness is needed to reliably manage and control the distribution system. Placing meters throughout the system connected through an effective communication structure through Supervisory Control and Data Acquisition (SCADA) infrastructure can accomplish this, however can become quite costly, so it becomes infeasible to place meters at every point in every distribution network. Instead, common methods for real-time monitoring in power systems that are utilized are power flow and state estimation (SE), more precisely weighted least square (WLS) based SE, where utilities only need to place a minimal number of measurements at optimal locations to estimate the operating point [1]. These methods allow for system operators to have a real-time awareness of the condition of the network without having to place meters at every point in the system.

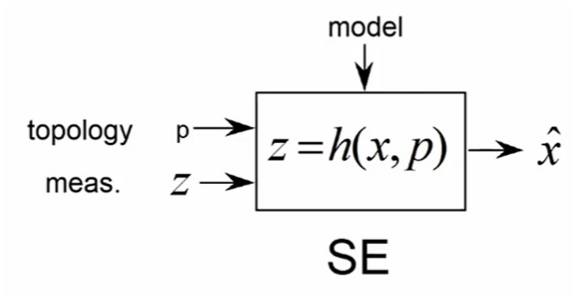

Figure 1.1 State Estimation Block Diagram

However, there are challenges associated with the nature and behavior of distribution systems that arise when applying these methods that must be addressed for accurate real-time monitoring. As seen in Fig. 1.1, WLS SE takes into consideration the system model, the system topology, and measurements, which all have unique characteristics in the distribution system.

1.1 Distribution Systems Challenges

Distribution systems, when compared to transmission systems, are more sparsely measured and radial in nature. Real-time monitoring methods that are applied to the distribution system must then be modified because of the requirement for some level of measurement redundancy to achieve an observable system. The loads are used in the method as pseudo-measurements and are derived through load estimation techniques or load allocation methods using the sparse measurements and customer information like billing history and connected capacity [2, 3].

Distribution systems are also more unbalanced in nature given loads can be single and two phase, so traditional WLS SE can experience issues with convergence when applied to these distribution systems [4]. Thus, an existing distribution system state estimation (DSSE) method has been developed based on the use of branch current values to address these challenges, which will be referred to as *C-DSSE* [5].

Another key limitation in practice is the accuracy of the estimation is reliant on the accuracy of the sparsely placed measurements. In distribution systems, there are very few real-time measurements and over time, meter accuracy can begin to degrade and become more inaccurate. Thus, a method of bad data detection (BDD) is utilized at completion of DSSE to detect these inaccuracies under the assumption that typical measurement error is within a Gaussian normal distribution [1]. Certain statistical tests like chi-squared test and largest normalized residual test are used for detection and identification of this standard bad data.

Another key challenge of DSSE in practice that impacts the performance of DSSE and BDD is the behavior of the radial topology of distribution systems. To maintain system reliability, distribution networks experience an "always changing" topology behavior. Some of these changes could be conducted with control center intervention, however most often these changes occur under automation through Distribution Automation Systems (DAS) to allow for efficient operation. Additionally equipment, such as capacitor banks, can degrade over time and fail, causing the device to be disconnected and inoperable. These automated changes and equipment failures are not always monitored or known, which creates issues for the accuracy of the DSSE results. DSSE must assume a correct model of the network, thus unknown changes in the system topology, called *topology errors*, will limit the ability of BDD to detect true measurement errors and ultimately negatively impact the accuracy of the DSSE results [4].

The focus of this work is on the challenge of topology error detection because keeping track of topology errors on a large distribution system is one of the most challenging tasks in practice. Some commonly occurring unmonitored topology errors will be discussed next.

1.1.1 Potential Topology Error Threats

1. **Network Reconfiguration**: In response to faults or line outages, switching devices are operated to reconfigure the network to maintain reliability while isolating the fault or outage. Often these devices are configured for automated action, but are also often monitored by SCADA [6].

2. **Fuse Operation**: Fuses are typically placed on laterals off of the main feeder to isolate any faults on the lateral and protect the overall reliability of the feeder. As a mechanical device, fuses are not monitored by SCADA, but these events are easily discovered by utilities when customers call to report an outage [7].

3. **Shunt Capacitor Bank Failure**: Shunt capacitor banks degrade over time. Typically utilities have scheduled maintenance when a capacitor bank is predicted to fail. However, some internal faults and failures can occur, causing the capacitor bank to be disconnected from the network [8, 9]. Capacitor bank switches are monitored by SCADA, however, if there is a failure, the switch will still appear closed but the capacitor bank will be essentially disconnected causing a change that is not modeled in DSSE.

4. **Load Phase Switching**: Load phase switching is often used for load balancing and is conducted to eliminate phase imbalance in a distribution system. This assists to eliminate the neutral current that is created by the imbalance. This is an alternative to redistributing branches through network reconfiguration [10], and can be done by automatically changing the phase of a single-phase customer transformer or single phase branch to obtain balance on across the three phases [11]. If these automatic changes are not monitored then this will create a topology change that is not represented in the model used in DSSE.

5. **Cyber Attacks**: We have seen the capability that cyber actors have in impacting a system topology, the largest coming in 2015 in Ukraine, when malicious actors took control of a utility's SCADA and opened breakers within 30 different distribution networks causing more than 200,000 customers to lose power [12]. This event was a more extreme case and was instantly

known of its occurrence, however we also see in literature the capability to stealthily impact SE [13], so there is potential that an adversary is capable of stealthily impacting measurement data and system topology [14]. For instance, it is possible adversaries can impact operations like load balancing and voltage regulation control [15, 16].

Ultimately, capacitor bank failures and load balancing are both the most common unmonitored topology changes that will impact the performance of DSSE and BDD and have not been approached yet, thus is the primary focus in this work with topology error detection in DSSE. We will also focus on a specific cyber threat that is meant to cause permanent damage to equipment that targets malicious modification of voltage regulator control.

1.2 Scope of Dissertation

It is shown that traditional WLS SE and BDD can be used to detect topology errors [1]. Recent research has also been applied to meter placement methods for improvement of performance in state estimation [17, 18]. However, focus on the more challenging and commonly occurring topology errors, such as capacitor failures and load transfers, is not found in literature. Thus, it becomes critical to assess the capability of using traditional BDD methods to monitor for these challenging topology errors. As a result, chapter 2 will take us through an assessment of the impact that two of the most common topology errors have on C-DSSE and will assess the minimum meter placement from common placement schemes needed for topology error detection. In this chapter we propose a new approach to using traditional BDD methods for topology monitoring [19]. Test results on a sample feeder are provided to illustrate the effectiveness of the proposed approach for detection of topology errors. However, limitations with the proposed approach in identification of the topology error source are examined and are addressed in the next chapter.

In traditional uses of C-DSSE and BDD, it is common to approximate the measurement variances that are used to weigh measurement residuals and used to obtain residual variances that are used for normalized residuals [20–22]. Improving such modeling of measurement variance/uncertainty can be shown to improve the overall quality and accuracy in the state estimates and an improved approach for this modeling is proposed in chapter 3 [23]. As a result, capability of BDD methods for topology error monitoring will improve. With improved BDD results a method is further proposed in chapter 3 for accurate topology error detection and identification. The method is broken into three steps: 1) Detection of presence and type of topology error using machine learning-based method, 2) Localizing the detected topology error to a smaller section of the system based on signature-based analysis, and 3) Identification of topology error through a modified search method. Case studies on test feeders that are based on real system data are provided to illustrate the effectiveness of the proposed method.

In practice, alternative real-time monitoring methods are used by certain utilities. Thus, it is important to develop a practical application for these alternative methods. Given the limitations of topology errors on the performance of state estimation on distribution systems, there are times when the specific parameters that impact the performance are not easily identifiable in practice. In this work, we focus on voltage mismatch as the performance metric. Certain data analysis tools can be employed to determine key differences in system parameters that could correlate to poorly performing cases that experience a large voltage mismatch on a consistent basis. These key differences can assist in a general suspicion of potential causes of large voltage mismatch, however, further signature-based analysis developed previous can be adopted to determine more specifically the cause for such poor performance. Chapter 4 addresses using data analysis tools, like clustering, to identify possible correlating parameters that could help explain the issues in performance on real distribution networks. An effort is also made in chapter 4 to adopt the signature-based analysis and propose guidelines to analyze signatures to help pinpoint the most likely cause for large voltage mismatches. The proposed analysis is performed using real distribution networks.

As mentioned, DSSE can also be negatively impacted resulting from malicious cyber interaction through cyber attacks targeting system topology. A malicious actor can stealthily exploit vulnerabilities in SCADA and either directly impact system topology or indirectly through misleading of automation control or control operators. One target in distribution systems is voltage regulators. Voltage regulator tap positions can be maliciously changed to cause overvoltage that can permanently damage equipment and cause outages. As a result, it must be correctly monitored. Chapter 5 will introduce this practical threat and propose an extension strategy to the method discussed in chapter 3 for analysis to improve the performance of detection of these attacks.

Ultimately, the main objective of this work is to propose methods for better real-time monitoring of distribution systems by improving topology error processing and address the following challenges:

1. Assessing the limitations of traditional methods of topology error processing using DSSE.
2. Developing methodologies for accurately processing of challenging and commonly occurring topology errors.
3. Developing a practical application for alternative real-time monitoring methods.
4. Developing a method for accurately processing cyber-attacks that impact system topology.

CHAPTER

2

ASSESSMENT OF DISTRIBUTION SYSTEM STATE ESTIMATION PERFORMANCE WITH PRESENCE OF TOPOLOGY ERRORS

As mentioned, distribution system state estimation (DSSE) performance is severely limited when the system topology is inaccurate. Previous works have considered topology error processing as part of the traditional bad data detection (BDD) process [1, 17, 18]. However, these works do not consider the most challenging common topology errors like capacitor bank failure and load transfers. Thus, it is critical to assess the capability of these traditional methods of BDD to be used for topology error processing in the presence of these topology errors.

In this chapter, the performance of BDD in the presence of topology errors is assessed. We first look at the development of C-DSSE and traditional BDD. In this chapter, we also propose an alternative approach to traditional BDD methods that is meant to improve the capability for topology error monitoring. Then several common measurement placement schemes are introduced to assess the minimum placement required to improve the topology monitoring scheme.

We evaluate the challenges and limitations of performance of BDD and the ability to detect the presence of topology errors through test results on a sample test feeder.

2.1 Current-Based Distribution System State Estimation

As previously stated, distribution systems are characterized by sparsely measured, unbalanced three-phase, radial networks, so basic WLS SE [1] must be extended to capture the unbalanced nature while remaining computationally efficient. Current-based methods are considered for this work [5].

In general, the objective of WLS methods seeks to reduce a weighted sum of square residuals between the m measurements and calculated values derived from the n system states. This objective function is defined by:

$$min f_0 = \sum_{i=1}^{m} W_{ii}(z_i - h_i(\mathbf{x}))^2 \tag{2.1}$$

where, W_{ii} the i^{th} term of the diagonal weighted matrix, $\mathbf{W}$, z_i is the i^{th} measurement value, and $h_i(\mathbf{x})$ is the calculated measurement function of the i^{th} measurement given state estimate vector, $\mathbf{x}$, of size $n x 1$. Often $\mathbf{x}$ contains node voltages and phase angles. The weight matrix, $\mathbf{W}$, is derived from measurement variances and defined as:

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_m^2} \end{bmatrix} \tag{2.2}$$

where σ_m^2 is the measurement variance of the m^{th} measurement.

Thus, the WLS approach is an optimization problem that can be solved using normal equations and iterative methods by adjusting the terms of **x** and converges within some pre-determined tolerance. The normal equation and update equation are defined below:

$$\Delta\mathbf{x}^k = \mathbf{G}^{-1}\mathbf{H}^T\mathbf{W}[\mathbf{z} - \mathbf{h}(\mathbf{x}^k)] \tag{2.3}$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k \tag{2.4}$$

where **H** is the measurement Jacobian relating the system states to the measurements and **G** is the gain matrix and is defined as, $\mathbf{G} = \mathbf{H}^T\mathbf{W}\mathbf{H}$.

To capture the unbalanced behavior of a distribution, the three-phase line impedance model is used. A diagram of a three phase feeder section is shown in Fig. 2.1 below:

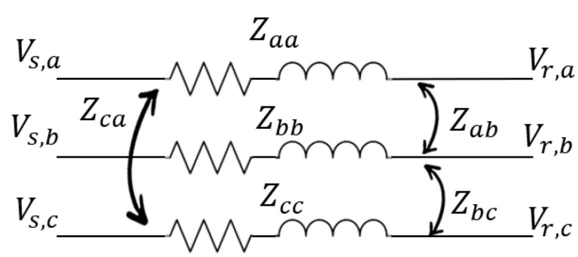

Figure 2.1 Three Phase Line Impedances

From this the 3x3 impedance matrix is defined as:

$$\mathbf{Z}_{ij} = \begin{bmatrix} Z_{aa} & Z_{ab} & Z_{ac} \\ Z_{ba} & Z_{bb} & Z_{bc} \\ Z_{ca} & Z_{cb} & Z_{cc} \end{bmatrix} \tag{2.5}$$

where the diagonal terms are the self impedances of each phase and off-diagonal terms are the mutual impedances between the phases of the line section. Given this impedance matrix, the three-phase admittance matrix is defined as the inverse of the impedance matrix, $\mathbf{Y}_{ij} = (\mathbf{Z}_{ij})^{-1}$. Then, the bus admittance matrix, $\mathbf{Y}_{bus}$, of the three-phase system is modeled by:

- Diagonal Terms: sum of $\mathbf{Y}_{ij}$ connected to the node
- Off-Diagonal Terms: $-\mathbf{Y}_{ij}$ between the two nodes

Finally, to improve the computational efficiency of this method, the various input measurements are converted into equivalent current measurements creating a current-based approach. For example, the equivalent current measurement for a power flow measurement is:

$$I_{ij}^{meas-eqv} = ((P + jQ)_{ij}^{meas} / V_k)^* \tag{2.6}$$

where V_k is the complex voltage state estimate at iteration k. This improvement creates a linear approximation of the system that allows the measurement Jacobian, $\mathbf{H}$, to partition its rectangular components and remain constant, increasing computational efficiency. Then with partitioned equivalent current measurement vector, $\mathbf{z}^{eq}$ and system state vector, $\mathbf{x}$, the model becomes:

$$\begin{bmatrix} \mathbf{z}_{real}^{eq} \\ \mathbf{z}_{imag}^{eq} \end{bmatrix} = \begin{bmatrix} \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{real} \\ \mathbf{x}_{imag} \end{bmatrix} \tag{2.7}$$

where $\mathbf{H}$ is defined now by:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial \mathbf{z}_{real}^{eq}}{\partial \mathbf{x}_{real}} & \frac{\partial \mathbf{z}_{real}^{eq}}{\partial \mathbf{x}_{imag}} \\ \frac{\partial \mathbf{z}_{imag}^{eq}}{\partial \mathbf{x}_{real}} & \frac{\partial \mathbf{z}_{imag}^{eq}}{\partial \mathbf{x}_{imag}} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & -\mathbf{B} \\ \mathbf{B} & \mathbf{G} \end{bmatrix} \tag{2.8}$$

where **G** and **B** are the real and imaginary rectangular components of the $\mathbf{Y}_{bus}$.

During each iteration of DSSE, the measurement residual needs to be calculated before updating the state estimates. The measurement residual vector, **r**, is defined by:

$$\mathbf{r} = \mathbf{z}^{eq} - \mathbf{h}(\mathbf{x}) \tag{2.9}$$

The following section will look at using these measurement residuals for detection of noisy measurements and bad data.

2.2 Bad Data Detection Methods

As we've seen through derivation, the performance of DSSE relies upon the accuracy of the measurement data it receives. Thus, it is crucial to know when there is inaccurate data present, which can be done through various methods with the assumption that measurement errors follow a normal Gaussian distribution with known measurement variance. We adopt common methods of bad data detection (BDD) and identification for this work [1].

For the detection of the presence of bad data we adopt the Chi-Squared test. While considering a χ^2 distribution with $(m-n)$ degrees of freedom, the method tests the likelihood that the sum of squared normalized residuals is larger than a defined tolerance threshold. Explicitly, the test is shown by:

$$P\{X \geq x_t\} = \int_{x_t}^{\infty} \chi^2(u)du \tag{2.10}$$

where X is:

$$X = \sum_{i=1}^{m} \frac{(r_i)^2}{\sigma_i^2} \tag{2.11}$$

where σ^2 is the measurement variance corresponding to the i^{th} measurement. If it is determined that X is larger than x_t, which is derived from $\chi^2_{(m-n),p}$ for a given probability, p, bad data is suspected of being present.

The relationship between $\boldsymbol{\sigma}^2$ and $\mathbf{r}^2$ however is not as direct and introduces an approximation that could cause some approximation error [24]. Thus, an improvement can be made by replacing the measurement variance with the residual covariance matrix defined by:

$$\boldsymbol{\Omega} = E[\mathbf{r}\mathbf{r}^t] = \mathbf{S}\mathbf{R} \tag{2.12}$$

Here $\mathbf{S}$ is the residual sensitivity matrix defined as $\mathbf{S} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ and $\mathbf{R}$ is the measurement error covariance matrix. With this, a new term is calculated for this test and is defined below:

$$\Psi = \sum_{i=1}^{m} \frac{(r_i)^2}{\Omega_{ii}} \tag{2.13}$$

Once the presence of bad measurement data is detected, identification of the bad measurement is crucial for correction.

For bad data identification, largest normalized residual (LNR) test is adopted [1]. The measurement residuals found at the conclusion of DSSE are normalized with respect to $\boldsymbol{\Omega}$ defined in the previously. The normalized residuals, r^N are calculated using:

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}}, i = 1, ..., m \tag{2.14}$$

With the assumption that these r^N follow a Normal distribution, N(0,1), it is assumed that normal measurement noise is within 3σ, so any r^N greater than 3 is assumed to be bad data and the largest value is identified as the source of the bad data

2.2.1 Limitations

BDD is not without flaws however and experiences some limitations. Some limitations include:

1. **Reliance on number of measurements**: BDD is limited in the presence of critical measurements, which are measurements that upon removal creates portions of the network that are

unobservable. Thus, it is critical to obtain a level of redundancy, defined by $(m - n)$. This can be difficult to obtain in distribution systems without the use of pseudo-measurements obtained from load estimation or forecasting.

2. **Sensitivity to multiple interacting bad data**: When multiple bad measurements are present, the performance of BDD can be hindered [25]. Especially in the presence of what is called interacting and conforming bad data BDD will not be capable of identifying the source of bad data. Multiple bad data in measurements i and k are defined as interacting if their correlating term in the residual sensitivity matrix, $\mathbf{S}$ is large. Additionally, these interacting bad data are determined to be conforming if they are consistent with one another and typically of the same magnitude [1].

3. **Sensitivity to topology errors**: It should be noted that the result of BDD is the detection and identification of measurement data that are erroneous. However, when the network topology has changed and the model used within DSSE is incorrect, BDD will identify true measurements as bad data because of the model error. Thus, topology errors pose a significant challenge to DSSE and BDD. This is challenge is especially difficult in distribution systems given the "always changing" behavior of the radial network as stated previously.

2.3 Proposed Topology Error Monitoring Scheme

Using the residuals obtained from a DSSE solution limits BDD as it considers only the current snapshot of the system. To improve the BDD capability, we can extend the approach by including past DSSE solutions. Hence, we propose a scheme which considers two DSSE solutions: the current solution and the most previous solution. DSSE is often executed in 15-minute intervals, thus the measurement normalized residuals is expected to be relatively similar between consecutive iterations. A key advantage of the proposed method is the ability to see a change more dramatically that can indicate a change in the system topology.

If there is a topology error present, the change between the normalized residuals of the current iteration and the normalized residuals of the previous iteration is expected to be more extreme, so the LNR threshold of 3 can be relaxed to 2 and any normalized residual violating this threshold can indicate the presence of topology errors. This correlates to the change in normalized residual between consecutive BDD intervals, Δr^N, existing within 2 standard deviations or a confidence of 95% in a normal distribution.

It should be noted that the bad data to be detected by this approach will identify the suspected measurements, and since the topology is not a measured quantity, the last challenging step involves using this suspect measurement to identify if and which of the possible topology changes is present. To distinguish between measurement bad data and a topology error, we must analyze the suspect Δr^N and use observations of the differences in characteristics of Δr^N between situations of measurement bad data and topology errors. If we can identify unique characteristics in the BDD results from topology errors compared to BDD from measurement error, we can identify the presence of topology errors using these characteristics. This issue will be highlighted in the case study section.

The steps of the proposed scheme can be summarized below [19]:

1. Using the current DSSE, calculate the normalized residuals. If there is a normalized residual, $|r_i^N| > 2$, then bad data suspected and go to step 2. If not, complete.

2. Calculate Δr_i^N using:

$$\Delta r_i^N = \left| r_i^{N[1]} - r_i^{N[0]} \right|$$

   where $r_i^{N[1]}$ and $r_i^{N[0]}$ are the i^{th} normalized residual of the current solution and previous solution respectively.

3. Bad data is suspected if any Δr_i^N exceeds the threshold of 2.

4. Analyze characteristics of Δr_i^N. For example, observe both real and reactive measurements at node i and the other phases at that bus to identify the characteristics of a topology error.

Distinguishing characteristic is determined, topology error is present.

5. Largest Δr^N will be identified as suspected location of topology error using the LNR test.

2.4 Impact of Meter Placement Schemes on Topology Error Processing

The topology detection capability of BDD can be improved by the use of measurement placement schemes. The amount, type, and placement location of measurements all improve the capability of BDD to indicate the presence of topology errors. Different measurement schemes have been proposed in literature for the improvement of DSSE [17, 18]. Those schemes are ordered in terms of increasing number of measurements and are outlined below:

1. **Scheme 1: Basic Measurement Scheme**: The basic scheme for DSSE include power flow measurements at the substation and power injection pseudo-measurements at every node. Typically, load estimation results are used for these pseudo-measurements. These pseudo-measurements are defined with less weight in DSSE given less accuracy with estimation error.

2. **Scheme 2: Voltage Measurements at Critical Points**: In addition to previous scheme, voltage measurements are placed at critical points in the system to increase state estimate accuracy. These critical points are areas of most variability and deviation and include the end of the main feeder, ends of large laterals, voltage regulators and capacitor banks.

3. **Scheme 3: Extended Current/Power Flow Measurements**: In addition to the previous two schemes, current measurements are placed around areas of large load. Doing so delimits similar amounts of load throughout the system creating load zones. If flow directional and phase information is necessary as it is in multi-phase models, power flow measurements replace these current measurements. Current measurements typically lack phase information in practice, however are used more commonly due to economic factors. We denote *scheme*

3.1 as the use of current flow magnitude measurements and *scheme 3.2* as the use of real and reactive power flow measurements.

With these schemes, placement of measurements can be further optimized with a measurement elimination process to satisfy cost constraints [17]. This process primarily considers cost constraints with measurement that will vary the estimation accuracy the least upon removal.

It is vital to assess the effectiveness in which DSSE and BDD is capable of detecting topology errors given these measurement schemes. The next section will look at a few case studies examining detection of capacitor bank failure and load balancing with these schemes [26].

2.5 Case Study

A case study was performed using DSSE and the proposed topology monitoring scheme was implemented in MATLAB. The IEEE 34 node distribution test feeder and simulated data are used for the simulations [27]. Line information, bus data, regulator tap position, and capacitor bank rating are all modeled given the information from [27].

Measurement values were generated using CYME CYMDIST power flow results. Loading values at every node are taken to simulate load estimates and are used as load pseudo-measurements and include a ±30% noise. Power flow measurements are taken from CYMDIST results as actual measurements and have a ±3% noise added. Additionally, voltage magnitude measurements and current magnitude measurements are taken from CYMDIST solutions as well with a ±3% noise.

To simulate the system at the most previous time interval DSSE execution, the each load is decreased by 10% to capture a 10% load growth between DSSE intervals, which is a common distribution system operating condition change within a 15 minute interval. DSSE and BDD is executed with these conditions and the normalized residuals are recorded.

Bad data detection was done by implementing Chi-squares test and bad data identification was done using LNR test with a normalized residual threshold of 3. The proposed topology monitoring scheme is implemented comparing the normalized residuals of the current DSSE iteration to the normalized residuals of the previous DSSE solution and a detection threshold of 2 is used.

To evaluate effectiveness of the proposed scheme, we observe first if BDD is capable of detection of the presence of bad data caused by the topology error, then we observe if the correct node can be identified by the BDD method. We must also note the characteristics of r^N in order to identify the type of topology error detected.

To determine a proper measurement scheme for this sample feeder, the measurement schemes can be seen in Fig. 2.2 and are defined below:

1. Add load pseudo measurements at every node and power flow measurement at feeder head. (*Scheme 1*)

2. Add voltage measurements areas of large voltage variation, i.e. the end of the main feeder, the end of laterals with more than 2 nodes, etc. (*Scheme 2*)

3. Add current measurements around large load areas (*Scheme 3.1*)

4. Change current measurements to P and Q measurements (*Scheme 3.2*)

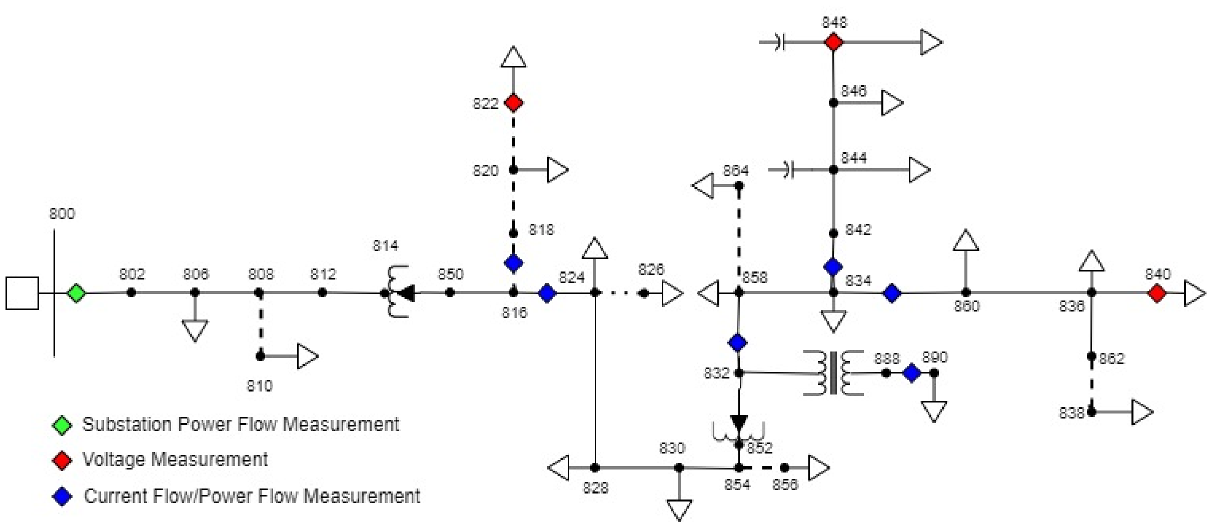

Figure 2.2 IEEE 34 Bus Test Feeder with Measurement Placement

Topology errors were simulated by changing the measurement values to reflect the system under these topology changes while keeping the model the same. Note that there are two capacitor banks (one 3 phase 100kVAR per phase bank at bus 844 and one 3 phase 150kVAR per phase bank at bus 848), so to simulate the first set of topology errors, capacitor failures associated with these CAP banks are simulated. The second set of topology changes simulate load transfers. Among the loads in the system, the loads at bus 860 and bus 844 will be used for load transfer scenarios. Thus, a practical case of a transformer switching phases to conduct load balancing through load transfer is simulated at these nodes. As a result, five different topology change scenarios are simulated.

- *Scenario 1*: 100kVAR per phase capacitor bank fails across all three phases and gets disconnected from bus 844
- *Scenario 2*: 100kVAR per phase capacitor bank fails on phase A and gets disconnected from bus 844
- *Scenario 3*: 150kVAR per phase capacitor bank fails across all three phases and gets disconnected from bus 848
- *Scenario 4*: 150kVAR per phase capacitor bank fails on phase B and gets disconnected from bus 848
- *Scenario 5*: 140kVA distribution transformer at bus 860 switches from phase C to phase A
- *Scenario 6*: 140kVA distribution transformer at bus 844 switches from phase A to phase C

The results for all five topology change scenarios are shown in tables 2.1-2.6 under all measurement schemes.

2.5.1 Case: Capacitor Failure

Table 2.1 and table 2.2 illustrate the results of the two scenarios involving the 100 kVAR per phase capacitor bank.

Table 2.1 Scenario 1: Three Phase Cap Failure at bus 844

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	χ^2-test Detected	–	–
Scheme 2	Not Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	Q_{844} ID	$P_a = 0.185$ $P_b = 0.012$ $P_c = 0.096$	$Q_a = 2.242$ $Q_b = 1.739$ $Q_c = 1.378$

Table 2.2 Scenario 2: Phase A Cap Failure at bus 844

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	χ^2-test Detected	–	–
Scheme 2	Not Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	Q_{844} ID	$P_a = 0.132$ $P_b = 0.032$ $P_c = 0.058$	$Q_a = 2.097$ $Q_b = 0.037$ $Q_c = 0.018$

From both of these tables, measurement scheme 3.2 with the proposed monitoring scheme is capable of detection and identification of bad data. For these scenarios, traditional BDD is incapable of identifying bad data as the detection threshold of 3 is not violated from the normalized residuals of the current iteration, whereas the change in normalized residual is proposed schemes threshold, showing the advantage of the proposed topology error monitoring scheme.

Figure 2.3 illustrates an example of the details of the method.

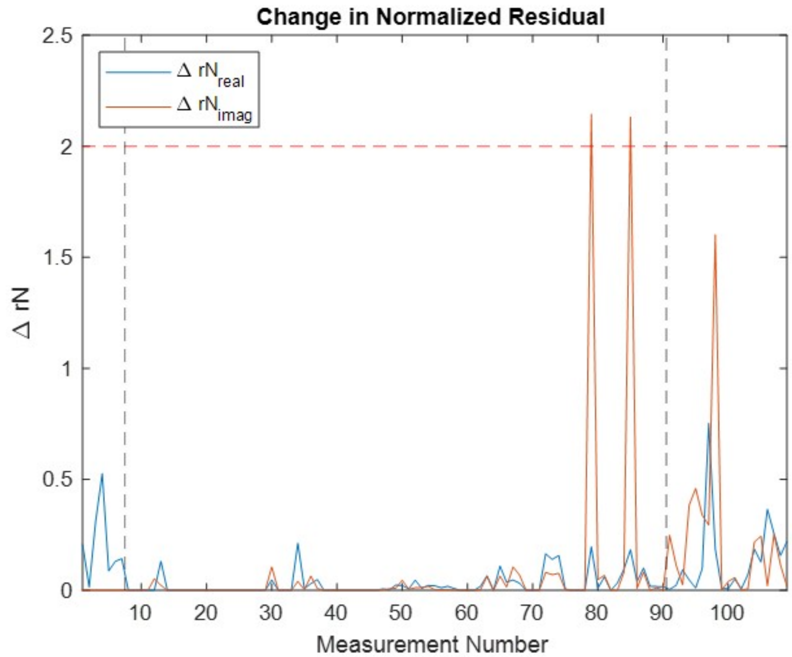

Figure 2.3 Change in normalized residual for scenario 2

From figure 2.3 the Δr^N is plotted for each residual corresponding to the given measurement number. Measurement numbers 1-7 refers to residuals of voltage magnitude measurement, measurement numbers 8-90 refer to the residuals of load pseudo-measurements, and measurement numbers 91-109 refer to the residuals of power flow measurements, with the blue plot referring to real power and the red plot referring to reactive power. As seen, Δr^N exceeds the threshold of 2 for the reactive load pseudo-measurement numbers 79 and 85 which refer to the phase A load on bus 844 and 848 respectively with measurement 79 having the largest Δr^N. For measurement number 79, the normalized residual of the previous DSSE solution, $r^{N[0]}$, is 0.0707 and the normalized residual of the current DSSE solution, $r^{N[1]}$, 2.1677, thus showing a sharp increase in normalized residual.

Table 2.3 and table 2.4 illustrate the results of the two scenarios involving the 150 kVAR per phase capacitor bank.

Table 2.3 Scenario 3: Three Phase Cap Failure at bus 848

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	χ^2-test Detected	–	–
Scheme 2	χ^2-test Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	Q_{844} ID	$P_a = 0.168$ $P_b = 0.175$ $P_c = 0.068$	$Q_a = 3.317$ $Q_b = 2.669$ $Q_c = 2.170$

Table 2.4 Scenario 4: Phase B Cap Failure at bus 848

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	χ^2-test Detected	–	–
Scheme 2	Not Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	Q_{844} ID	$P_a = 0.166$ $P_b = 0.157$ $P_c = 0.097$	$Q_a = 0.133$ $Q_b = 2.538$ $Q_c = 0.072$

From table 2.3 and table 2.4, it can be seen again that using measurement scheme 3.2 with the proposed monitoring scheme is capable of detection and identification of bad data. For scenario 3, which is a three phase failure, traditional BDD is able to identify bad data, however in scenario 4, which is a single phase failure, traditional BDD is unable to identify bad data. Scenario 4 is a more practical scenario as it is more likely only a single phase of a capacitor bank will fail at a time rather than all three phases, thus if the proposed monitoring scheme is capable of identifying bad data in this more realistic case, this again shows the advantage of the proposed monitoring scheme.

In both scenarios, it should be noted that the reactive load at the incorrect bus is identified. The change in normalized residual at bus 848 does violate the detection threshold but is not the largest, thus is not identified. However, this information and the proximity of bus 844 to the correct location can be used to reduce the network to the general area for further search of the source of the bad data.

It should also be noted that traditional use of a single snapshot of BDD could miss the detection of the presence of a capacitor failure. However, taking the change in normalized residual into consideration using the proposed scheme, the capacitor failure is detected in all scenarios using measurement scheme 3.2.

For all four capacitor failure scenarios, it should be noted the difference between the change in normalized residual of real and reactive power. From tables 2.1-2.4 the change in normalized residual is much larger for reactive power than for real power, especially for the phases impacted (all three phases for tables 2.1 & 2.3, phase A for table 2.2, and phase B for table 2.4). This indicates only the reactive power is erroneous and will be used for further analysis.

2.5.2 Case: Load Transfers

The final two scenarios involving load transfers are shown in tables 2.5 & 2.6.

Table 2.5 Scenario 5: 140 kVA from Phase C to Phase A at bus 860

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	χ^2-test Detected	–	–
Scheme 2	χ^2-test Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	P_{860} ID	$P_a = 2.127$ $P_b = 0.008$ $P_c = 2.241$	$Q_a = 1.087$ $Q_b = 0.029$ $Q_c = 1.189$

Table 2.6 Scenario 6: 140 kVA from Phase A to Phase C at bus 844

Measurement Scheme	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
Scheme 1	$P_{800,802}$ ID	$P_a = 0.982$ $P_b = 0.000$ $P_c = 2.515$	$Q_a = 1.097$ $Q_b = 0.093$ $Q_c = 0.000$
Scheme 2	χ^2-test Detected	–	–
Scheme 3.1	Not Detected	–	–
Scheme 3.2	P_{848} ID	$P_a = 2.259$ $P_b = 0.058$ $P_c = 2.001$	$Q_a = 1.678$ $Q_b = 0.035$ $Q_c = 1.456$

As seen from tables 2.5 and 2.6, scheme 3.2 with the proposed monitoring scheme is capable of detection and identification of bad data. Note that in this case the large residual corresponds to the real power. Using the residuals obtained from a DSSE limits BDD as it is shown again traditional BDD is unable to identify the bad data.

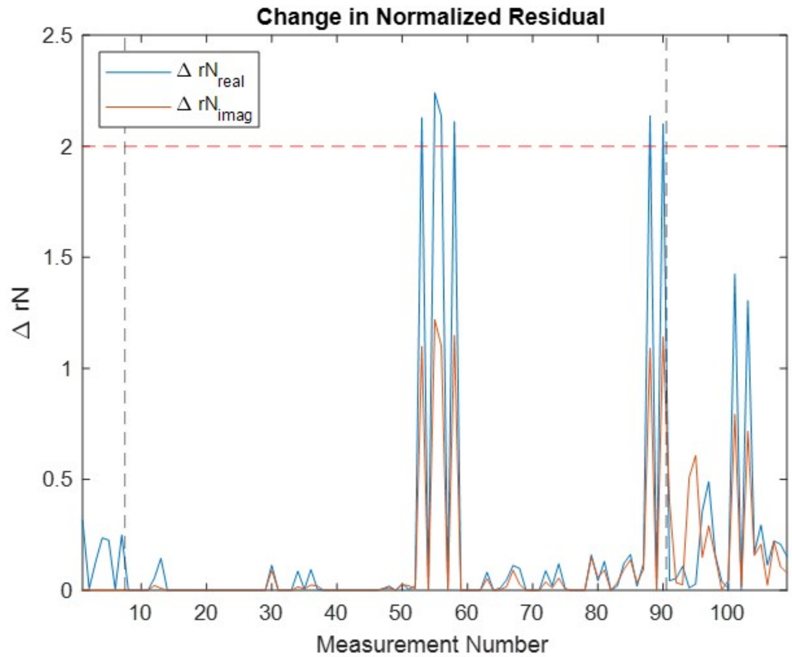

Figure 2.4 Change in normalized residual for scenario 5

Figure 2.4 shows an example of the details for the load transfer scenarios. As seen, Δr^N exceeds the threshold of 2 for the real load pseudo-measurements. Measurement numbers 53, 56, 88 refer to the phase A load at bus 860, bus 836 and bus 840 respectively and measurement numbers 55, 58, and 90 refer to the phase C load at bus 860, bus 836, and bus 840 respectively. This shows that phase A and phase C is impacted within this section of the network delimited by the power flow measurement from bus 834 to 860. For measurement number 55, which is the largest change in normalized residual, $r^{N[0]}$ is 0.0834 and $r^{N[1]}$ is 2.3244, again showing a large change in normalized residual.

Through the proposed monitoring scheme with measurement scheme 3.2, for scenario 5 (table 2.5), the correct location of the load transfer is identified (bus 860) and the change in normalized residual for phases A and C are comparable and much larger than phase B, indicating these two phases are impacted. However, in scenario 6 (table 2.6) the location identified (bus 848) is not the correct location (bus 844). Similar to scenario 4 (table 2.4), the change in normalized residual at bus 844 does violate the detection threshold, however is not the largest, thus is not identified. Again, this information along with the proximity of 844 and 848 can allow for further search in a reduced general area of the network.

As with capacitor failures, the proposed scheme is capable of detection using the change in normalized residual that would be missed in traditional methods of using a single snapshot, thus showing the improvement the proposed scheme provides.

With measurement scheme 3.2, the change in normalized residual between the two phases experiencing the load transfer having similar magnitudes and being much larger than the third phase indicates that the two phases are experiencing an error in similar manner, which can be used for further analysis. Additionally, the larger real change in normalized residuals than reactive change in normalized residuals indicates that the real power is more erroneous than the reactive power.

It is noted that BDD identifies measurements that are suspect of bad data, but not directly indicating the topology change that may have caused the error. Thus to address this issue, a comparison needs to be made between the results of BDD in these topology scenarios and the results of BDD with bad measurement data. We must identify the unique characteristics in the BDD results to distinguish the presence of topology errors instead of measurement bad data.

In practice, a three phase power measurement, real and reactive power are impacted in the same manner due to how the meter generates this data, meaning if there is a error in the measurement, both the real and reactive power are impacted. To simulate this, in measurement scheme 3.2, an error of 5σ is added to $P_{832,842}$ and $Q_{832,842}$. The results of BDD are for single phase measurement error and three phase measurement error are shown in table 2.7 and table 2.8 respectively.

These results indicate that these actual bad measurements impact both real and reactive normalized residuals and make the associated LNRs larger than the threshold value. For the single phase A measurement error, the real and reactive normalized residuals for only phase A are large, and all three phases have large normalized residual for the three phase measurement error. Previously in

Table 2.7 BDD Results Caused by Single Phase Bad Measurement

Scenario	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
$S_{834,842}$Single phase Error	$P_{834,842}$ ID	$P_a = 2.476$	$Q_a = 2.625$
		$P_b = 0.034$	$Q_b = 0.009$
		$P_c = 0.019$	$Q_c = 0.029$

Table 2.8 BDD Results Caused by Three Phase Bad Measurement

Scenario	Proposed BDD	Δr^N_{real}	$\Delta r^N_{reactive}$
$S_{834,842}$Three phase Error	$P_{834,842}$ ID	$P_a = 2.497$	$Q_a = 2.614$
		$P_b = 3.507$	$Q_b = 2.001$
		$P_c = 2.883$	$Q_c = 1.831$

this section we noted that in the presence of topology errors, either real or reactive power normalized residuals were large and violated the threshold and not both. Thus, these differences can help us to distinguish a topology error from an actual bad measurement.

2.6 Conclusions

In this chapter, an assessment conducted of traditional C-DSSE and BDD performance in the presence of challenging and common topology errors, like capacitor failures and load transfers. A new approach is proposed to use the traditional BDD methods in order to be used for topology error detection. A case study is presented on a sample test feeder and test results indicate that there are overall limitations present in using these traditional methods for topology error monitoring. These limitations will be used as motivation for new approaches for further enhancement in the next chapter.

The main contributions of the chapter are:

- A new approach to extending traditional BDD methods to incremental analysis for enhanced performance needed for topology error monitoring.

- An assessment of the performance of C-DSSE and BDD methods in the presence of common and challenging topology errors and the minimum measurement placement scheme to enhance BDD capabilities is identified from common practical placement schemes.

- Preliminary signatures are observed that will be used for topology error detection.
- Limitations are identified in the identification of these topology errors that will encourage further development in the next chapter.

CHAPTER

3

METHODS FOR MONITORING OF TOPOLOGY ERRORS

In chapter 2, we observed that limitations still exist with standard methods of BDD for topology error processing in the presence of common topology errors in distribution systems. In the presence of capacitor bank failures and load phase switching, detection is improved and possible in most cases with the proposed approach, yet there is still limitations in identifying the source of the respective topology change. We take two steps to address these limitations and to propose a topology error monitoring method.

In this chapter, we first focus on improvements that can be made to the capability of traditional BDD-based methods to be used for topology error processing. This will be our first step in development of our proposed method. In the traditional method, the modeling of measurement uncertainty is approximated to save computational burden, especially with very large distribution networks. However, this approximation severely limits BDD and the overall accuracy of the C-DSSE solution. To address this and enhance BDD's capability, we propose the adoption of a new approach to model uncertainties in measurements more accurately in C-DSSE. The improvement in C-DSSE accuracy is assessed using a comprehensive set of performance metrics on a test feeder that is based on a real distribution system.

As a result, the capability of using BDD for topology monitoring is enhanced. From this, we can generalize the specific signatures of these topology errors through a numerical observations of residual spread-based analysis. This will be our second step and basis for our proposed method.

Ultimately, using these signatures the main contribution is the proposal of a new methodology to accurately monitor for these challenging topology errors through the signatures in the enhanced BDD that are provided by the improvements to measurement uncertainty modeling.

3.1 Improving Measurement Uncertainty Modeling

As we have seen, DSSE has multiple components that impact the overall performance including the system modeling, system topology, and the measurements and measurement model. In an effort to monitor the system topology for errors, we focus on improving another component which is the modeling of measurements. It becomes critical to accurately model the limited amount of measurements that are available in order to improve the capability to monitor for topology errors. WLS-based state estimation utilizes the modeling of the measurement uncertainty to properly weigh measurement residuals to prioritize minimization of residuals. The weights are derived from the inverse of the measurement variances [1], thus measurements with smaller variance will have a higher weight and its residual will be prioritized to be reduced as close to zero as possible.

In practice, the measurements available are voltage magnitudes, current magnitudes, and complex branch power flow, including real and reactive power. Each type of meter has a constant and known percentage uncertainty, provided by the manufacturer of the device. For computational simplicity, it is common to approximate the measurement uncertainty based upon the known meter accuracy[5, 20–22]. However, this approximation does not consider the true measurement uncertainty that is dependent upon the measured value, thus creating an inaccurate uncertainty modeling and limits the ability to calculate appropriate weights for the measurement residuals, which can lead to less accurate and lower quality state estimates.

Load modeling is another important issue in DSSE, as the loads are not monitored in real-time. In DSSE, to increase redundancy of measurements and ensure observability, load estimates are used as pseudo-measurements [28]. Pseudo-measurement accuracy is reliant on the load estimates, thus accurate modeling of load estimate uncertainty becomes important as it can limit the quality of DSSE [2].

Ultimately, the goal of this effort to improve measurement uncertainty modeling is aimed to improve the overall quality and accuracy of the state estimates and enhance the capability of developing a topology monitoring method using BDD-based methods.

3.1.1 Accurate Measurement Model

Field measurements received from SCADA contain an inherent noise that is usually modeled with a normal Gaussian distribution:

$$z = z_{true} + e \tag{3.1}$$

where z is the measurement value received through SCADA, z_{true} is the true field value, and $error$ is Gaussian measurement noise $\sim N(0,\sigma)$ [29]. If σ is the measurement standard deviation, as mentioned, it is traditional to approximate σ using known meter accuracy, a [5]. This a is typically given as a ±% and is the maximum expected error, which correlates to 3σ in a normal distribution, thus σ is given by:

$$\sigma = \frac{a}{3} \tag{3.2}$$

In practice however, the true measurement accuracy is dependent on z_{true} as well [30]. Specifically, the maximum expected error, 3σ is the ±% times z_{true}, creating a normal distribution centered around z_{true} with sigma:

$$\sigma = \frac{z_{true} * a}{3} \tag{3.3}$$

Additionally, it is common to model all measurements as independent random variables. Power measurements of different meters are indeed independent of each other however, a meter calculates the real and reactive power measurements based on voltage and current measurements. Therefore, measurements of the same meter should not be modeled independent of each other [31, 32]. In other words, if a real power measurement experiences % noise, the reactive power measurement of that PQ pair will experience the same % noise.

The accuracy of these pseudo-measurements is reliant on the estimation method and is much less accurate than field measurements. Accurate modeling of load estimate uncertainty should then consider a wider variance that is dependent on factors like values for P and Q, which is typically gathered by historical load curves [2]. Using this relationship demonstrated in [2], we can extend the variance modeling into a similar modeling of pseudo-measurement uncertainty as (3.3) as the variance changes in proportion with the historical load curve value.

It should be noted that these measurement variances are for field measurements and load pseudo-measurements. However, in C-DSSE, measurements are converted to equivalent current measurements. Therefore, the variances of the equivalent current measurements must be determined [33].

To convert R to R^{eqv} for equivalent measurements, the general form of the variance sum law must be used [34]. For two random variables X and Y, and function $Z = aX \pm bY$:

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 \pm 2ab\sigma_{XY}^2 \tag{3.4}$$

where σ_Z^2, σ_X^2, and σ_Y^2 are the variances of Z, X, and Y respectively and σ_{XY}^2 is the covariance between X and Y. Using the relationship between given power measurements and equivalent current measurements given by:

$$I^{real} = \frac{V_r}{V_r^2 + V_i^2} P + \frac{V_i}{V_r^2 + V_i^2} Q \tag{3.5}$$

$$I^{imag} = \frac{V_i}{V_r^2 + V_i^2} P - \frac{V_r}{V_r^2 + V_i^2} Q \tag{3.6}$$

where V_r and V_i are estimated real and imaginary rectangular components of node voltage at the current iteration of C-DSSE. If we assume P and Q random variables, we can rearrange in the form of (3.4) to obtain the equivalent measurement variance.

3.1.2 Performance Metrics

To assess the improvement in quality and accuracy of the state estimates when the more accurate measurement modeling is used, it is critical to define statistical performance metrics that can be used to properly assess the performance improvement. We have adopted the following main statistical performance metrics from [35, 36]:

3.1.2.1 Bias

Bias is the tendency of the expected value of an estimator to be different than what is true. In other words, for a state estimator to be defined as unbiased, the expected value of error in the state estimate is zero, meaning the expected state estimate should be equivalent to the true state. We can mathematically define an unbiased estimator as

$$E\left[(x_t - \hat{x})\right] = 0 \tag{3.7}$$

where x_t is the true state vector and $\hat{x}$ is the estimated state vector. This metric provides an understanding of the general accuracy of the state estimator.

To estimate this metric and the others introduced below using simulations, we adopted the most commonly used approach -the Monte Carlo method- as it is the most effective process of estimating sample statistics of state estimation [37]. To obtain the statics, we assume measurement error follows a normal distribution, and hence, measurement samples are obtained by perturbing the measurements following the distribution of each measurement and C-DSSE state estimates are obtained for each measurement sample.

To estimate the bias using Monte Carlo simulations, the state error, $(x_t - \hat{x})$, is observed for each iteration of Monte Carlo and if the state error varies about zero mean, the state estimator is said to be unbiased. Hypothesis testing can be utilized to further verify more definitively that there is no bias by testing the following null hypothesis, H_0, and alternative hypothesis, H_a, on a 5% significance level:

- H_0: $\mu_{(x_t - \hat{x})} = 0$
- H_a: $\mu_{(x_t - \hat{x})} \neq 0$

For a sample that comes from a normal distribution with unknown population mean, a one-sample, two-sided T-test can be used as the hypothesis test method [38]. The t-statistic is the test statistic found in t-tests and can be calculated by:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \tag{3.8}$$

where $\bar{X}$ is the sample mean, μ is the population mean, which is 0 in the null hypothesis, s is the sample standard deviation, and n is the sample size. From a t-distribution with n-1 degrees of freedom, the t-statistic has a corresponding statistical likelihood that it fits the null hypothesis, called the p-value. If the p-value is smaller than the significance level, then the null hypothesis is rejected and therefore there is significant evidence that the expected error in the state estimates is not zero, implying there is bias.

3.1.2.2 Quality/Accuracy

The quality of an estimate is inversely associated with the variance of the estimated quantity. In other words an estimate is said to have higher quality if the variance is smaller. For a multivariate case, such as C-DSSE, the trace of the error covariance matrix, P_x, can be used. Thus, quality is defined by:

$$Q_{trace} = log\left(\frac{1}{trace(P_x)}\right) \tag{3.9}$$

where $P_x = inv\left(H^T W H\right)$. It must be noted that Jacobian and weight matrices in P_x must be in terms of the equivalent current measurements since it is used in C-DSSE.

Q_{trace} will increase with increasing accuracy (lower variance) in the estimation, hence can be used in comparative analysis to show the improvement in the more accurate measurement modeling.

3.1.2.3 Error Rate

As C-DSSE is used for real-time monitoring in a distribution management system (DMS), it becomes critical that C-DSSE accurately monitors for any voltage violations that must be addressed by a utility's Volt/VAR control, meaning C-DSSE must estimate potential voltage violations as accurately as possible, minimizing the number of false negatives and false positives. To illustrate this, error rate is defined keeping track of the false positives (α risk) and false negatives (β risk) [36]:

$$Q_{violation} = \frac{\alpha + \beta}{N_{simulation}} \tag{3.10}$$

where $N_{simulation}$ is the number of Monte Carlo simulations. Monte Carlo is then used to estimate the likelihood of C-DSSE giving a false positive or a false negative. Thus, when comparing, a smaller $Q_{violation}$ is desired.

3.1.3 Distribution System State Estimation Performance Evaluation

Simulations were performed with C-DSSE using MATLAB. Several models of measurement uncertainty are utilized to show the impact of each improvement in the more accurate measurement modeling approach. The models tested include:

1. **Model 1**: Constant weights based on $\sigma = \frac{a}{3}$ of given measurements [5]
2. **Model 2**: Constant weights based on $\sigma = \frac{z*a}{3}$ of given measurements [30]
3. **Model 3**: Weights based on $\sigma = \frac{a}{3}$ converted to W^{eqv} [33]
4. **Model 4**: Weights based on $\sigma = \frac{z*a}{3}$ converted to W^{eqv} (proposed more accurate measurement model)

It should be noted that when W is converted to W^{eqv}, it needs to be updated every iteration of C-DSSE because of the dependence on the state estimate of the current C-DSSE iteration as seen in (3.5) and (3.6). This increases the computational burden, however if the initial state estimates are based off of a flat start, W^{eqv} will be inaccurate. Thus, in our comparative analysis, W^{eqv} in Model 3 is updated once after the first update of state estimates and held constant for the rest of C-DSSE to be more accurate than a flat start while providing the computational advantage of an approximated model.

To illustrate the effectiveness in a practical case, a large sized, realistic distribution network based on an actual feeder is also used. In this realistic distribution network, system modeling including line/cable modeling, bus modeling, etc., is based on an actual feeder. Measurement values were obtained using CYME CYMDIST power flow results for the realistic distribution network as well. The load values are estimated by the using a load allocation method and are provided at every node. We used these as load pseudo-measurements and are given a ±30% noise. Power flow and voltage magnitude values are taken from CYMDIST results as actual measurements and both have a ±3% noise added. The feeder one-line diagram and placement of measurements are shown in Fig. 3.1 [18, 36, 39, 40].

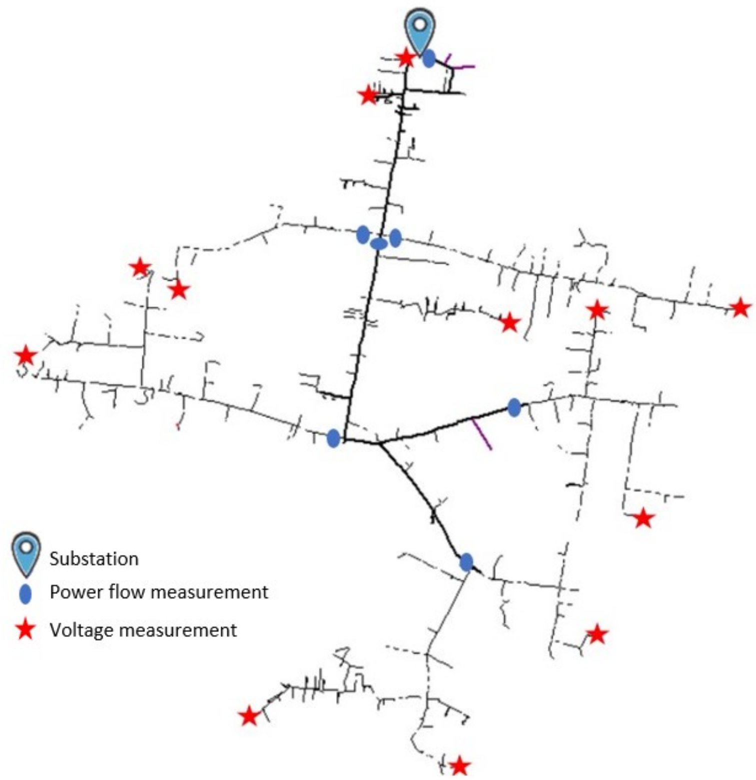

Figure 3.1 Realistic System One-Line Diagram with Measurement Placement

Within the system model, there are 1522 nodes for 3038 states that must be estimated in C-DSSE. Monte Carlo simulations are conducted by adding a random noise to each measurement that follows a normal distribution specific for each measurement and executing C-DSSE. In Monte Carlo it is important to have enough simulations to provide an accurate enough sample to estimate the statistical metrics. For a guide of the minimal number of Monte Carlo simulations need, the following relationship is defined [41]:

$$d = sqrt\left(\frac{\alpha(1-\alpha)}{N_{simulation}}\right) \tag{3.11}$$

where d is the acceptable standard deviation for the estimate and α is the target confidence interval.

It is determined for a .05 α and acceptable d of 0.01, that about 500 simulations is necessary to assure accurate results.

3.1.3.1 Bias

Figure 3.2 shows the estimation error across Monte Carlo for the all measurement models using the realistic distribution network.

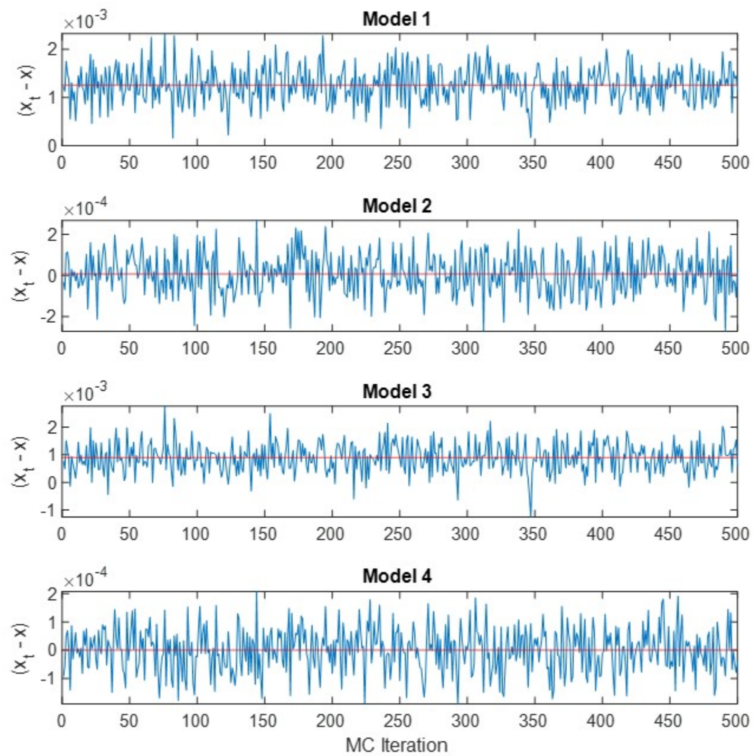

Figure 3.2 Estimation Error for Each Measurement Model Using Realistic Distribution Network

From these figures, visually it is shown that the estimation error is distributed around a zero mean for constant weights based on $z * a$ (model 2) and the proposed more accurate measurement model (Model 4). Visually, it is also clear that models based on only considering constant meter accuracy (Model 1 and Model 3) are not distributed around a mean of zero and are clearly biased.

The estimation error seems to vary around the sample mean of the estimation error of 1.2609e-03, 7.3943e-06, 8.9435e-04, and 3.5048e-07 for models 1, 2, 3, and 4 respectively. Again here, we see improvement in average estimation error from Model 1 to Model 2. Model 3 does not improve as much, however, Model 4 shows the best improvement as it has the lowest average estimation error.

To further check if there is bias, we also conducted hypothesis testing using the sample statistics. Given that the estimation error is normally distributed, under a 5% significance level with null hypothesis that the population mean estimation error is zero, the results of the one-sample t-test are provided in table 3.1.

Table 3.1 One-Sample T-test Results Using Realistic Distribution Network

Measurement Model	P-value	Decision ($\alpha = 0.05$)
Model 1	1.204e-276	Reject H_0 (Biased)
Model 2	0.0914	Fail to Reject H_0 (Unbiased)
Model 3	4.359e-147	Reject H_0 (Biased)
Model 4	0.9212	Fail to Reject H_0 (Unbiased)

It is determined that the null hypothesis is rejected with a p-value much smaller than 0.05 for the traditional approximated measurement model (Model 1) and the equivalent current weights based on constant meter accuracy (Model 3). This means that the true population mean of estimation error is not zero, which then indicates that using these models with real implementation on the realistic distribution network creates a biased estimator. As seen with the average estimation error, there is improvement for the constant weights based on $z * a$ (Model 2) as the t-statistic is 1.6913 and the p-value of 0.0914 is larger than the significance level, so it is likely that it is unbiased. However, for the more accurate measurement model (Model 4), the t-statistic is 0.099 and the p-value of 0.9212. The t-statistic is significantly smaller, which means it is much more likely that the estimator is unbiased.

These results show that as the system model gets larger and more complex as in the realistic distribution network, the more accurate measurement models 2 and 4 have greater impact on reducing the bias in C-DSSE estimates.

3.1.3.2 Quality/Accuracy

The state error covariance matrix is obtained for all measurement models and the quality is calculated analytically using (3.9) and gathered numerically from the Monte Carlo sample similar to the IEEE 34 bus case study. Table 3.2 shows the resulting values for quality for the realistic distribution network analysis.

Table 3.2 C-DSSE Quality Using Realistic Distribution Network

Measurement Model	Trace(Px) (Analytical)	Quality (Analytical)	Trace(Px) (Numerical)	Quality (Numerical)
Model 1	0.03288	3.415	0.00652	5.032
Model 2	0.000112	9.095	0.00013	8.930
Model 3	0.02312	3.767	0.00959	4.647
Model 4	0.000092	9.296	0.000095	9.261

From the table, it is evident that the more accurate measurement model (Model 4) improves the quality of C-DSSE when using the realistic distribution network. Numerically the state error variance found from the Monte Carlo simulation is large for Models 1 and 3, which use only constant meter accuracy for the original given measurement variance, resulting in a small quality value. For Model 2, which is constant weights based on original given measurement $z * a$ the state error variance decreases, resulting in a better quality value of around 9. However, for the proposed more accurate measurement model (Model 4), the numerical variance is much smaller than the others, which results in the largest quality value of 9.296. This is clear indication that the more accurate measurement model improves the quality of C-DSSE especially in a practical case.

3.1.3.3 Error Rate

For error rate, a voltage violation is considered if the voltage magnitude is outside of the standard range of 0.95 p.u.-1.05 p.u. The number of false positives, α, and the number of false negatives, β are recorded across the Monte Carlo simulation. Table 3.3 presents the amount of false positives and false negatives and the error rate for the approximated measurement model and the more accurate measurement model for the realistic distribution network case.

Table 3.3 C-DSSE Error Rate Using Realistic Distribution Network

Measurement Model	False Pos. (α)	False Neg. (β)	Error Rate
Model 1	15,843	8,644	48.974
Model 2	1,897	1,887	7.568
Model 3	23,129	10,166	66.590
Model 4	1,652	1,442	6.188

From the table we can see a significant improvement in the error rate for the more accurate measurement model (Model 4) compared to the traditional approximated measurement model based on constant weights from original constant meter accuracy (Model 1). For the more accurate measurement model, in the 500 Monte Carlo iterations, there was a total of 1,652 false positives and 1,442 false negatives for an error rate of 6.188 errors per Monte Carlo iteration. This is an acceptable rate because there are 1522 nodes with an operating point close to 1.05, so a large amount of the false positives and false negatives are due to rounding and smaller distance from the true voltage to the violation threshold. Even given this, for the traditional approximated measurement model based on constant weights (Model 1), there is a significantly larger amount false positives and false negatives. Through the Monte Carlo simulation, there were 15,843 false positives and 8,644 false negatives for a error rate of 48.974 errors per Monte Carlo iteration, which is not an acceptable rate. Thus, 14,191 false positives and 7,202 false negatives are eliminated when implementing the more accurate measurement model.

In Model 2, which is the use of constant weights based on original given $z * a$, there is also good improvement from Model 1. However, in Model 3, which is using equivalent current weights based on constant meter accuracy, it seems there are more errors. This shows that a most of the improvement comes from more appropriate given measurement variance coming from $z * a$.

These results show that including both improvements in the proposed approach (Model 4) provides the best performance in error rate. Model 4 eliminates 245 false positives and 445 false negatives (error rate of 6.188) as compared to Model 2 (error rate of 7.568).

Considering all performance metrics, there is clear and evident improvement in the quality/accuracy in C-DSSE with improved measurement models. We see similar results between the Model 2 and proposed Model 4, but marginal improvement in bias and quality in Model 4. Model 4 however does demonstrate significant improvement in error rate as compared to Model 2, which is the most significant performance in distribution system state estimation. This shows that both modeling measurement accuracy and equivalent current variance modeling is critical in improving performance of C-DSSE. Results from this realistic distribution network test case are especially significant to show that the improvement is scalable to practical system sizes and the improvement in quality is more apparent for practical implementation.

3.1.3.4 Computational Requirement

Given the size and complexity of the realistic distribution feeder, it becomes critical to analyze the computational effort required by the more accurate measurement model. As stated previously, the advantage of using the approximated model is for computational simplicity as the weight matrix does not have to be updated every iteration. To verify this computational advantage, several statistics across the Monte Carlo simulations are recorded including total execution time, average number of iterations for C-DSSE to converge and average time taken per C-DSSE iteration. All values are recorded from running Monte Carlo on a 1.8GHz Intel Core i7-8550U processor personal computer. The results are provided in table 3.4

Table 3.4 C-DSSE Computational Requirements Using Realistic Distribution Feeder

Measurement Model	Total Time (sec)	Avg. # of Iter. to converge	Avg. Time per Iter (sec)
Model 1	25194	8.022	6.2811
Model 2	23644	7.378	6.4103
Model 3	32210	10.248	6.2869
Model 4	21072	6.746	6.2486

From table 3.4, it appears that the more accurate measurement model (model 4), where the weight matrix is updated every iteration, actually has the smallest total time. This appears as there is overall reduction in the average number of iterations C-DSSE needs to converge. Thus again showing that the computational burden is not increased significantly using the proposed approach to modeling measurement uncertainty.

It should be noted that computational effort can be made more efficient with more sophisticated matrix computational methods. With this, the main limitation of this approach is the recalculation of the weight matrix at every iteration of C-DSSE, thus if a more sophisticated matrix computational method is used, this limitation can be addressed.

3.2 Residual Spread-Based Analysis

In the previous chapter, observations are made on specific signatures that are present in BDD results in the presence of either a capacitor failure or load transfer. For capacitor bank failures, we observe that the resulting normalized residuals are largest amongst the reactive power measurements and pseudo-measurements and contained within the measurement island, which again is a subsection of the system bounded by a branch power flow meter, that bounds the failed capacitor. For load transfers coming from load balancing, we observe that the resulting normalized residuals are largest for real power measurements and pseudo-measurements and slightly smaller for reactive power measurements and pseudo-measurements also contained within the measurement island that bounded the load transfer.

To ensure the generality of these observations, we must analyze the technical residual spread from a WLS-based state estimator especially with the improved more accurate measurement model. It is considered that these topology errors of capacitor bank failure and load transfer are similar to a pseudo-measurement error at its respective node because each impact the estimated power injection at that node. Because of this, we can analyze how WLS-based state estimation spreads that measurement error to its measurement residuals. It is shown that the relationship between measurement error, e, and measurement residuals, r, is as follows [1]:

$$r = Se \tag{3.12}$$

Where $S = I - K = I - \left(HG^{-1}H^T R^{-1}\right)$ is the residual sensitivity matrix, I is an identity matrix, K is the projection matrix, H is the measurement Jacobian, R is the measurement covariance matrix, and $G = H^T R^{-1} H$ is the gain matrix. If there is a single measurement with error ($e_i \neq 0$, $e_j = 0$ for all $j \neq i$), then the i^{th} column of S will determine the spread of that e_i to the r vector.

In literature, areas of the system that confine the spread of a specified e_i are called residual spread areas and can be determined analytically [42]. To show this, authors take advantage of the theorem that critical measurements, which are measurements that upon removal of creates a portion of the system that is unobservable, cannot have detectable measurement error residual. As a result, S can be structured in into block-diagonal form where each diagonal block of S represents a residual spread area and where an error in the j^{th} does not affect the residual in the i^{th} measurement if they are not in the same residual spread area ($S_{ij} = 0$). This is done by partitioning the non-critical and critical measurement in the measurement Jacobian, H, such that:

$$H = \begin{bmatrix} H_{11} & & \\ & \ddots & \\ & & H_{kk} \\ & H_c & \end{bmatrix} \tag{3.13}$$

Where k is the number of non-critical measurements and H_c refers to the rows of H relating to the critical measurements. As a result, H of the non-critical measurements becomes block-diagonal and the resulting projection matrix, K, also becomes block-diagonal for non-critical measurements:

$$K = \begin{bmatrix} H_{11}\left(H_{11}^T R_{11}^{-1} H_{11}\right)^{-1} H_{11}^T R_{11}^{-1} & & & \\ & \ddots & & K_{kc} \\ & & H_{kk}\left(H_{kk}^T R_{kk}^{-1} H_{kk}\right)^{-1} H_{kk}^T R_{kk}^{-1} & \\ & K_{ck} & & K_{cc} \end{bmatrix} \tag{3.14}$$

Where K_{ck} refers to the rows of K relating to the critical measurements, K_{kc} refers to the columns of K relating to the critical measurements, and K_{cc} is the diagonal block referring to the critical measurements.

However, this is not applicable in DSSE because load estimates are used as pseudo-measurements at every node in the distribution system, including nodes with no load. Though these pseudo-measurements are not highly accurate, this creates the scenario where removal of any measurement does not create an unobservable system, meaning there are no critical measurements. Therefore, S cannot be structured in block-diagonal form and an error in the j^{th} measurement will have spread throughout the system.

Though it is more challenging to distinguish the residual spread areas analytically in DSSE, the residual spread can be shown *numerically* to localize most of the error to some residuals more than others. This refers to a larger sensitivity value for some measurement residuals than others given a specified measurement error. Again, $S = I - H\left(H^T R^{-1} H\right)^{-1} H^T R^{-1}$, so the more appropriate weights determined by the more accurate measurement uncertainty modeling will be shown to impact this.

Assume a simple single-phase 4 bus test system as an example for this analysis. This test system and the measurements are shown in Fig. 3.3

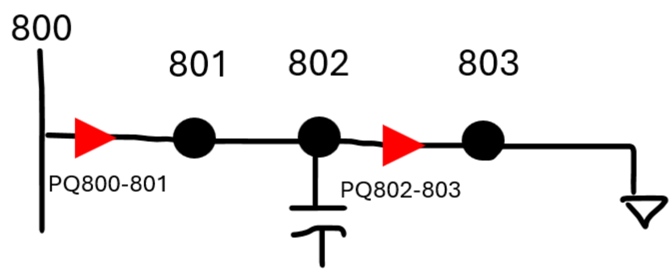

Figure 3.3 Simple 4 Bus Example One-Line Diagram

In this example, node 800 is treated as the reference bus and branch power flow measurements are placed on branches 800-801 and 802-803 creating two measurement islands. Pseudo-measurements are placed on node 803 for real and reactive power load and on node 802 for reactive power injection from the shunt capacitor. Zero-injection pseudo-measurements are placed otherwise. The resulting S matrix is shown in Table 3.5:

Table 3.5 Simple 4 Bus Example Sensitivity Matrix

	P800-801	Q800-801	P801	Q801	P802	Q802	P802-803	Q802-803	P803	Q803	W
P800-801	0.507228	-0.19261	0.504514	-0.19367	0.501606	-0.19459	-0.49629056	0.188801	0.004979	-0.00595	316.218
Q800-801	-0.03867	0.492902	-0.00202	0.507054	0.037245	0.519272	-0.03734914	-0.51347	0.000364	0.006021	1574.966
P801	1.77E-05	-3.5E-07	1.81E-05	-2E-07	1.85E-05	-7.4E-08	-1.8347E-05	-5.3E-08	1.84E-07	-1.3E-07	9000000
Q801	-6.8E-06	8.87E-05	-2E-07	9.13E-05	6.87E-06	9.35E-05	-6.8871E-06	-9.2E-05	6.76E-08	1.1E-06	9000000
P802	1.76E-05	6.52E-06	1.85E-05	6.87E-06	1.95E-05	7.17E-06	-1.9321E-05	-7.2E-06	1.94E-07	-4.3E-08	9000000
Q802	-0.00886	0.117768	-9.6E-05	0.121153	0.009297	0.124076	-0.00931863	-0.12266	9.21E-05	0.001468	6944.444
P802-803	-0.47032	-0.17629	-0.49485	-0.18576	-0.52114	-0.19394	0.526222397	0.188316	0.004754	-0.00578	333.6775
Q802-803	0.024846	-0.33655	-0.0002	-0.34618	-0.02703	-0.35449	0.026150237	0.362248	-0.00027	0.008042	2402.922
P803	0.459152	0.167219	0.483117	0.177442	0.508863	0.186446	0.462541157	-0.19192	0.995359	0.005738	3.429355
Q803	-0.06092	0.307253	-0.0377	0.31932	-0.01261	0.3303	-0.06243683	0.626096	0.000638	0.991837	30.8642

Where the column of the matrix refers to the j^{th} measurement with error and the row of the matrix refers to the i^{th} measurement residual. In the table, the darker red cells within a column indicate the largest sensitivity value in the column and the lighter red shaded cells refer to the moderate sized sensitivity values. In the last column labeled "W", the weights of each measurement residual are shown to illustrate the impact the weight has on the sensitivity value. As seen with the more accurate measurement uncertainty modeling, measurement variances are dependent on the measured value, thus a larger measurement or load will result in a larger variance, which will equate to a smaller weight.

Through numerical analysis, several key observations can be made. Looking first at the columns of the last measurement island containing node 803, the largest sensitivity values can be seen to be contained almost exclusively within the rows that refer to measurement residuals in that same measurement island. This means that any error in P803 or Q803 will spread predominantly to the residual of P803 or Q803 respectively. It can also be seen that these measurement residuals have the smallest weights. Thus, the lower weighted residuals with respect to surrounding residual weights will localize the residual much more as the sensitivity value is much larger.

However, looking at the columns referring to the first measurement island containing nodes

801 and 802 (P801, Q801, P802, and Q802), any error in real power pseudo-measurements will mostly spread to the two bounding branch flow measurements (P800-801 and P802-803) as well as downstream into the downstream measurement island (P803). As seen the weights in P801 and P802 are much larger (9e6) than P800-801, P802-803, and P803 (316.218, 333.6775, and 3.429355 respectively), again showing that in a WLS-based method, the measurement residuals will be spread predominantly to the much lower weights. It is emphasize that an error in this beginning measurement island spreads a near majority into the downstream measurement island. This indicates that because there are not enough lower weighted residuals in that measurement island to absorb most of the spread, the residual spread will not be localized within a measurement island as seen in our previous observations of the BDD signatures.

As a result, residual spread itself is not enough to be used as a verification of the localization of the topology error based signature. If there are not enough large loads (which result in small weights) in a measurement island, residual spreads between measurement islands and cannot localize residual spread to a measurement island, which does not verify the signatures observed with BDD signatures in the presence of capacitor failures and load transfers.

However, the signatures observed in the simulations are in terms of *normalized residuals*, so it becomes critical if the spread related to normalized residuals can be used to verify the signatures that these topology errors can be localized to measurement islands that contain the topology error.

3.2.1 Normalized Residual Spread-Based Analysis

As mentioned, normalized residual can be found using:

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}}, i = 1, ..., m \tag{3.15}$$

Where, $\Omega = SR$ is the measurement residual covariance matrix and m is the number of measurements. Using (3.12), (3.15) can be re-written as:

$$r_i^N = \sum_{j=1}^{m} \frac{\left|S_{ij}\right|\left|e_j\right|}{\sqrt{S_{ii}}\sqrt{R_{ii}}}, i = 1, ..., m \tag{3.16}$$

Hence, we can define a normalized sensitivity matrix, S^N, where,

$$S_{ij}^N = \frac{\left|S_{ij}\right|}{\sqrt{S_{ii}}\sqrt{R_{ii}}}, i = 1, ..., m, \& j = 1, ..., m \tag{3.17}$$

Employing the numerical analysis to S^N that was applied to S, we can observe how normalized residuals are spread and localized based on a given measurement error in the j^{th} measurement. Using the same example illustrated in Fig. 3.3, Table 3.6 shows the resulting normalized sensitivity matrix.

Table 3.6 Simple 4 Bus Example Normalized Sensitivity Matrix

	P800-801	Q800-801	P801	Q801	P802	Q802	P802-803	Q802-803	P803	Q803	W
P800-801	12.6647	4.809125	12.59694	4.835666	12.52433	4.858611	12.39160972	4.714079	0.12433	0.14846	316.218
Q800-801	2.185974	27.86223	0.114387	28.66219	2.10536	29.35284	2.111232404	29.02487	0.020582	0.340357	1574.966
P801	12.4937	0.249589	12.76936	0.143518	13.06471	0.051962	12.93108332	0.037351	0.129746	0.091133	9000000
Q801	2.136674	27.86201	0.063937	28.66242	2.157046	29.35347	2.162549514	29.02214	0.02123	0.343849	9000000
P802	11.96891	4.426368	12.58851	4.665277	13.25241	4.87151	13.12160114	4.901292	0.13168	0.029374	9000000
Q802	2.09626	27.86148	0.022604	28.66226	2.199362	29.35363	2.204590855	29.01898	0.021782	0.347298	6944.444
P802-803	11.84335	4.439191	12.46108	4.677683	13.12301	4.883618	13.25098423	4.742061	0.119706	0.145428	333.6775
Q802-803	2.023574	27.41028	0.016166	28.19483	2.201571	28.87162	2.129818314	29.50345	0.022308	0.654973	2402.922
P803	0.852262	0.310386	0.896744	0.329362	0.944534	0.346075	0.858552352	0.356242	1.847549	0.01065	3.429355
Q803	0.339827	1.713969	0.210329	1.781284	0.070358	1.842536	0.34829584	3.492598	0.003556	5.532833	30.8642

In Table 3.6, looking again at the columns relating to the last measurement island that contains node 803, it is again clear that the rows within the same measurement island are have the largest normalized sensitivity value within the columns. However, now looking at the columns of P801, Q801, P802, and Q802, it is shown that the largest normalized sensitivity values remain in the rows of the residuals in that same measurement island containing nodes 801 and 802. This indicates that a large majority of the normalized residual spread will be localized and confined within the same measurement island with a large drop-off in normalized sensitivity value in rows that are outside of that measurement island.

From this numerical analysis, it is clear that in a WLS-based state estimator, such as C-DSSE, the normalized residual can be localized to within the measurement island that the error exists, even if the actual residual is shown to spread between measurement islands, which verifies the signatures observed with topology error simulations. If a capacitor failure is treated as an error in reactive load pseudo-measurement, it will predominantly spread amongst the reactive power residuals and remain localized within the measurement island that the capacitor failure occurs. Alternatively, if load transfers are treated as more real load pseudo-measurement error, it will also predominantly remain localized within the measurement island that the load transfer occurs.

It should be noted that this is shown on a very simple sample feeder. However, this normalized residual spread-based numerical analysis can be shown as an effective strategy to be used as a verification tool of the signature even on larger more complex systems. Application of this strategy can show that if a topology error is suspected, going to the column of the suspected error in the normalized sensitivity, the normalized residual spread can be numerically observed and the verification can be made if it matches the signature present in the localized normalized residuals from BDD.

The normalized residual spread-based analysis will be the basis of the proposed topology error monitoring method that will be described next.

3.3 Proposed Topology Error Monitoring Methodology

The proposed topology error processing method is broken into three steps:

1. Detection of presence and type of topology error using machine learning-based method.

2. Localizing the detected topology error to a smaller section of the system based on signature-based analysis.
3. Identification of topology error through a modified search method.

3.3.1 Topology Error Detection

After obtaining the improved normalized residuals, it must be further determined if a topology error is detected rather than an erroneous measurement because topology is not a measured quantity. Additionally, if a topology error is detected the type of topology error should be suspected. To do so, we can utilize Support Vector Machines (SVM) to assist in this detection [43–45]. In SVMs, given a set of data with different classifiers, a model is trained to find the most optimal separating hyperplane(s) between data of different classes. An example of a linear based SVM is shown in Fig. 3.4

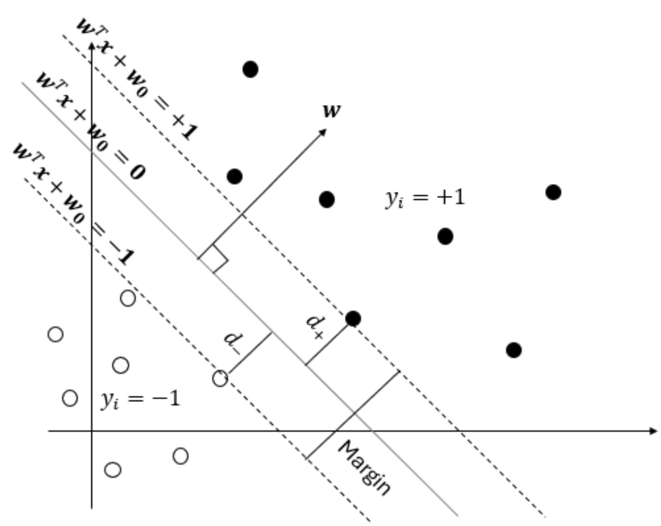

Figure 3.4 Example Support Vector Machine hyperplane

Where **w** is a weight vector defining the coefficients of the hyperplane and support vectors, **x** is the vector of features, y is the classifier, d_+ and d_- are the distances of the closest data point of each classification to the hyperplane and whose sum determines the margin. To find the optimal hyperplane, the maximum margin is determined.

In the general case the decision boundary can be determined by:

$$\mathbf{w}^T \mathbf{x} + w_0 = \sum_{i \epsilon SV} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0 \tag{3.18}$$

Where α is a Lagrange multiplier and $K(\cdot)$ is a Kernel function meant to map the feature space to a high dimensional space at no extra cost. A general Kernel function that is commonly used is the Radial Basis Function (RBF) and is defined as [46]:

$$K(X, X') = exp\left[-\frac{\left\|X - X'\right\|^2}{2\sigma^2}\right] \tag{3.19}$$

Where σ is the variance and $\left\|X - X'\right\|^2$ is the Euclidean distance between X and X'. From this the decision function is then:

$$f(\mathbf{x}) = sign\left(\sum_{i \epsilon SV} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0\right) \tag{3.20}$$

As seen in [19] and in the normalized residual-based analysis, there are distinct characteristics in the normalized residuals as it relates to the real and reactive power components for capacitor failure and load transfer topology errors. For capacitor failures, the normalized residuals relating to reactive power are large and are much larger than its real power counterpart. Alternatively, for load transfer, both real and reactive power normalized residuals are determined as large, with a larger normalized residual relating to the real power component because of the typical customer load power factor that gets transferred between phases.

Given a large enough sample size, we can train the SVM model on cases that include capacitor failure, load transfer, and no topology error (including only measurement error) so that it can accurately predict if there is a topology error and specifically if it is a capacitor failure or load transfer scenario. Ultimately, this step aims at predicting with high certainty a suspected topology error type for further investigation.

It should be noted that this model is trained on capacitor failure, load transfer, and no topology error scenarios that an included fourth case should be considered in the training that includes unmodeled topology errors, thus if a given case does not fit capacitor failure nor load transfer, it should be considered either no topology error or a topology error not considered in training.

3.3.2 Localizing Detected Topology Error

In this step, the goal is to localize the detected topology error to a reduced subsection of the system to create a general idea of where the error could have occurred. To do so, we must utilize the signatures in the normalized residuals to localize the detected topology error to a specific measurement island. As seen from the Normalized Residual Spread-Based Analysis, the largest normalized residuals will be localized within a measurement island that the error occurs and the normalized sensitivity matrix can be used as a verification.

Additionally, each topology error has specific signatures in addition to its signature of localization. First, the signatures for capacitor failure are as follows:

1. Largest normalized residuals are observed as reactive power measurements (branch flow & load pseudo-measurement)
2. Only the phases that the capacitor fails on will experience large normalized residual.

Alternatively, the signatures for load transfer are as follows:

1. Largest normalized residuals are observed for real power measurements.
2. The two phases involved in the load transfer (the original phase and the new phase) will experience large normalized residual.

Thus, using the detected topology error type and signatures related to localization and topology error type, the suspected area of topology error can be reduced to the measurement island and the phases that are suspected of this topology error.

This will create an advantage in further pinpointing the exact source of the topology error in the next step.

3.3.3 Topology Error Identification

Once the type of topology error is detected and the localized subsection of the system is determined, a search method can be used to further pinpoint the most likely source of the topology error. In this search, all suspected components within a localized subsection are iteratively modified and re-evaluated with DSSE. As a metric for comparison, sum of squared normalized residual is used and determined by:

$$\Psi_m(\hat{x}) = \sum_{i=1}^{m} \frac{(r_i)^2}{\Omega_{ii}} \tag{3.21}$$

The iteration resulting in the smallest Ψ is then identified as the most likely source of the detected topology error.

In distribution systems, there are a finite number of capacitors placed with further reduced amount contained within a localized subsection and phase, thus the search area is improved with the localization step. We have used a similar approach in [47] for the identification of failed capacitors based on results of traditional normalized residuals rather than the improved normalized residuals. We extend this to load transfers.

For load transfers, as identified in localizing step II, a signature is observed where only two phases within a localized subsection experience large normalized residuals. To identify suspected components for the search method to iterate, any single phase load or branch that is connected to a three or two phase bus is selected and included within the search.

3.4 Case Study

To assess the performance of the proposed method, case studies are provided to show the effectiveness. We first assess the performance of the proposed topology error detection step using Support

Vector Machines. We then provided sample analysis for the localization and identification steps for capacitor failure cases and load transfer cases on two realistic feeders.

3.4.1 Topology Error Detection

To assess the performance of the detection step of the proposed method, we evaluate the performance of the proposed SVM method with RBF kernel prediction accuracy and F1-score compared to other common kernel function and another supervised machine learning classifier, random forests. F1-score can be determined by:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.22}$$

Where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

To collect a large enough sample of data, we simulate 14 capacitor failure cases and 7 load transfer cases across the two realistic feeders. For all cases simulated included, 10 Monte Carlo samples each are collected where the measurement noise is perturbed according to a normal distribution. Additionally, we include a balanced amount of Monte Carlo samples where there is no topology error. In total, 250 samples are collected.

Each data point collected includes the resulting averages of the 10 largest normalized residuals of real power measurements and 10 largest normalized residuals of reactive power measurements. This is an effort to ensure that this detection method is robust to any measurement placement scheme with varying numbers of measurements.

Each model is trained on 70% of the data called training data and the performance is tested on the remaining 30% of the data called the test data.

Table 3.7 provides the test performances for the various models.

Table 3.7 Detection Model Performance

Detection Model	Accuracy	F1-Score
SVM w/ Linear Kernel	92.00	92.00
SVM w/ polynomial Kernel (Degree: 3)	72.00	71.55
SVM w/ RBF Kernel	100.00	100.00
Random Forest	96.00	95.91

As a result, it is shown in Table 3.7 that the proposed method of SVM with RBF kernel has the best performance amongst other SVM models and another similar supervised machine learning classifier, random forests. The resulting confusion matrix for the SVM with RBF kernel is shown in Table 3.8.

Table 3.8 Confusion Matrix of SVM w/ RBF Kernel

	Pred. 0	Pred. 1	Pred. 2
Obs. 0	13	0	0
Obs. 1	0	42	0
Obs. 2	0	0	20

Where the classifier 0 refers to no topology error cases/unmodeled topology error, classifier 1 refers to capacitor failure cases, and classifier 2 refers to load transfer cases. Thus, with extensive data of various examples of capacitor failure cases and load transfer cases across two different large realistic feeders, the SVM with RBF kernel proves to be an effective model for detecting the presence and the type of topology error.

As a result, with accurate detection of the topology error type, the next steps are to localize the detected topology error and identify the source.

3.4.2 Case: Capacitor Failure

As we have shown an effective detection methodology using SVMs with the RBF kernel, we now simulate the localization and identification steps first for detected capacitor failure cases. As an example of the analysis, we provide applications of the method on two realistic feeders that are based on real feeder data, Feeder AM and Feeder WA.

3.4.2.1 Realistic Feeder AM Capacitor Failure Case

We simulate a single phase capacitor failure first on phase A on realistic Feeder AM and implement the localization and identification steps of the method. The system diagram with measurement placement is shown in Fig. 3.5. The simulated failed capacitor is a 400kVAR per phase capacitor bank and its location is indicated in the figure.

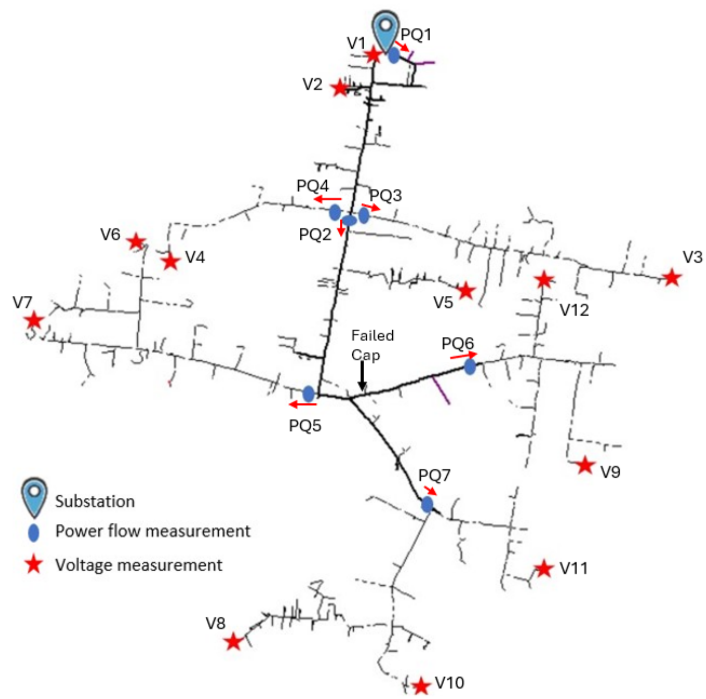

Figure 3.5 Realistic Feeder AM Capacitor Failure Case

As a result of the BDD normalized residuals, we confirm that the detection step detects the presence of capacitor failure topology error and begin the localization step. To localize the detected capacitor failure, we analyze heat maps of the normalized residuals for reactive power across all three phases. For this case, a signature is identified for reactive power on phase A Fig. 3.6 shows the heat map of normalized residuals for reactive power for phase A.

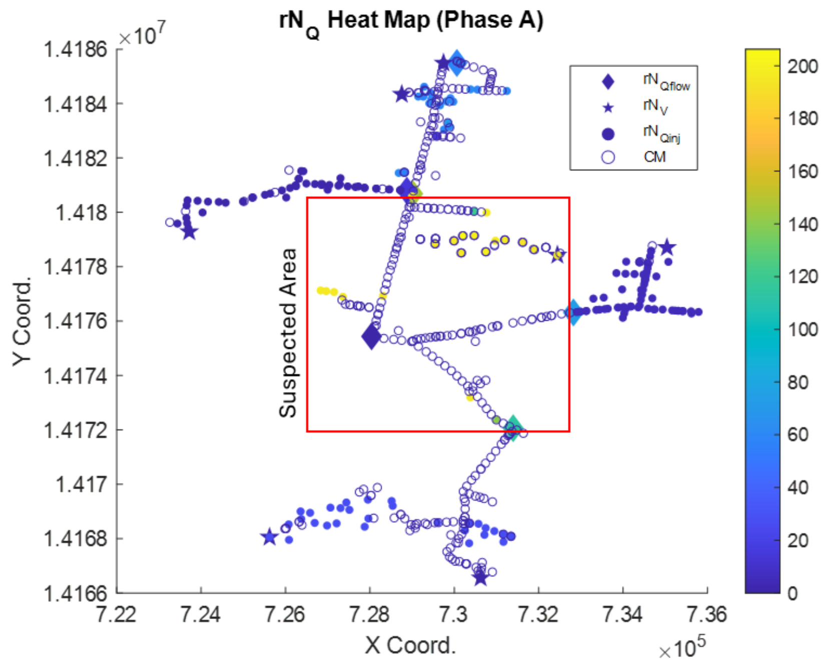

Figure 3.6 Realistic Feeder AM r_Q^N Heat Map (Phase A)

Examining the other heat maps for the other phases, the normalized residual remains small, but normalized residual for phase A gets as large as 200 as seen in the figure. Utilizing the normalized residual spread-based analysis approach, we see that the large normalized residuals, indicated by the yellow filled circles, can be localized to the middle measurement island, thus the suspected localized area is indicated by the red box. It should be clarified that there is a blue diamond that refers to a power flow measurement in the middle of the section indicated. This is a measurement on a lateral off of the left side of the main backbone, thus does not create a new measurement island and the measurement island in the red box is one measurement island.

As a result of localizing the detected failed capacitor to phase A in the middle measurement island, the identification step can begin with the search method. The search method automatically investigates the measurement island for available capacitor and only a single capacitor on phase A in the measurement island, so it is likely the source of the failed capacitor topology error. However, to verify this, the search method adjusts the capacitor model and re-executes C-DSSE/BDD and obtains a new Ψ value as defined by (3.21) and compares it to the original where the topology error is present. Table 3.9 presents the results.

Table 3.9 Realistic Feeder AM Capacitor Failure Search Results

Case	Ψ
W/ Topology Error	9.182e5
Corrected Topology	**140.656**

As seen in Table 3.9, after localizing the capacitor failure to the middle measurement island, only a single capacitor on phase A is indicated from the search method. As a result, the capacitor is disconnected from phase A to correct the topology and using Ψ (the sum of squared normalized residual) as a metric, the corrected capacitor topology improves Ψ significantly from $9.182e5$ to 140.656.

This shows the effectiveness of both the localization step to reduce the search surface area down to a single measurement island and the search method to prove the correct identification of the failed capacitor.

3.4.2.2 Realistic Feeder WA Capacitor Failure Case

As an alternative example to show this methods application across different large feeders, we now simulate a capacitor failure on realistic Feeder WA and implement the localization and identification steps of the method. The system diagram with measurement placement is shown in Fig. 3.7. The simulated capacitor is a single phase, 150kVAR capacitor that is connected to phase B. This capacitor is indicated in the figure.

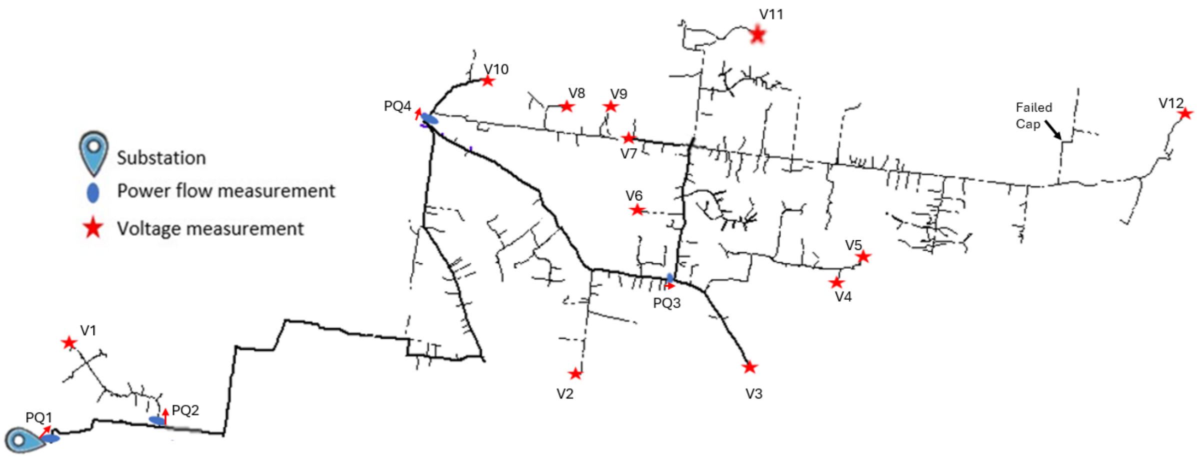

Figure 3.7 Realistic Feeder WA Capacitor Failure Case

We confirm the detection step detects the presence of a topology error and detects a capacitor failure topology error type. As a result, the localization step begins and the heat maps for reactive power normalized residuals are analyzed across all three phases. A signature is identified on phase B and the heat map is shown in Fig. 3.8.

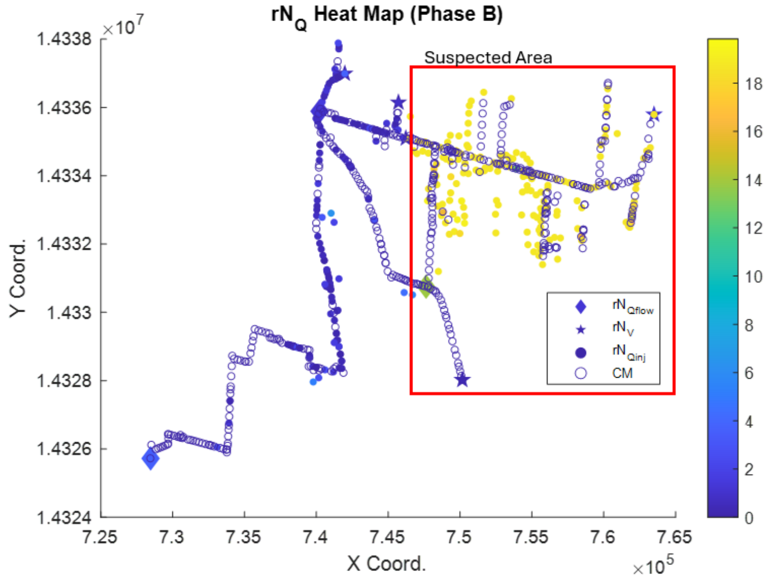

Figure 3.8 Realistic Feeder WA r_Q^N Heat Map (Phase B)

Utilizing the normalized residual spread-based analysis approach, we see that the large normalized residuals, indicated by the yellow filled circles, can be localized to the end measurement island, thus the suspected localized area is indicated by the red box.

As a result, the search method reduces the search surface area to this measurement island on phase B. The search method identifies 3 capacitors available and calculates Ψ to identify the most likely source of the capacitor failure. The results are shown in Table 3.10.

Table 3.10 Realistic Feeder WA Capacitor Failure Search Results

Case	Ψ
W/ Topology Error	6.996e4
Cap 1	2.058e5
Cap 2	6.996e4
Cap 3 (Actual Cap)	**72.918**

From Table 3.10, it is clear that the actual failed capacitor is identified by the search method as Ψ is much lower than the other simulated corrections at 72.918, which is also much improved from the case with topology error of $6.996e4$. This again shows the effectiveness and the advantages provided by the localization and identification steps to accurately and efficiently identify the correct topology error across different large realistic feeders.

3.4.3 Case: Load Transfer

We now focus on implementing the method on load transfer cases to illustrate the effectiveness on this topology error type. As an example of the analysis, we provide applications of the method on the same two realistic feeders that are based on real feeder data.

3.4.3.1 Realistic Feeder AM Load Transfer Case

In this case, we now simulate a load transfer case on realistic Feeder AM and implement the localization and identification steps of the method. The branch that is simulated as a load transfer is shown in Fig. 3.9 in the red box and switches the entire branch from phase A to phase C.

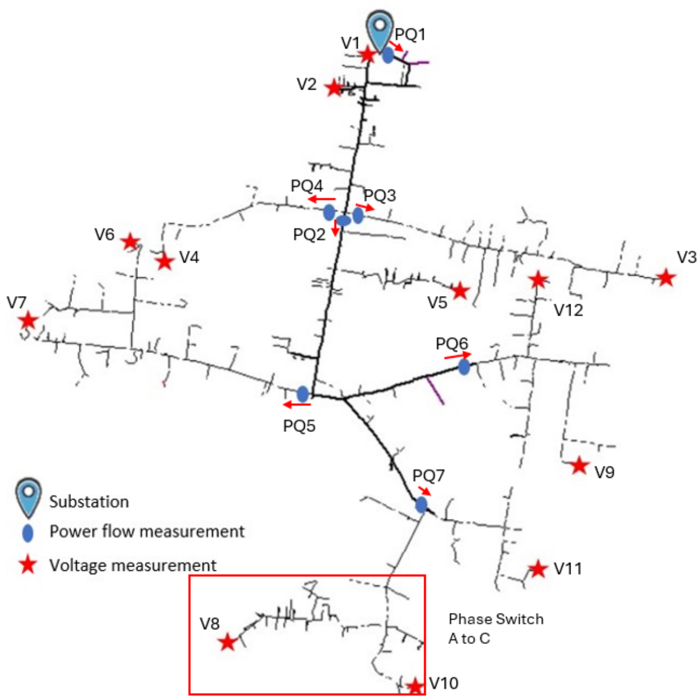

Figure 3.9 Realistic Feeder AM Load Transfer Case

It is confirmed that the detection step detects a topology error and detects the load transfer topology error type. As a result, the localization and identification steps can begin. Examining the normalized residuals for real power across the three phases, a signature is identified in phase A and C. Fig. 3.10 shows the real power normalized residual heat maps for phases A and C.

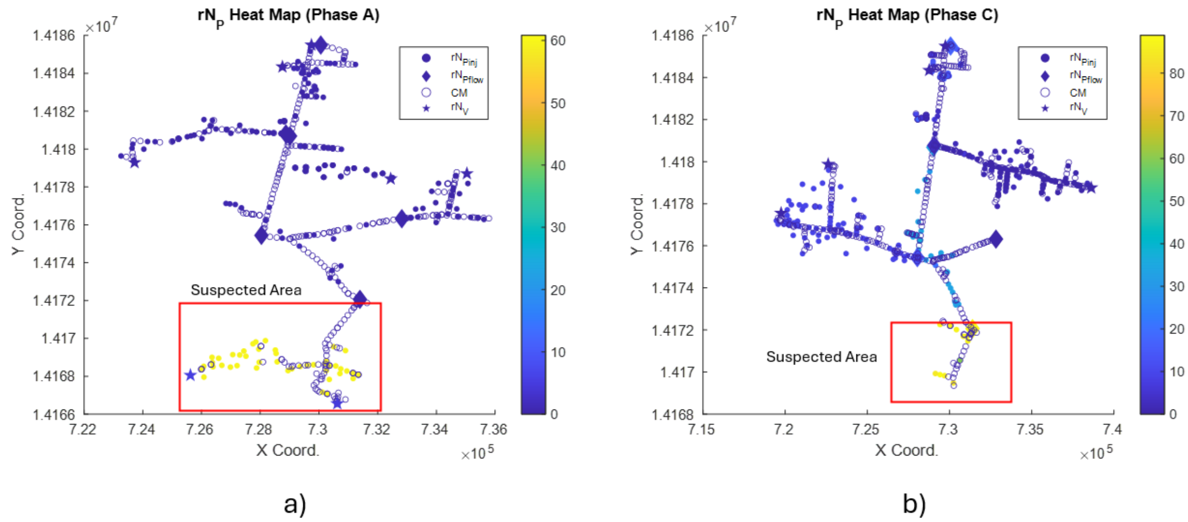

Figure 3.10 Realistic Feeder AM r_P^N Heat Map (a) Phase A (b) Phase C

From Fig. 3.10, we can observe that normalized residuals are large for phases A and C for real power measurements and can localize the detected load transfer between phases A and C in the end measurement island that is indicated by the red box.

As a result, in the identification step, the search method begins searching for branches and loads where the switch is possible (i.e. is modeled on one of the phases and disconnected on the other phase). Of the 1136 branches in the system, the localization step and search method reduce the search surface area to 4 possible branches. Thus, the search method begins attempting each possible branch and uses Ψ to identify the most likely source of the topology error. The results are shown in Table 3.11

Table 3.11 Realistic Feeder AM Load Transfer Search Results

Potential Branch	Ψ
W/ Topology Error	4.427e5
871-872	6.223e5
890-891	5.167e5
895-900	5.047e5
906-907 (actual branch)	**846.582**

From Table 3.11, it can be seen that the search method identifies the correct branch (branch 906-907). Ψ for this simulated correction is 846.582, which is much lower than the other possible branches and is much lower value than the original Ψ where the topology is present (4.427e5).

These results shows the effectiveness of the localization and identification steps now for the load transfer case on a large realistic feeder. The reduction of the search area from 1136 branches to just 4 branches also shows the key advantage of the localization step.

3.4.3.2 Realistic Feeder WA Load Transfer Case

Finally, a load transfer case is simulated on realistic feeder WA and the localization and identification steps of the method are implemented. Fig. 3.11 displays the simulated load transfer, which switches the phase of the branch from phase B to phase C.

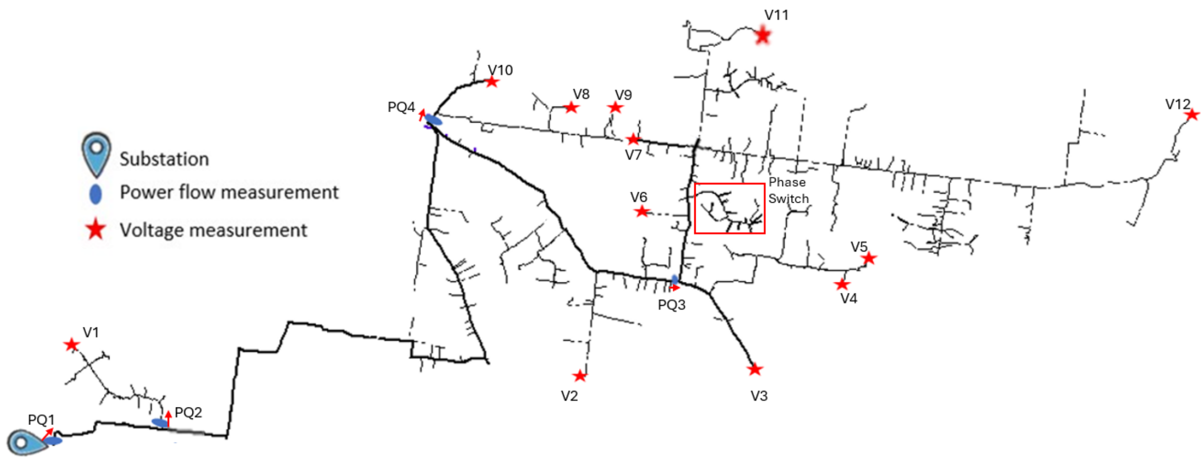

Figure 3.11 Realistic Feeder WA Load Transfer Case

It is confirmed that the detection step detects a topology error and detects the load transfer topology error type. As a result, the localization and identification steps can begin. Examining the normalized residuals for real power across the three phases, a signature is identified in phase B and C. Fig. 3.12 shows the real power normalized residual heat maps for phases B and C.

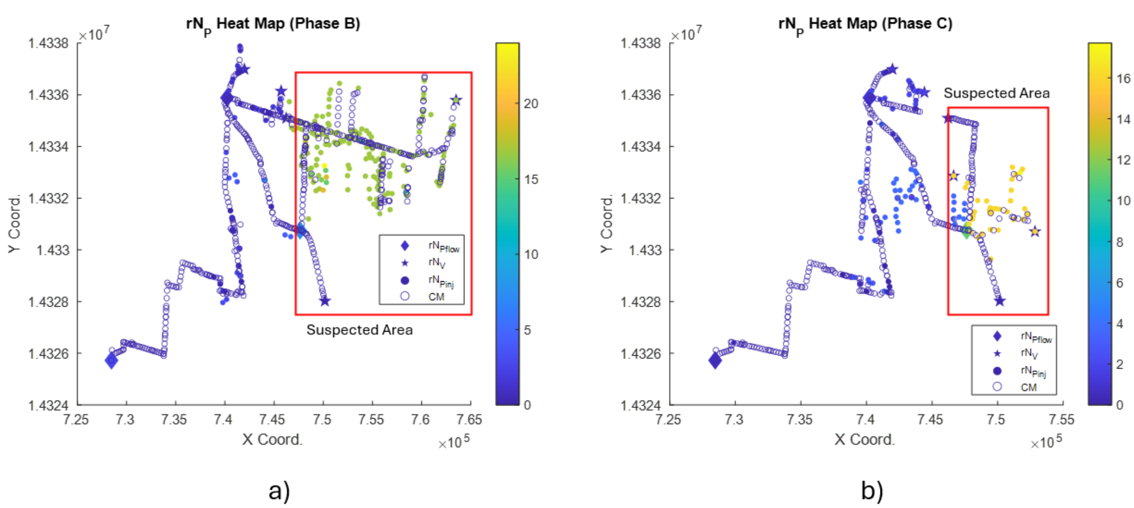

Figure 3.12 Realistic Feeder WA r_P^N Heat Map (a) Phase B (b) Phase C

From Fig. 3.12, we can observe that normalized residuals are large for phases B and C for real power measurements and can localize the detected load transfer between phases B and C in the end measurement island that is indicated by the red box.

As a result, the search method begins searching for branches and loads where the switch is possible. Of the 1005 branches in the system, this localization and search method reduce the search surface area to 19 possible branches. Thus, the search method begins attempting each possible branch and uses Ψ to identify the most likely source of the topology error. The results are shown in Table 3.12

Table 3.12 Realistic Feeder WA Load Transfer Search Results

Potential Branch	Ψ
W/ Topology Error	8.395e4
370-390	4.428e8
394-395	1.057e5
398-399	9.441e4
402-403	1.023e5
406-407	1.143e5
366-367	8.349e4

Potential Branch	Ψ
412-413	1.076e5
419-420	8.089e4
419-422 (actual branch)	**626.774**
442-443	8.221e4
446-447	8.294e4
450-451	8.386e4

Potential Branch	Ψ
455-456	8.279e4
458-459	7.569e4
463-519	5.567e7
804-805	8.267e4
806-807	7.904e4
811-812	7.777e4
816-819	8.355e4

From Table 3.10, it can be seen that the search method identifies the correct branch (branch 419-422). Ψ for this simulated correction is 626.774, which is much lower than the other possible branches and is much lower value than the original Ψ where the topology is present (8.395e5).

These results reiterate the effectiveness of the localization and identification steps for load transfers across multiple large realistic feeders.

3.5 Conclusions

In this chapter, we address the limitations of use of traditional BDD methods for topology error monitoring. Development of our method follows two steps: 1) Improvement of C-DSSE and BDD through a novel approach to modeling the measurement uncertainties more accurately, and 2) Improve the detection and localization through signatures from the normalized residual spread-based approach.

The method proposed is implemented on several example cases of both capacitor failure and load transfers. Test results on realistic feeders that are based on real feeder data show the effectiveness of the proposed method and the scalability to large feeder sizes that are expected in practice. Across the comprehensive set of examples of the localization and identification steps for capacitor failure cases and load transfer cases, it is clear that the advantage of the localization step is the reduction of the search surface area and the identification step shows easily identifies the source as there is improvement in the sum of squared normalized residuals between the topology error case and identified correction in the order of magnitude of 10^2 10^3 in all cases, showing a clear distinction in the correct system topology.

The main contributions of this chapter are:

- Proposal of a new approach to modeling of measurement uncertainty meant to improve the overall quality/accuracy of C-DSSE estimates.
- Verification of normalized residual signatures through numerical analysis of the normalized residual spread.
- Proposal of a three step method that uses the signatures in normalized residuals from BDD to 1) Detect the presence and type of topology error, 2) Localize the detected topology error to a specific smaller portion of the system called measurement islands, and 3) Identify the specific source of the detected topology error.

CHAPTER

4

DATA ANALYSIS METHODS TO IDENTIFY PARAMETERS THAT IMPACT PRACTICAL REAL-TIME MONITORING PERFORMANCE

A Distribution System State Estimator is now being used by many utilities to provide the situational awareness required for applications like voltage monitoring and control [48]. Power flow-based distribution system state estimation is one of the techniques used in practice [49]. In this work, we will also refer to it as near real-time power flow (RTPF). In practice, RTPF is integrated with a

Distribution Management System (DMS). To operate the distribution system safely and effectively, accurate RTPF data is essential.

Only a small number of real measurements are available in distribution systems. A method of load estimating is therefore required [31]. Bus Load Allocation (BLA), a method of load estimation, utilizes the limited measurements, estimates the allocation of load from these measurements, and then passes these load estimates to power flow.

A common issue that affects the performance of RTPF is errors in the distribution system model used in the power flow program. Most commercial RTPF programs use detailed three phase circuit models. However, any error in the models of lines or other devices will have an impact on the accuracy of the estimated node voltages. Because of the large amount of data involved it becomes very challenging to detect and identify factors that contribute to the issues when the performance of the RTPF is not satisfactory [50, 51]. This is the problem we focused on in this work.

There is no literature directly addressing this issue in practice that we could identify. State estimation has been used mostly in research and publications to identify and detect bad data [1, 52, 53]. Given that there are a few real-time measurements, the methods utilized for bad data identification will not be very effective for this issue [54]. Furthermore, because bad data detection only considers one picture or sample, it makes it more difficult to spot long-term consequences or patterns of behavior.

This was work done as part of an industry sponsored project in collaboration with Duke Energy and the University of North Carolina at Charlotte (UNCC) through the facilitation from the Center for Advanced Power Engineering Research (CAPER) [55]. From feedback from Duke Energy engineers, we focused on the performance of RTPF voltage estimation accuracy. Duke Energy engineers have observed that while some feeders have low voltage mismatches (calculated as measured voltage – RTPF calculated voltage), there are some feeders which have large voltage mismatch on a consistent basis. We will call these feeders large voltage mismatch feeders. It is common to assume that the voltage mismatch on the feeder should be within $\pm 2V$ on a 120V base and anything outside of this range is considered large voltage mismatch. Existing methods of analysis regularly fail to determine the cause of the inaccuracy due to the size and complexity of the distribution system model, and the time-consuming nature of circuit-by-circuit analysis. Thus, this chapter focuses on this practical application of the signature-based methodology for identifying the factors that contribute to poor voltage estimation accuracy performance and to identify suspect cause(s) on a given feeder.

4.1 RTPF Algorithm

The DMS RTPF algorithm is explained to provide insights on the components that have possible impact the RTPF solution and ultimately the voltage mismatch. The procedure begins with a load estimation procedure called bus load allocation (BLA). BLA uses real-time real and reactive power (kW and kVAR) measurements from SCADA to scale the distribution system loads. To increase the accuracy of the allocated load, measurement islands can be formed from the real-time kW and kVAR branch flow measurements, where BLA can then allocate load within the island from the measured demand of that given measurement island. As a result loads are adjusted within the measurement island and do not spread to the other measurement islands.

Within each measurement island, BLA initialization begins with the system model and topology and initial load is selected as either the nominal loads or the loads of the previous RTPF solution. The load values are then iteratively adjusted by scaling them based on the real-time SCADA measurements. Additionally, the system topology can be adjusted, such as locally controlled voltage regulator estimated tap position and locally controlled capacitor switch status. In each iteration, power flow across the entire feeder is executed to estimate the observed power flow along the feeder given the current iterations adjusted load. The mismatch between such estimated power flow and the measured power flow is calculated and compared against a tolerance threshold to determine the acceptability of the load allocation. As a result, if the BLA mismatch is lower than the acceptable tolerance and the number of iterations has not exceed the max iterations allowed, the solution is said to be converged. Otherwise, the solution will be determined as suspect as BLA did not converge.

[56] Figure 4.1 illustrates the RTPF procedure flowchart.

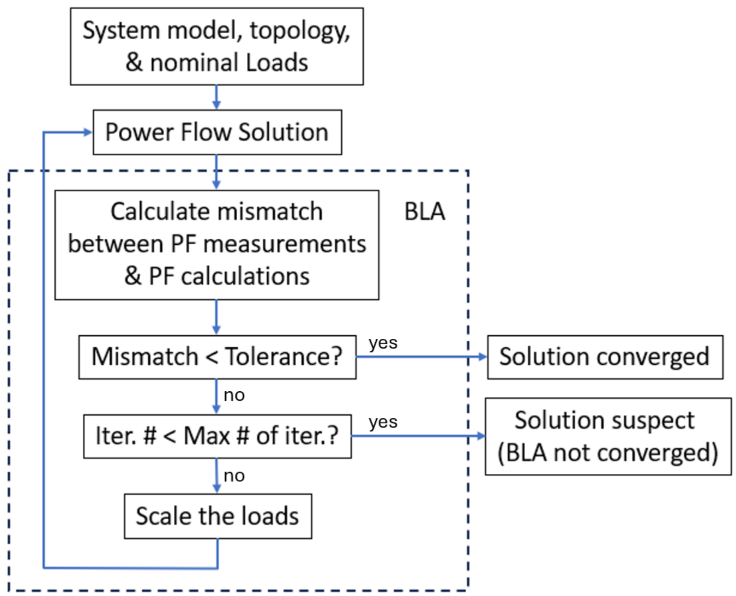

Figure 4.1 RTPF Algorithm Flowchart

4.2 Data Extraction

As seen in the RTPF, there is a significant amount of various inputs and parameters that can impact the solution and the voltage estimation accuracy. To conduct the proper analysis, a data gathering method must be used and proper parameters must be selected.

The data gathering process is described in Fig. 4.2. As the figure indicates, the first step involves performing the near real time power flow on a real substation model using save cases from the utility partner. The substation model comprises the system components and line connections, while the save case is a file which contains the system configuration, such as switch statuses, regulator tap positions, and capacitor statuses and available field measurements at a particular timestamp. To obtain the data for each save case, a python script has been developed. The python script eliminates errors, replaces missing data, removes duplicated data, ensures the data points are in the right format and calculates a set of additional parameters from the existing data.

Figure 4.2 Data Extraction Process

Table 4.1 shows the initial selected parameters for the analysis in consultation with the engineers of Duke Energy. The parameters can be classified into model and non-model based. The model-based parameters, such as total line length, number of transformers and number of capacitors, are independent of system power flow results, while the non-model-based parameters, such as capacitor injection and feeder head measurement, are those coming from field measurements or power flow results.

Table 4.1 Set of Input Parameters to Clustering Analysis

Model-Based Parameters	**Non-Model-Base Parameters**
1. Number of loads	20. Top of feeder meas. (kW, kVAR)
2. Number of load types (single phase, two phase, and three phase)	21. Total load (kW, kVAR)
3. Number of customer types (industrial, commercial, and residential)	22. Total industrial load (kW, kVAR)
4. Number of suspect loads	23. Total commercial load (kW, kVAR)
5. Number of transformers	24. Total residential load (kW, kVAR)
6. Number of unbanked transformers	25. Distribution Xfmr losses (kW, kVAR)
7. Total line length (ft.)	26. Line losses (kW, kVAR)
8. Number of feeder head voltage regulators(VR)	27. Cable and line capacitance (kVAR)
9. Primary Meter Size (kW, kVAR)	28. Three phase load (kW, kVAR)
10. Number of downstream VR	29. Two phase load (kW, kVAR)
11. Number of VR control type (local,remote,etc.)	30. Single phase load (kW, kVAR)
12. Number of capacitors	31. Feeder head power flow per feeder length
13. Number of capacitor control type (local,remote,etc.)	32. Avg. voltage drop between primary and secondary side
14. Number of measurement islands at the feeder head	33. Number of ON capacitors
15. Number of partially energized devices	34. Total capacitor injection (kVAR)
16. Number of looped lines	35. DER injection (kW)
17. Total overhead line length (ft.)	36. Measured voltage imbalance
18. Total underground cable length (ft.)	36. Amp flow imbalance
19. Number of capacitors in various control schemes (fixed, variable)	37. Number of suspect capacitors
	38. Number of suspect regulators
	39. Number of downstream meas. islands

4.3 Proposed Methodology

The proposed method has two parts. In the first part, we focus on identifying the factors that correlate to the differences in consistently large voltage mismatch feeders and the consistently small voltage mismatch feeders. To do so, the first part comprises of two steps. In the first step we use a clustering method to determine if we can separate cases having large voltage mismatch from those that have small mismatch based on the selected parameters. Another advantage of clustering in this step allows us to also identify large voltage mismatch feeders that are similar in parameters so that a common analysis can be applied across these feeders in a given cluster. In the second step, we use a binary logistic regression method to identify unique features for the clusters having large voltage mismatch feeders. These features indicate which predictors within the logistic regression model have a statistically significant impact on the likelihood of a binary outcome, which in other words, the contribute the most to the separation of large voltage mismatch clusters from the small voltage mismatch clusters.

In the second part, we focus on signature-based approach for diagnosing suspected cause(s). In the previous chapter, residual spread based analysis was observed and determined normalized residual spread was more appropriate in C-DSSE. In this RTPF, because of the use of measurement islands in BLA, BLA does not spread adjusted load between measurement islands. This ensures that residual does not spread between measurement islands, thus Residual Spread-Based Analysis is enough to determine the signatures and diagnose the suspected cause. As a result, we recognize that the different possible causes for large voltage mismatch have specific signatures in the measurement residuals that are present within the data and RTPF solution. In this part, we propose a guideline for analysis to identify these signatures in the measurement residuals which will assist in diagnosing suspected cause(s) of large voltage mismatch in specific feeders.

4.4 Part I: Identifying Unique Features Correlating to Large Voltage Mismatch

4.4.1 Step I: K-Means Clustering

Clustering is adopted in the first step to check whether the feeders with large voltage mismatch indeed have features that are different from the feeders that do not experience large voltage mismatch. Researchers have used k-means clustering for similar purposes [57]. K-means clustering is used. This method is an unsupervised learning algorithm that iteratively groups data points into the most likely group of data points given a set of features/parameters. It should be noted that a key advantage of clustering is that we do not use the a priori information on whether a feeder has large voltage mismatch or not, meaning that these clusters that are formed are not biased by this information. Thus, this clearly shows that the separation between large voltage mismatch and small voltage mismatch is impacted by the feature set.

Another advantage of clustering is the ability to group multiple feeders with large voltage mismatch that are similar together. This allows us to determine the common features these feeders exhibit, and thus help the engineers to identify other similar feeders. It can be very time consuming to analyze each feeder at a time and identify the unique features that separate each from those that have small voltage mismatch, thus this provides an improvement in the computational and time requirement to complete the analysis.

For K-means clustering, the number of k clusters is a hyper-parameter that must be tuned for optimal performance. As K-means clustering iteratively groups data points into clusters, it divides the whole sample into increasing number of clusters and repeats up to the maximum number of k clusters. Given the different cluster models with different k number of clusters, we must select the most optimal one. To do so, we have adopted two metrics in order to determine the most optimal k. These metrics are sum of square error (SSE) and silhouette score:

1. **Sum of Square Error (SSE)**: SSE finds the sum of the squares of the Euclidean distance of a data points within a given cluster to that cluster's center [58]. SSE can be calculated using:

$$SSE = \sum_{i=1}^{k} \sum_{x_j \epsilon S_i} ||\mathbf{x}_j - \mathbf{u}_i||^2 \tag{4.1}$$

where, $\mathbf{x}_j$ is the j^{th} data point within the i^{th} cluster and $\mathbf{u}_i$ is the center of the i^{th} cluster. SSE therefore measures how spread out the data points within a given cluster are and therefore a smaller SSE indicates a well fit cluster. Often, the relation between k and SSE creates an elbow plot where the optimal k selection is at the elbow point.

2. **Silhouette Score**: Silhouette score is often referred to as the "goodness of fit score" of clustering. This metric considers both how well the clusters are grouped within themselves and the distance between one cluster and the others around it [59]. Fig. 4.3 displays an example of the silhouette score visually.

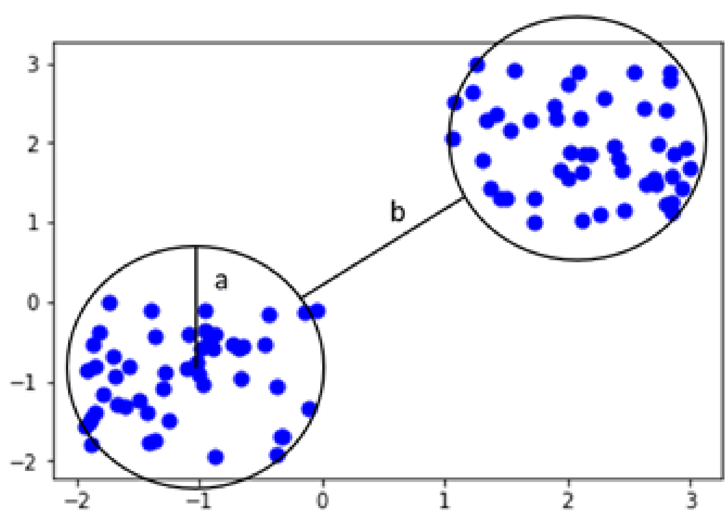

Figure 4.3 Illustration of Silhouette Score

Numerically, silhouette score can be defined by:

$$Silhouette\ score = \frac{b-a}{max(a,b)} \tag{4.2}$$

Where b is the average inter-cluster distance and a is the average intra-cluster distance, as shown in Fig. 2. Analytically, this score can range from -1 to 1, with the most optimal performance closest to 1. This is because a score of 1 would result from a large inter-cluster distance, b, and a small intra-cluster distance, a. From these two metrics, an optimal model can be selected from the conjunction of the two metrics that can be used for further analysis of large voltage mismatch and small voltage mismatch feeders.

Finally, once the clusters are obtained, we can "label" each cluster as a large voltage mismatch or small voltage mismatch cluster based on a simple majority rule. I.e. if majority of a cluster is filled with large voltage mismatch cases, we will call this cluster a large voltage mismatch cluster. These definitions do not impact the overall performance of the clustering as clustering does not use a priori information like this, however it will be used in the logistic regression step.

4.4.2 Step II: Logistic Regression

Binary Logistic Regression is a supervised machine learning technique utilized for the purpose of binary classification. It enables the separation of discrete values, such as Yes/No or Pass/Fail, within a given set of observations. Logistic regression involves the fitting of a regression curve, represented as y = f(x), to a sample of observations. In this context, the variable y represents a categorical variable, such as Yes/No or Pass/Fail. The approach is visually depicted in Fig. 4.4.

Logistic regression is a classification algorithm that predicts a binary outcome based on a series of independent variables. Another main output of the logistic regression method is that it can pinpoint which of the parameters used are the so-called significant parameters; these are the parameters that contribute the most to the estimated outcome. In our case, we used binary logistic regression

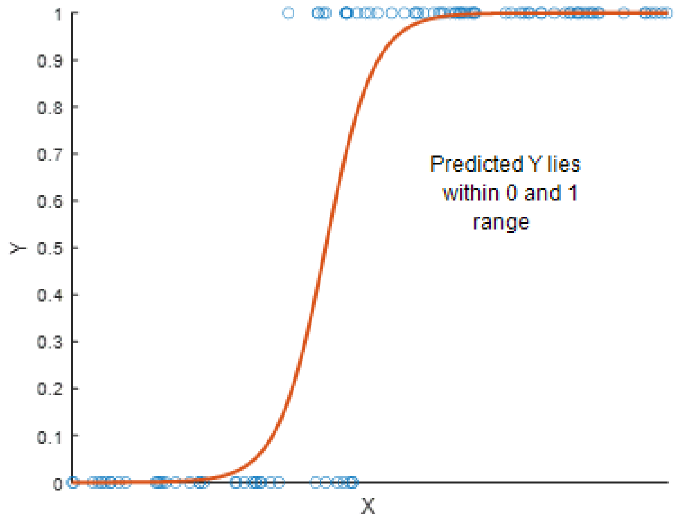

Figure 4.4 Binary Logistic Regression Model

to extract the significant parameters rather than predicting the outcome although it is frequently used for prediction of binary outcome with given independent input variables.

In the Logistic Regression model, we need to choose a logistic function [60]. The standard logistic function, for predicting the outcome of an observation given a predictor variable (X), is an s-shaped curve defined as:

$$p(X) = \frac{e^y}{1+e^y} \tag{4.3}$$

Where $p(X)$ is the probability of the event to occur given X, and $y = \beta_0 + \beta_1 X$. To fit the model (4.3), we use a method called maximum likelihood [61].

When we consider the problem of predicting a binary response using multiple predictors, the logistic function looks like:

$$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{4.4}$$

The coefficients β_1, β_2, ..., β_n are are unknown and need to be estimated based on the available data. The goal is to find estimates for β_1, β_2, ..., β_n such that plugging these estimates into the model for p(X) yield values close to 1 for individuals having low voltage mismatch, and values close to 0 for individuals having high voltage mismatch. Typically, a probability threshold of 0.5 is chosen. Based on whether the probability of an element exceeds the threshold or not, it is classified into one category or the other.

Running a regression model with too many variables, especially with irrelevant ones will lead to a needlessly complex model. In many cases, several variables used in a multiple logistic regression model may not actually have any association with the response variable. Incorporating such irrelevant variables introduces unnecessary complexity into the resulting model. Additionally, an excessive number of variables can lead to overfitting and high variance in the coefficient estimates for logistic regression. Hence, step-wise methods offer alternatives to the best subset selection [61], as they explore a more restricted set of models and facilitate the creation of a more interpretable model.

The process begins with a regression model which includes all candidate variables. Then, variables are gradually eliminated from the model until the remaining variables satisfy certain criteria. At each step, the variable that contributes the least improvement to the model is removed.

One criterion to select the optimal set of features is Akaike Information Criterion (AIC) [62]. The AIC is:

$$AIC = -2logL + 2k \tag{4.5}$$

Where L is the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate, and k is the number of estimated parameters in the model.

This value is calculated for every candidate model after each variable elimination on the logistic regression model and the best model is the model with the smallest AIC [63]. AIC estimates the quality of each model relative to the other models and provides a single number score that can be used to determine which of the models is the best fit to the data.

As a result, the logistic regression model will identify significant parameters that we will call "unique features." These unique features identified are the variables that contribute the most influence on the predicted binary outcome. To do so, the statistical significance is determined by using the predictors' p-value. The p-value, which is calculated by comparing the estimated coefficient of the predictor to its standard error, tests the null hypothesis that a predictor's has no effect on the binary outcome. A p-value smaller than a predetermined significance level or confidence suggest that there is a strong likelihood that this parameter has a significant impact on the outcome. Thus, these significant parameters are critical in the logistic regression model in determining the difference between large voltage mismatch and small voltage mismatch, hence determining the unique features of the large voltage mismatch feeders.

4.5 Part II: Identifying Cause for Large Voltage Mismatch

At this point in the methodology, the goal is to directly pinpoint the suspected cause(s) for a given feeder to be experiencing a large voltage mismatch. In the previous part, we identify unique features correlating to the feeders who consistently experience large voltage mismatch. However, these unique features are not directly the cause for the large voltage mismatch.

Leveraging the unique features identified in the previous part as a guide, further examination must be conducted. To begin, we identify potential causes within the model, the topology, and the RTPF algorithm that have the highest impact on the estimated voltage solution. Those potential causes are as follows:

1. **Bus Load Allocation (BLA) Convergence:** As a preliminary step in RTPF to distribution power flow (DPF), a correctly allocated load from BLA is critical for accurate voltage estimates. If BLA does not converge, then it is clear that load is not allocated properly and can lead to disparities in the calculated voltages.

2. **Feeder Head Voltage Mismatch:** In traditional RTPF, it is common to use the feeder head

as the reference node where the voltage is known and constant. However, in DMS RTPF the reference node is selected as the high side of the substation transformer at the interconnection between the transmission system and the substation, which will be called the interface bus. Fig. 4.5 shows an example of the substation model with the interface bus. It is common for

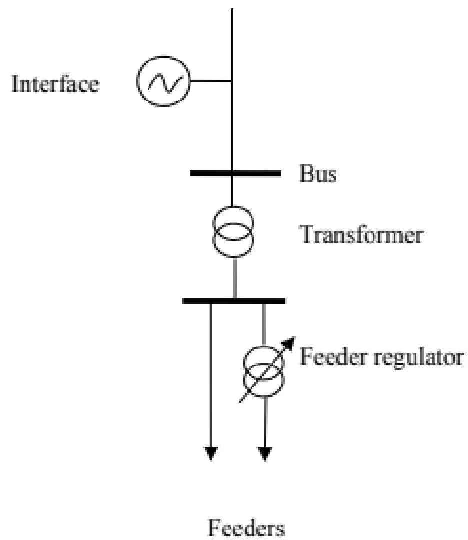

Figure 4.5 Substation Model with Interface Node

utilities to place meters on the load side of the substation and not on the high side, thus this interface bus voltage must be calculated through a method called back-calculation [56]. This back calculation method utilizes the measured power flow and substation model including transformer impedance and transformer and voltage regulator tap positions to calculate the interface voltage. Once complete, it is used as the reference voltage and the feeder head voltage is calculated along with the other system voltages. If there are any disparities in the substation model or even BLA solution, it will cause the voltage mismatch at the feeder head to be larger than the acceptable ±2V range. If a feeder converges in BLA but shares a station bus with a feeder that does not converge, this may also lead to large voltage mismatch at the feeder head.

This large voltage mismatch at the feeder head will propagate downstream and cause the other voltage measurements to be large, thus is a significant cause to voltage mismatch on the feeder.

3. **DER Status:** It is common for utilities to place DERs like photovoltaics (PVs) in their distribution networks and effort must be made to model them correctly to ensure accurate RTPF solution. However, it is not always the case where the DER penetration is not modeled correctly causing a disparity in the estimated demand on the feeder. As a result, voltage drop will be impacted and voltage estimates will experience a large mismatch.

4. **Primary Meter Loading:** For utilities, there are some customers that require a large demand that is easier to monitor by placing a real-time meter. As it is a measured load, BLA does not adjust it and allocate load here. However, it is possible to incorrectly model this load and if there is a error in the modeling of the primary meter, it could lead to a disparity in estimated demand at that point in the system, which can have an impact on voltage calculations.

5. **Capacitor Status:** In most cases, capacitor switch status is monitored through SCADA. However, in some control schemes, RTPF must estimate the switch status of locally controlled capacitor banks. Additionally, there is a case where the capacitor bank fails and the internal fuse disconnects the capacitor from the feeder but the switch status will still show the capacitor as connected [8, 9]. In these cases, capacitor status can be wrong and cause an incorrect calculation of reactive power injection from the capacitor bank, which has a significant impact on voltage calculation.

6. **Voltage Regulator Tap Status:** In RTPF, voltage regulator tap position is an estimated value if there is no available tap postion measurement from SCADA. Thus, if this tap position in the voltage regulator model is incorrect, there will be a disparity between the actual regulated voltage in the system and the estimated voltage from the RTPF solution.

7. **Line Modeling:** Line modeling is a critical need for accurate voltage estimation in RTPF. If the

modeling of line parameters is incorrect, this can lead to incorrect voltage drop calculations across the lines as the RTPF is a nodal admittance matrix-based power flow, hence leading to large mismatches in the voltages.

8. **Distribution Transformer Losses:** Customer distribution transformers losses must be modeled correctly for accurate voltage estimation. If there is a disparity in the no load losses of the modeled customer transformers, this can lead to a widespread error in the estimated demand on the system, commonly leading to non-convergence in BLA.

9. **Load Modeling:** Load modeling can also highly impact voltage estimation. If there are any disparities in the modeling of the load being constant power, constant impedance, constant current, or any combination, it can lead to an error in the demand causing a large mismatch in voltage estimation.

Considering the causes identified above, we focus on the first six as they are most common in practice. We propose a systematic guideline-based approach designed to expedite the process of identifying particular causes for large voltage mismatch given any new feeder exhibiting similar issues. This approach is intended to assist utilities in diagnosing the cause(s) that a feeder is experiencing large voltage mismatch so that the utility may be able to resolve any issue to improve the accuracy of their real-time monitoring. The guidelines utilize specific signatures in the measurement residuals that exist with each cause to further identify the specific suspected cause. The systematic guideline steps are as follows:

1. **Check BLA convergence:** Ensure that BLA converges. Continue to next step if converging, correct otherwise.

2. **Check voltage mismatch at feeder head:** Ensure that voltage mismatch at the feeder head is acceptable. Feeder head voltage mismatch can be caused by:

    i) Interface voltage back-calculation method failure.

ii) Mismatch between calculated and measured power flow (on a station basis if feeders share a station bus)

iii) Incorrect station regulator tap position

After identifying the cause for feeder head voltage mismatch and correcting, if downstream voltage mismatch does not resolve, continue to next step to downstream component evaluation.

3. **Use BLA convergence results and power flow measurement residuals to: a) determine suspect cause and b) pinpoint suspected cause:** Signatures for each of the possible downstream component causes exists in the mismatches of the feeder power flow measurements. First, validate all power flow meters that will be used in the analysis by checking: 1) validity status as determined by SCADA, 2) power factor, 3) polarity, and 4) phasing.

   i) If large residual exists in real power: suspect DER or missing load (primary meter or incorrect tie switch status).

   ii) If large residual exists in reactive power: suspect capacitor status issue or missing reactor problem.

   iii) If all voltage measurements downstream of a voltage regulator have large measurement residual: suspect voltage regulator tap position issue.

   In the second part of this step, to pinpoint suspected cause, modify the model of the suspected issue and verify improvement in voltage mismatches. If large voltage mismatches exist continue to next step

4. **Check system modeling:** Validate line modeling, distribution transformer loss modeling, and load modeling.

As a systematic step by step approach, the process of diagnosing and correcting the possible system topology and modeling issues becomes streamlined and the efficiency to intervene to resolve these issues improve.

4.6 Case Study

Through collaboration with Duke Energy and UNC at Charlotte, work has been conducted to identify parameters that impact the performance of an established distribution management system (DMS) within Duke Energy. The cases under consideration are actual distribution networks within the Duke Energy's jurisdiction. Again, the criteria used for performance evaluation within DMS is voltage mismatch and is taken as the difference between measured voltage and the calculated voltage at measurement points resulting from the load flow analysis. For voltage measurements, it is expected the mismatch is within 2V. If there are any mismatches outside of these expectations, the case is determined to have large voltage mismatch and be poorly performing.

Sample data includes 340 RTPF samples (save cases) from the sponsoring utility. The data used for the analysis consists of near real-time power flow results from DMS (save cases). We have used consistently large and small voltage mismatch feeders for our analysis. This includes 14 feeders that have small voltage mismatch and 10 feeders that have large voltage mismatch. For each feeder there are about 10-15 samples at different days and times

4.6.1 Identifying Unique Features of Large Voltage Mismatch Feeders

4.6.1.1 K-means Clustering to Separate Large Voltage Mismatch Feeders

K-means clustering was performed in the first step using the described sample. Parameter k is iterated from 2 to 21 and the resulting SSE and silhouette scores are recorded at each iteration. The resulting plots are displayed in Fig. 4.6 and Fig. 4.7

From the plots of the two metrics, it is decided that the optimal k is the 11 cluster model. We can see that the elbow point in the SSE curve is around when k is 6 or 7, however the rate of decrease in SSE is still relatively big until k is 11 when the curve begins to level out. From Fig. 4.7, silhouette score seems to increase as the number of clusters increases. As mentioned, the most optimal silhouette score is closest to 1, however, as we split more clusters, the clusters begin to only contain single

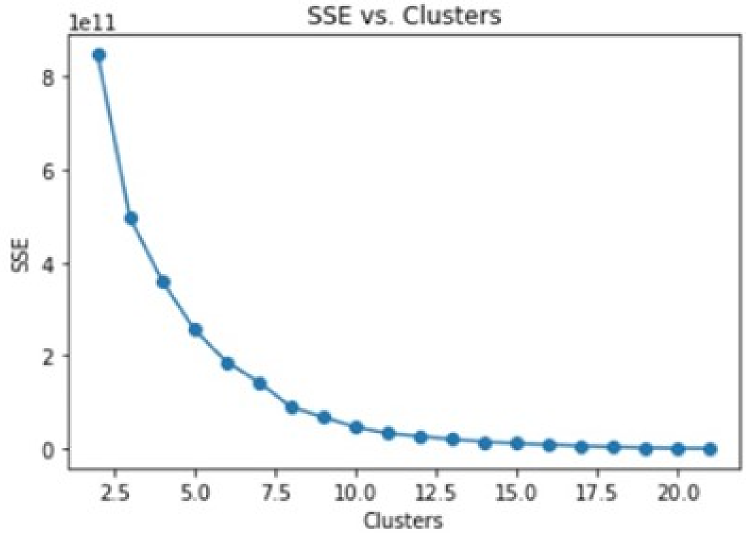

Figure 4.6 *k* vs SSE Curve

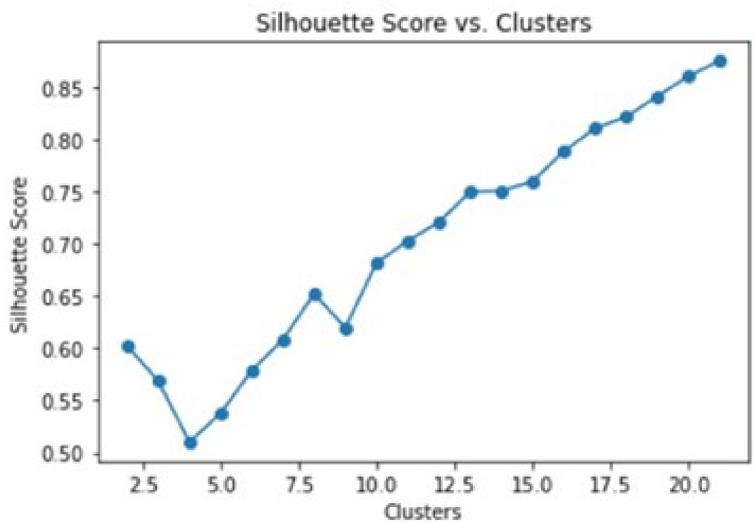

Figure 4.7 *k* vs Silhouette Score Plot

feeders, which eliminates the secondary advantage of the clustering approach that is previously discussed. Thus, combining the selection process of the two metrics, 11 clusters is selected as it has a low SSE and an acceptable silhouette score.

Table 4.2 breaks down the contents of each cluster and its label that is determined by simple

Table 4.2 Cluster Breakdown

		# of Observations	
Cluster Number	Cluster Label	Large Voltage Mismatch	Small Voltage Mismatch
0	Small Mis.	0	24
1	Small Mis.	0	12
2	Large Mis.	40	3
3	Small Mis.	1	56
4	Small Mis.	4	25
5	Large Mis.	15	0
6	Large Mis.	27	3
7	Large Mis.	15	0
8	Small Mis.	16	40
9	Small Mis.	1	14
10	Large Mis.	32	12

majority rule.

The selected clustering model achieves separation between large mismatch and small mismatch cases as seen in Table 4.2. For a majority of the clusters, they are clearly a small mismatch cluster or clearly a large mismatch cluster. Cluster 2 contains three different large voltage mismatch feeders, cluster 6 contains two different large voltage mismatch feeders and clusters 5 and 7 each contain a single large mismatch feeder. This is significant as it verifies the key advantage of using clustering to show that indeed feeders with large voltage mismatch have features that are different than feeders with small voltage mismatch. Also, the clusters with multiple large voltage mismatch feeders show the secondary advantage where any features identified can be applied to any of the feeders within the cluster.

It should be noted that though the majority of the clusters are well separated, two clusters appear to have some mix and might require further analysis. Cluster 8 contains two feeders that have small voltage mismatch and one feeder that has large voltage mismatch. This will require further analysis into the reason that this large voltage mismatch feeder is grouped with small mismatch feeders and what separates them. Cluster 10 contains one feeder with small voltage mismatch feeders and two large voltage mismatch feeders thus is determined as a large voltage mismatch cluster.

4.6.1.2 Cluster-Based Logistic Regression to Identify Unique Features

Next, logistic regression is used on a cluster-by-cluster basis to identify the unique features that correlate to the separation between large mismatch and small mismatch clusters. Hence, we choose a selected large voltage mismatch cluster and use all of the small voltage mismatch clusters to identify the unique features associated with this cluster. The set of data is split data into training and testing data based upon a 70%/30% split respectively. The logistic regression model is trained on the training data set and the coefficients of the model are analyzed for the unique features. If the coefficient of a certain feature is reported to have a statistically significant relationship to the voltage mismatch classification, we determine these as the unique features as it has statistical significance towards the separation between large voltage mismatch and small voltage mismatch. For an example of this analysis, we consider two select clusters, cluster 2 and the further analysis into cluster 8. Table 4.3illustrates the logistic regression model for cluster 2.

From Table 4.3, we can see that of the features selected by the model selection process, four of them can be determined as statistically significant given a significance level, $\alpha = 0.01$: distribution transformer kVAR losses, top of feeder kVAR flow measurement, total kVAR load, and total capacitor kVAR injection. These four features have a strong statistical relationship to separation between the large voltage mismatch and small voltage mismatch for the feeders within cluster 2 specifically. For example, the average value for distribution transformer kVAR losses for small voltage mismatch clusters is 272.863 kVAR and 162.012 kVAR for large mismatch cluster 2 but for total kVAR load the average values are 636.347 kVAR and 882.836 kVAR for small and large mismatch respectively. These values show the difference between the two clusters and can be used further in identifying the cause of large voltage mismatch.

As mentioned, cluster 8 would require further analysis as a large voltage mismatch feeder is

Table 4.3 Cluster 2 Logistic Regression Model

Feature	Estimate	P-Value	Significance ($\alpha = 0.01$)
Distribution Xfmr Losses (kVAR)	106.1715	4.21e-05	Statistically Significant
Cable & Line Shunt Cap Injection (kVAR)	-1.74041	0.039	Not Significant
Top of Feeder Measurement (kVAR)	3.8148	0.00355	Statistically Significant
Total Load (kVAR)	-36.9240	1.24e-05	Statistically Significant
Capacitor Injection (kVAR)	-3.9605	0.003366	Statistically Significant
Number of Suspect Capacitors	16.8011	0.993660	Not Significant
Number of Suspect Regulators	19.7582	0.992021	Not Significant
Single Phase Load (kVAR)	-6.1803	0.029685	Not Significant

mixed within a cluster with two small voltage mismatch feeders. For this analysis we isolate the large mismatch feeder in this cluster and train a logistic regression model on its data with the small mismatch feeder data of the cluster. The resulting logistic regression model is shown in Table 4.4.

From Table 4.4, we can see that three of the features are identified as having a statistically significant relationship with the separation between the large voltage mismatch feeder and the small voltage mismatch feeders of cluster 8. Distribution transformer kVAR losses, cable and line shunt capacitance injection, and total kVAR load are identified as unique features and can be further used in identifying the cause of the large mismatch in this feeder in cluster 8. For example, the average value for total kVAR load for small voltage mismatch cases in this cluster is 376.207 kVAR and 1789.769 kVAR for the large voltage mismatch feeder. The impact of total kVAR load on the voltage mismatch can be analyzed further.

The above analysis is done for all 5 large mismatch clusters and cluster 8 further analysis. Table 4.5 provides the various significant parameters that are identified for all clusters and the average

Table 4.4 Cluster 8 Further Analysis Logistic Regression Model

Feature	Estimate	P-Value	Significance ($\alpha = 0.01$)
Distribution Xfmr Losses (kVAR)	3169.663	0.00326	Statistically Significant
Cable & Line Shunt Cap Injection (kVAR)	-7.339	0.00278	Statistically Significant
Top of Feeder Measurement (kVAR)	5.889	0.01163	Not Significant
Total Load (kVAR)	-77.459	0.00159	Statistically Significant
Number of Suspect Regulators	23.035	0.99228	Not Significant

normalized values for each cluster are shown and compared to the average value of the small mismatch cases. Cells of the table indicated with "**" refer to the unique feature of the cluster as determined by the significant parameters.

As seen in the table, each cluster has different significant parameters as determined by training the logistic regression model for each. This shows that these features are unique to the cluster and the feeders contained within them. Additionally using these unique features for each cluster and observing how they relate to the cases of small voltage mismatch feeders (i.e. larger or smaller average value) does not distinctly pinpoint the cause for the large voltage mismatch feeder. Thus, further analysis is needed for identifying the cause(s).

4.6.2 Signature-Based Analysis for Identifying Cause

In this step, we use the Signature-Based Analysis to further examine the feeders and determine a suspected cause. Leveraging the list of possible causes, we search the measurement residuals for signatures and use the proposed guidelines to determine the suspected cause. For a comprehensive study, examples of several suspected causes are provided below. We will exclude the step regarding BLA convergence as a cause and only regard cases where BLA converges, as determining the cause for BLA convergence requires its own analysis and procedure.

Table 4.5 Cluster Unique Feature Average Values

Unique Feature	Small Mis. Avg.	Cluster 2 Avg.	Cluster 5 Avg.	Cluster 6 Avg.	Cluster 7 Avg.	Cluster 10 Avg.	Cluster 8 Avg.(Large Mis. Only)
Distribution Xfmr Losses (kVAR)	0.0409	0.0396**	0.0402**	0.0384	0.363**	0.0406**	0.0389**
Top of Feeder Measurement (kW)	0.4214	0.4038	-0.5752	0.2996	0.2940	0.5032**	0.6054
Top of Feeder Measurement (kVAR)	0.0839	-0.1156**	1.1364	-0.1825	-0.2834	-0.1863	0.091
Total Load (kVAR)	0.0951	0.1857**	0.0372	0.0733	0.0617	0.199	0.1599**
Capacitor Injection (kVAR)	-0.3754	-0.6357**	-0.7381	-0.4827	-0.4960	-0.6428**	-0.1875
Cable & Line Shunt Cap Injection (kVAR)	-0.5591	-0.5798	-0.5054**	-0.5608**	-0.3445	-0.4098	-0.4139**
Number of Suspect Capacitor	0.1397	0	1**	0	2**	1.0227**	0
Total UG Cable Length (ft.)	0.4377	0.5080	0.8515**	1.0257**	0.4884**	0.4184**	0.2605
Total OH Line Length (ft.)	0.8506	0.8431	3.3323**	1.6299	2.2960**	0.5439	0.6979

4.6.2.1 Feeder VE42

The approach was implemented on this example feeder VE42. We have verified that BLA converges in all samples and move to the next step examining the voltage mismatch at the feeder head. Table 4.6 provides the voltage mismatch at the feeder head and voltage mismatch downstream for each sample.

Table 4.6 Feeder VE42 Voltage Mismatches

	Feeder Head Voltage Mismatch			Downstream Voltage Mismatch		
Sample Time	Phase A Vmismatch (V)	Phase B Vmismatch (V)	Phase C Vmismatch (V)	Phase A Max Vmismatch(V)	Phase B Max Vmismatch(V)	Phase C Max Vmismatch(V)
1.49	4.393168	4.706504	4.943168	4.710035	–	–
1.52	4.39984	5.423168	5.80984	4.535465	–	–
12.15	3.34984	3.943168	4.106504	3.740936	–	–
12.19	5.956504	6.35984	7.066504	6.527744	–	–
12.55	5.043168	5.186504	6.153168	5.345738	–	–
3.34	5.053168	5.246504	6.32984	5.235485	–	–
4.06	6.25984	6.673168	7.223168	6.722753	–	–
4.244	4.163168	4.80984	4.796504	4.442423	–	–
4.24	3.223168	4.413168	4.493168	3.555876	–	–
5	4.633168	5.38984	5.346504	5.031377	–	–
8.02	5.406504	6.023168	6.223168	5.894557	–	–
8.16	1.98984	3.24984	3.126504	3.555876	–	–
8.35	5.623168	6.016504	6.41984	6.066399	–	–
9.14	4.536504	5.756504	5.956504	4.995997	–	–
9.31	3.2665	3.8065	3.8665	3.60725	–	–

From Table 4.6, it is apparent that every sample has both a large voltage mismatch at the feeder head and downstream on the feeder as every case has voltage mismatch at the feeder head and downstream outside of the acceptable ±2V. Hence, it is possible that the large voltage mismatch on the feeder is caused by the voltage mismatch at the feeder head. Further verification is needed.

As a result the substation model is examined and shown in Fig. 4.8.

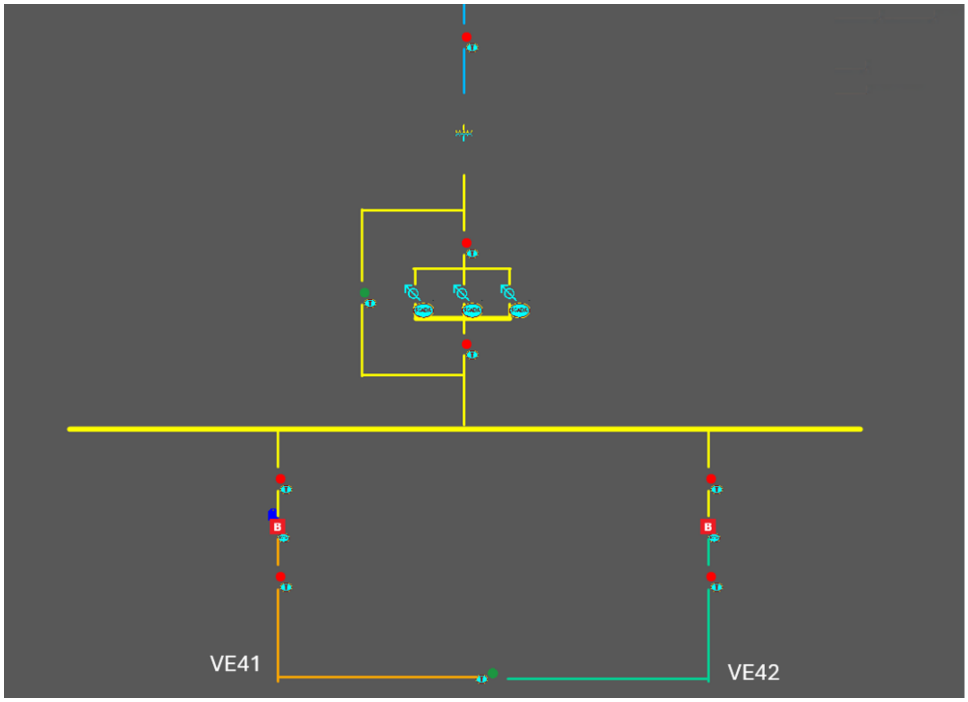

Figure 4.8 VE42 Substation Model

From Fig. 4.8, it can be seen that included the substation model is the substation transformer, followed by the substation voltage regulators, and a substation bus that is shared by two feeders, VE41 and VE42. As mentioned previously, the feeder head voltage (voltage at the station bus) can have mismatch if there is any error in the back-calculation method that obtains the interface bus voltage (high side of the substation transformer). BLA mismatch is one such cause, but VE42 has already been confirmed to accomplish convergence for every sample. However, as it shares a station bus with VE41, the BLA mismatch of VE41 also impacts this shared feeder head voltage. We further confirm that VE41 has large BLA mismatch for all samples, thus creating the large voltage mismatch at the feeder head for both feeders.

We can further verify that feeder head voltage mismatch is the primary suspected cause for large voltage mismatch downstream by forcing the voltage mismatch at the feeder head to be small. For one sample (sample 4.24) we override the control of the substation voltage regulator and force the voltage mismatch to a small magnitude and observe the resulting downstream max voltage mismatch. The results are shown in Table 4.7 and prove that feeder head voltage mismatch is the primary cause for large voltage mismatch on the feeder as indicated by the approach.

Table 4.7 Feeder VE42 Correction Voltage Mismatches

	Feeder Head Voltage Mismatch			Downstream Voltage Mismatch		
Case	Phase A Vmismatch (V)	Phase B Vmismatch (V)	Phase C Vmismatch (V)	Phase A Max Vmismatch(V)	Phase B Max Vmismatch(V)	Phase C Max Vmismatch(V)
4.24 Original	3.22	4.41	4.49	3.56	–	–
4.24 Corrected	0.07	0.059	0.22	0.30	–	–

4.6.2.2 Feeder FI43

This approach was applied to the feeder FI43. Beginning the procedure, we confirm that the BLA converges and begin by investigating the feeder head voltage mismatch as a potential cause. Table 4.8 provides the voltage mismatches at both the feeder head and downstream on the feeder for all samples available.

Table 4.8 Feeder FI43 Voltage Mismatches

	Feeder Head Voltage Mismatch			Downstream Voltage Mismatch		
Sample Time	Phase A Vmismatch (V)	Phase B Vmismatch (V)	Phase C Vmismatch (V)	Phase A Max Vmismatch(V)	Phase B Max Vmismatch(V)	Phase C Max Vmismatch(V)
1.49	-0.86016	0.20984	0.713168	-6.69	-2.37	-2.72
1.52	-0.55016	0.02984	0.87984	-7.31	-2.68	-2.7
12.15	-0.54683	-0.0035	1.003168	-7.33	-2.68	-2.56
12.55	0.376504	1.336504	1.703168	-6.86	-1.37	-1.5
3.34	-0.4835	0.223168	0.726504	-6.54	-1.70	-2.05
4.244	-0.74016	0.083168	0.91984	-6.93	-2.66	-2.88
4.24	-0.6535	-0.0235	1.066504	-6.68	-2.12	-2.47
5	-0.7635	0.116504	0.876504	-6.78	-2.71	-3.37
8.02	-0.57683	0.003168	0.916504	-6.84	-2.1	-2.65
8.16	-0.5135	0.366504	0.556504	-6.54	-2.18	-2.85
8.35	-0.49683	0.013168	0.976504	-6.88	-2.54	-2.59
9.14	-0.43683	0.18984	0.916504	-6.35	-2.12	-2.69
9.31	-0.52683	-0.15683	0.926504	-7.09	-2.99	-2.82

From Table 4.8, feeder FI43 does not exhibit large voltage mismatch at the feeder head for any sample. However, for every sample there is large voltage mismatch outside of the acceptable ±2V. This indicates that the voltage mismatch at the feeder head cannot be the cause and we move on in the procedure to analyzing the signatures in measurement residuals to determine the cause that is downstream on the feeder.

The one-line diagram and measurements are shown in Fig. 4.9.

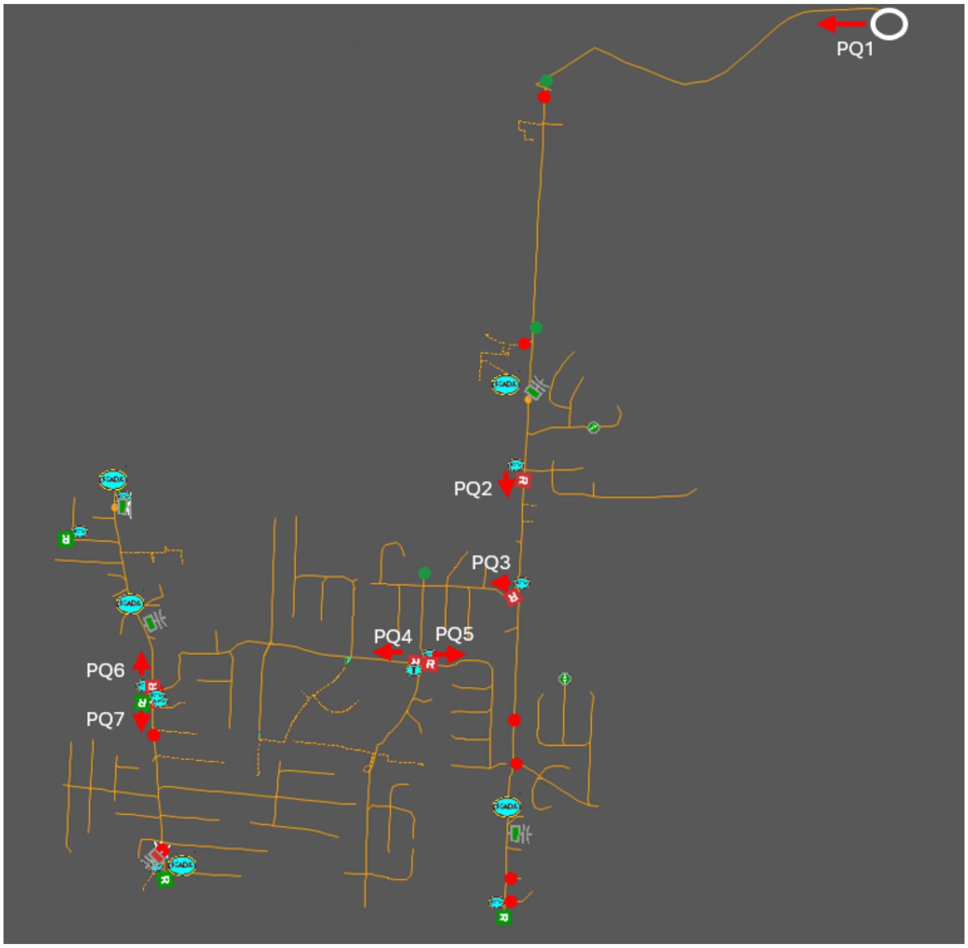

Figure 4.9 FI43 One-Line Diagram and Measurement Placement

In this one-line diagram shown in Fig. 4.9, the large white circle in the top of the figure represents the substation location. Each branch flow measurement is labeled and direction is indicated by the red arrows. Table 4.9 provides the measurement residuals of the branch flow measurements.

Table 4.9 Feeder FI43 Measurement Residuals

Meas. ID		Pa	Pb	Pc	Qa	Qb	Qc
PQ1	Meas.	660.01	630.01	690.01	-299.99	120.01	-289.99
	Calc.	659.3	630.01	687.04	-305.36	121.97	-279.03
	Mismatch	0.71	0	2.97	5.37	-1.96	-10.96
PQ2	Meas.	579	540	624	-329	99	-322
	Calc.	568.96	498.95	605.47	-329.58	6.42	-312.05
	Mismatch	10.04	41.05	18.53	0.58	92.58	-9.95
PQ3	Meas.	431	417	566	-353	90	-324
	Calc.	422.06	388.65	550.85	-344.98	-91.26	-313.94
	Mismatch	8.94	28.35	15.15	-8.02	181.26	-10.06
PQ4	Meas.	272	384	378	-392	55	-380
	Calc.	327.62	347.17	367.26	-360.53	-132.19	-348.93
	Mismatch	-55.62	36.83	10.74	-31.47	187.19	-31.07
PQ5	Meas.	96	–	–	20	–	–
	Calc.	94.44	–	–	14.06	–	–
	Mismatch	1.56	–	–	5.94	–	–
PQ6	Meas.	124	79	131	40	26	38
	Calc.	138.31	129.42	95.59	32.39	78.1	26.44
	Mismatch	-14.31	-50.42	35.41	7.61	-52.1	11.56
PQ7	Meas.	58	204	176	-458	22	-424
	Calc.	87.3	132.76	188.98	-419.29	-294.67	-392.05
	Mismatch	-29.3	71.24	-12.98	-38.71	316.67	-31.95

Analyzing the residuals in Table 4.9, it can be seen that there is a very large residual in reactive power measurements on phase B, thus indicating the signature of a capacitor error. With the largest residual seen at PQ7 and the measured value is more positive than the calculated reactive power, this indicates that the capacitor is disconnected on this phase downstream in the field, however is modeled as connected. This exemplifies the practicality of capacitor failures topology error and the challenging task to monitor, causing inaccuracies in the voltage estimates.

To further verify, we can simulate the capacitor downstream of PQ7 failing on phase B and observe any improvement in the voltage estimates. Table 4.10 provides the system max voltage mismatches on each phase in the original case and after the simulated correction.

Table 4.10 Feeder FI43 Correction Voltage Mismatches

Case	Va	Vb	Vc
9.48 Original	-6.69	-2.37	-2.72
9.48 Corrected	-6.72	-1.51	-2.76

From simulated correction, we see that the voltage mismatch on phase B improves to the acceptable range. Additionally, it should be noted that the mismatch on PQ7 for reactive power on phase B improves from 316.67kVAR to 1.05kVAR, thus proving this correction is necessary to resolve this topology error. However, because the simulated correction was only on phase B, only phase B voltage mismatch improves. Phases A and C require further analysis as there are no clear signatures indicated within the measurement residuals.

4.6.2.3 Feeder SA1273

SA1273 was also analyzed using the proposed approach. We confirm that BLA converges and in the first step of the procedure, begin with investigating the voltage mismatch at the feeder head as a possible cause. Table 4.11 provides the voltage mismatches at the feeder head and downstream for each sample available.

Table 4.11 Feeder SA1273 Voltage Mismatches

	Feeder Head Voltage Mismatch			Downstream Voltage Mismatch		
Sample Time	Phase A Vmismatch (V)	Phase B Vmismatch (V)	Phase C Vmismatch (V)	Phase A Max Vmismatch(V)	Phase B Max Vmismatch(V)	Phase C Max Vmismatch(V)
7.39	-0.39683	0.44984	1.82984	-1.72465	6.20757	4.065674
8.28	0.663168	1.23984	2.1898	1.438998	4.114249	2.610997
8.33	-0.20683	0.473168	2.1465	1.072398	4.694545	2.306067
8.43	-0.56016	0.523168	1.70984	-1.16548	6.502472	1.922469
9.06	-0.33016	0.30984	1.206504	-1.26483	4.312859	1.296009
9.29	0.876504	1.536504	1.763168	1.588581	4.630875	1.837436
9.39	-0.4035	0.593168	1.99984	-1.18306	6.550346	2.950431
9.46	0.713168	1.576504	2.9832	1.679621	5.715047	3.713616
10.05	-0.23016	0.63984	1.746504	-1.7611	5.374512	2.420781
10.07	-0.1635	0.643168	1.873168	-0.58012	6.159505	2.202739
10.53	0.643168	1.343168	2.9498	1.535706	3.447305	2.966234
12.14	-0.4835	0.396504	1.35984	-1.70355	4.993934	1.246663
14.28	-0.30016	0.68984	1.113168	-1.36512	4.821385	1.47498
15.5	0.553168	1.283168	1.236504	-0.70755	4.187304	1.712319
15.51	-0.4935	0.496504	1.20984	-1.23683	4.094553	1.274388

From Table 4.11, we see that there are some samples where the voltage mismatch is large at the feeder head, however, it is not the case for every sample, yet every sample experiences a large voltage mismatch downstream. Hence, the voltage mismatch at the feeder head cannot be the primary cause for the voltage mismatch downstream and we move onto investigating the signatures in the measurement residuals.

With this feeder, we begin by examining the voltage measurement residuals as we will see another unique signature from the list of common possible causes. Fig. 4.10 shows the one-line diagram of SA1273 and the placement of the voltage measurements.

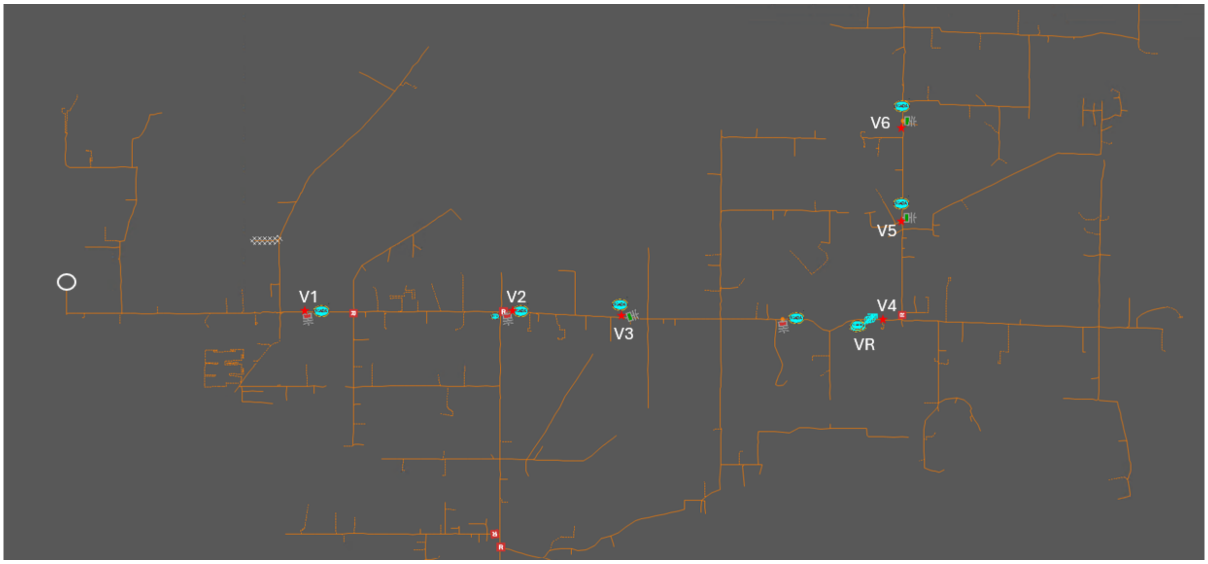

Figure 4.10 SA1273 One-Line Diagram and Measurement Placement

Again, the large white circle on the left of the figure indicates the location of the substation and the voltage measurements are indicated in the figure. Also, there is a voltage regulator (indicated by "VR") present just before voltage measurement V4. Table 4.12 provides the measurement residuals of the voltage measurements.

Table 4.12 Feeder SA1273 Voltage Measurement Residuals

Meas. ID		Va	Vb	Vc
V1	Meas.	122.5	122.7	121.5
	Calc.	123.42	121.33	121.36
	Mismatch	-0.92	1.37	0.14
V2	Meas.	122.4	121.7	122.9
	Calc.	123.76	120.69	121.57
	Mismatch	-1.36	1.01	1.33
V3	Meas.	122.7	121	123.1
	Calc.	123.81	120.35	121.63
	Mismatch	-1.11	0.65	1.47
V4	Meas.	123.4	123.7	122.9
	Calc.	123.7	119.63	121.79
	Mismatch	-0.3	4.07	1.11
V5	Meas.	123	123.9	122.5
	Calc.	123.23	119.08	121.19
	Mismatch	-0.23	4.82	1.31
V6	Meas.	123	123	121.8
	Calc.	123.15	118.71	120.66
	Mismatch	-0.15	4.29	1.14

From Table 4.12, there appears to be a large voltage mismatch in voltage measurements V4, V5, and V6, all on phase B. As all of the voltage measurements that are downstream of the voltage regulator in the system have large measurement residual, this suggests the signature that there is a topology error in the voltage regulator. It is also clear that the measured voltage in the field is at these measurements are much higher than their estimated counterparts, indicating that the actual voltage regulator tap position is boosting the voltage much more actually in the field than what is being modeled. Investigating the model, it is observed that the voltage regulator is set as inoperable in the model and there is a bypass switch closed in the model, bypassing the voltage regulator altogether.

We can verify that the large voltage mismatch is caused by the voltage regulator topology error by adjusting the model and simulating where the voltage regulator is operable. Table 4.13 provides the results of the original max voltage mismatches and the new voltage mismatches after simulated correction.

Table 4.13 Feeder SA1273 Correction Voltage Mismatches

Case	Va	Vb	Vc
14.28 Original	-1.37	-4.19	1.47
14.28 Corrected	-1.37	1.36	1.50

As a result, there are no voltage mismatches outside of the acceptable range of ±2V, thus it is clear that the voltage regulator was the primary cause for the large voltage mismatch as indicated by the signature.

4.7 Conclusions

In this chapter, we investigate a practical application of the signature-based analysis approach that can be applied to measurement residuals of real-time monitoring methods utilities use in practice. We begin by proposing a data analytics-based tool to provide insights and identify unique features correlating to feeders that consistently perform poorly with large voltage mismatches. As a result, a list of common causes for poor performance is identified and used with the signature-based approach to further propose guidelines meant to assist utility engineers diagnose specific suspected causes for large voltage mismatch. This approach is applied to real distribution systems and the main observations show the effectiveness and applicability of approach in practice.

The main contributions of the chapter are:

- A new approach to help identify unique features that contribute to poor real-time monitoring performance on distribution systems.
- A new practical application that uses the signature-based analysis in a guideline based process to help streamline the process for utility engineers to diagnose the cause(s) for poorly performing real-time monitoring.

CHAPTER

5

METHODS FOR MONITORING OF TOPOLOGY ERRORS RESULTING FROM CYBER-ATTACKS

In the previous chapters, we have looked at the limitations of existing standard methods of topology error processing in the presence of the most common unmonitored topology errors: capacitor bank failures and load phase switching for load balancing. An alternative threat to system topology is as a result of malicious actions from a cyber-attack. Unfortunately, critical infrastructure like electric power systems have become a larger target amongst cyber actors. The growing interface between cyber systems and physical systems to alleviate control and operation of many large and complex systems has created vulnerabilities in the security of the system that must be addressed [64].

One of the first types of attacks studied that negatively impacts SE is through the integrity of the data. From [13], we can see that these cyber actors can execute false data injection attacks (FDIA) that inject false data into field measurements in a way that can exploit limitations of standard BDD methods to remain hidden from detection and create a false sense of awareness of the system states. This can create situations of equipment damage and overloading and ultimately create outages within the system. This is especially critical in distribution systems given the reliance of DSSE on the minimal number of meter measurements available.

In this work, we focus on the cyber threat against voltage regulator control. This type of threat can have a significant impact on the system topology and serious physical impact if not carefully monitored. The ultimate goal of such an attack is to cause disruption in normal operations by increasing the voltage to dangerous levels. Consequences of overvoltage in a distribution system include permanent damage to system devices as device insulation will deteriorate rapidly and ultimately burn and cause failures or cause arcs and flashovers [65]. Alternatively, overvoltage can cause failures and damage to customer devices/appliances [66]. The modeling of such an attack and the vulnerabilities that are exploited are discussed in this chapter.

Additionally, it will be shown that an extension can be made to the signature-based analysis proposed in chapter 3 that can include the definition of a signature in measurement normalized residuals as a result of such an attack of system topology through voltage regulator tap error.

5.1 Voltage Regulator Attack Modeling

In the field of cyber-security, the first critical step is developing a well-defined attack model. As it relates to the threat towards voltage regulator control, there are two possible strategies malicious cyber actors can utilize to accomplish the goal of raising the regulated voltage to dangerous levels. Those two strategies include:

1. **Direct modification of voltage regulator tap through malicious tap change commands:** As more communication-able devices are added to the system, the number of entry points and

vulnerabilities which can be exploited to compromise the system are shown to increase [67, 68]. Once the communication system is compromised, cyber actors have the ability to send malicious tap change commands to voltage regulators to accomplish their goal of raising the system voltage to dangerous levels, especially if the Human Machine Interface (HMI) is compromised [69].

2. **Indirect modification of voltage regulator tap through data falsification attack:** Voltage regulators are often times equipped with automated control for efficient operation in the field. Control settings are implemented where the voltage regulator will receive measurements from a point of interest and will react and change its tap position to either increase the regulated voltage or decrease the regulated voltage so that this voltage remains within the specified range determined in the control settings. This creates a closed feedback control loop. This control loop can be exploited however through data falsification attacks [16, 70]. Fig. 5.1 shows the voltage regulator closed feedback control loop and the point of data falsification attack.

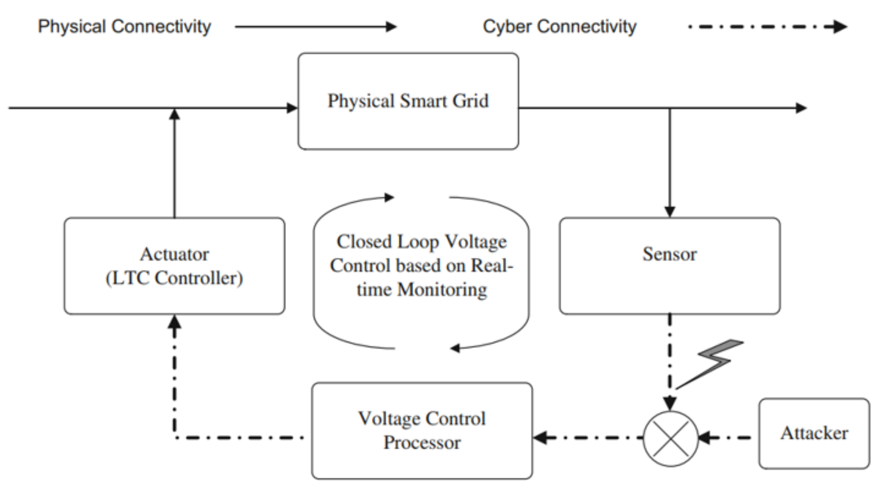

Figure 5.1 Closed Loop Voltage Regulator Control

In this data falsification attack, cyber actors can exploit the integrity of the data (the measured voltage) and manipulate the voltage to appear lower and outside of the control setting range. This forces the voltage regulator to react and change tap position. Because this automated control is due to a false measurement, the true system voltages will begin to increase outside of the acceptable range of operation and can result in permanent damage.

5.2 Signature-Based Detection Methodology

After considering the attack model and the potential topology error threat from wrong voltage regulator tap modeling, the signature-based analysis proposed in chapter 3 can be extended to consider such a threat. Topology errors from voltage regulators have distinctly unique signatures compared to other topology errors and will be addressed here.

Previously, capacitor failures and load transfers were considered and had a significant impact on the estimated branch flow measurements and injection pseudo-measurements, hence most of the normalized residual spread involved the power measurements and pseudo-measurements. However, with incorrect voltage regulator tap modeling, significant impact will be observed in the estimated voltage downstream of the voltage regulator. This will create significant errors in the voltage measurements, thus large residuals will result in the voltage measurements. This is the first insight of the new signature of voltage regulator topology errors and can be verified numerically in the normalized residual spread analysis.

Additionally, as the voltage regulator will boost or buck the voltage for the entirety of the system downstream of it, the voltage measurements in the entire system downstream of the voltage regulator will have large normalized residuals. Even if there are multiple separate measurement islands downstream of the voltage regulator, voltage measurements in each will still experience this large normalized residual. Hence, another critical insight is localization is no longer based upon measurement island as it was with capacitor failures and load transfers. However, if all of these voltage measurements downstream of the voltage regulator experience large normalized residual, it is then localized as part of the downstream portion and the voltage regulator that bounds this section will be suspected. It should also be noted that normalized residuals of voltage measurements upstream of the voltage regulator are not impacted by such a voltage regulator topology error.

Finally, loads that are defined from load estimate pseudo-measurements are modeled as constant power loads in C-DSSE. As such, estimated power injections from the load estimates will not be impacted by the incorrect voltage level. However, if devices that are modeled as constant impedance shunt elements like shunt capacitors, the estimated injection from such a constant impedance model is voltage dependent and will be impacted by the incorrect voltage level. As a result, a moderate signature will be observed of capacitor topology error from this and may skew the analysis. It becomes critical then that the signature of voltage regulator topology error is distinguished apart from this and verification is needed that the observed capacitor signature is just a symptom of the voltage regulator topology error. Thus, a comparison is needed that will display this distinction.

Ultimately, the signature-based detection can utilize the following signature for detection of voltage regulator topology error:

1. Large normalized residuals exist in voltage measurements.
2. The large voltage measurement normalized residuals are localized to the portion of the system downstream of the voltage regulator and are not confined to measurement islands. Voltage measurement normalized residuals upstream of the voltage regulator are not impacted.
3. Moderate signatures of capacitor topology error may be observed, but can be distinguished as a symptom of voltage regulator topology error through further examination.

Case studies illustrating such signatures are provided next.

5.3 Case Study

To validate the signature-based analysis extension in measurement normalized residuals as a result of voltage regulator topology error caused by cyber attack, case studies are performed on the two large realistic feeders based on real distribution system data. We first validate the signature in voltage regulator topology errors using realistic feeder AM, then distinguish the differences between voltage regulator topology error and capacitor failure topology error using realistic feeder WA. It should be noted that for a standard voltage regulator with a max boost of 10% and 32 taps (16 taps in the boost direction, 16 taps in the buck direction), the minimum tap error to cause disruptions should be 8 taps. For this type of voltage regulator, 8 taps refers to a voltage change of 5%. When it is desired to utilities to maintain system voltages within 95% to 105% of the nominal voltage, a tap change of 8 can easily raise the voltage level outside of the acceptable range for the cyber adversary and bring the voltages to dangerous levels. Hence in both cases, the minimum modeled tap error will be 8 taps.

5.3.1 Case: Voltage Regulator Data Falsification Attack

We first validate the signatures in measurement normalized residual present with a voltage regulator topology error. We utilize realistic feeder AM for this case. In this feeder, there is not a voltage regulator originally placed, thus one must be placed and we place one at a common location further downstream on the main feeder closer to the end of the feeder. As a result, any existing capacitor in this portion of the feeder is removed to simulate the trade-off of placement of voltage regulators instead of capacitors for Volt-VAR control schemes. Except in extreme cases when the voltages are extremely low, it is unlikely to operate both capacitor and voltage regulator to control the system voltages.

Fig. 5.2 shows the one-line diagram of the realistic feeder AM with measurement placement and location of placed voltage regulator. The placed voltage regulator is a standard 32 tap regulator with a max boost of 10%. It should be noted that a voltage measurement is usually placed on the receiving side of the voltage regulator for automatic control, however, in the case of data falsification attack, this measurement will be falsified, thus we neglect it from the analysis.

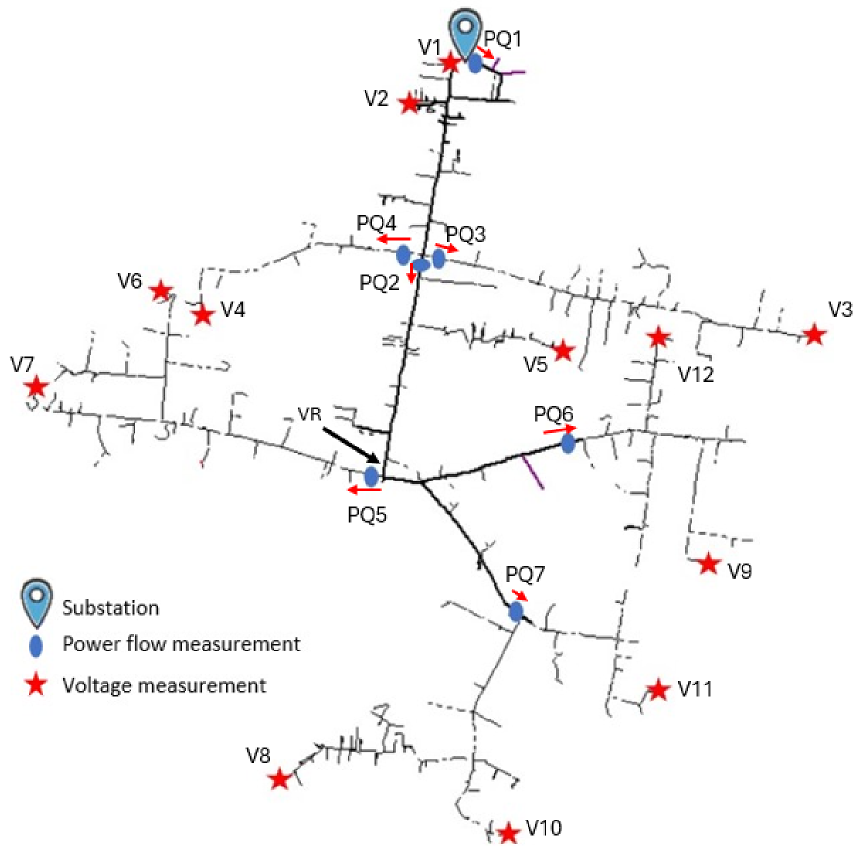

Figure 5.2 Realistic Feeder AM One-Line Diagram

As mentioned we simulate a tap error of 8 taps as the minimum attack model. We simulate this error on phase A and the resulting voltage measurement normalized residuals are shown in Table 5.1

Table 5.1 Feeder AM Voltage Measurement Normalized Residual w/ Tap Error 8 on Phase A

Meas. ID	r_a^N	r_b^N	r_c^N
V1	0.066	0.259	0.465
V2	0.686	0.395	0.327
V3	-	-	0.254
V4	0.111	-	-
V5	0.358	-	-
V6	-	-	0.800
V7	-	-	1.771
V8	4.347	-	-
V9	-	0.114	-
V10	4.269	-	-
V11	-	0.097	-
V12	4.511	-	-

From Table 5.1, it is clear that there are normalized residuals larger than our standard normalized threshold of 3. The voltage measurements with large normalized residuals are V8, V10, and V12 which are all phase A measurements and all downstream of the voltage regulator. These 3 measurements exist in two separate measurement islands formed by PQ6 and PQ7. As a result, the identified signature shows large normalized residuals in voltage measurements, on phase A, and localized to the entire portion of the system that is downstream of the voltage regulator. This indicates a voltage regulator topology error.

For completeness, the normalized residuals for branch flow measurements are also investigated for signatures. Table 5.2 provides the normalized residual for real and reactive power branch flow measurements as a result of the simulated voltage regulator topology error.

Table 5.2 Feeder AM Branch Flow Measurement Normalized Residual w/ Tap Error 8 on Phase A

	Real Power			Reactive Power		
Meas. ID	r_a^N	r_b^N	r_c^N	r_a^N	r_b^N	r_c^N
PQ1	0.159	0.328	0.465	0.053	0.045	0.097
PQ2	0.553	0.429	0.418	0.063	0.187	0.24
PQ3	-	-	0.757	-	-	0.198
PQ4	0.334	-	-	0.567	-	-
PQ5	0.0001	0.002	0.449	0.007	0.002	0.262
PQ6	0.267	0.205	0.001	1.283	0.010	0.010
PQ7	0.394	0.689	1.083	0.827	0.158	0.549

From Table 5.2, it is clear that real and reactive power branch flow do not have large normalized residual as expected. To further prove that real and reactive power normalized residual is small even in the load pseudo-measurements, Fig. 5.3 and Fig. 5.4 shows the normalized residual heat maps for real and reactive power on phase A respectively.

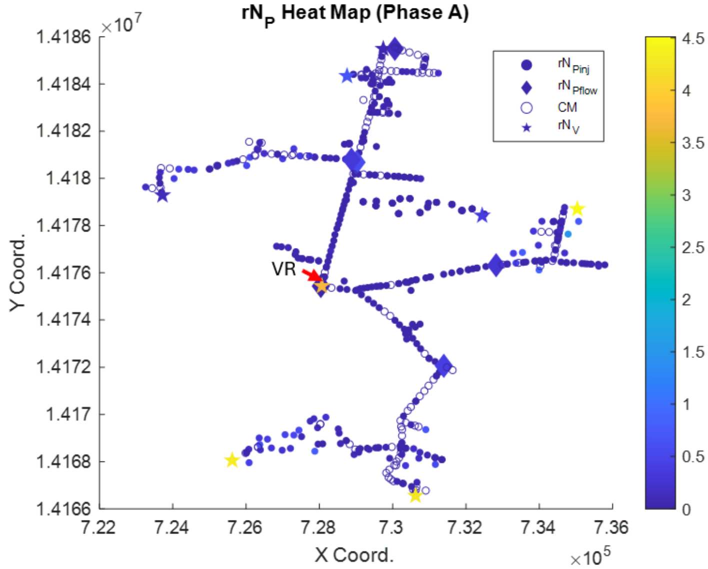

Figure 5.3 Feeder AM Real Power Normalized Residual Heat Map for Voltage Regulator Error

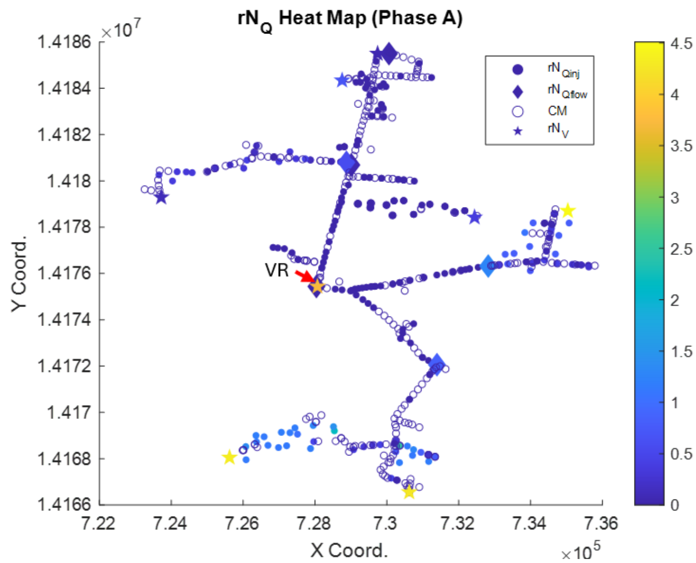

Figure 5.4 Feeder AM Reactive Power Normalized Residual Heat Map for Voltage Regulator Error

From these figures, it is clear that there are no large normalized residuals for any real and reactive load pseudo-measurements shown in the heat map as the filled circles.

Ultimately, this case proves that a signature can be identified for voltage regulator topology errors by identifying large normalized residuals in voltage measurements, observing that all measurements downstream of the impacted voltage measurement have this large normalized residual, and there is little impact on the branch flow measurements and load pseudo-measurements, specifically if there are no voltage dependent loads/capacitors modeled downstream of the impacted voltage regulator.

We will see that this last point will change when there is a capacitor downstream of the impacted voltage regulator in the next case.

5.3.2 Case: Voltage Regulator Data Falsification Attack with Downstream Capacitor

As mentioned in C-DSSE, the load estimates are all used as pseudo-measurements at every node. In this case, every load will essentially be modeled as constant power loads and will not be impacted by the level of the estimated voltage.

Shunt devices, like shunt capacitors, are modeled as a constant impedance and the resulting injection will be voltage dependent, meaning estimated injection from the shunt device will change based on the estimated voltage. If there is a data falsification attack that creates a topology error in a voltage regulator, this creates inaccuracies in the estimated voltage downstream of the voltage regulator, thus the injection from the shunt device will be erroneous because of its dependency on the estimated voltage. As a result, there will be a signature present from this indicating a capacitor error, thus it is critical to distinguish the signature of the voltage regulator topology error. This will be the focus in this case.

In this case, we use realistic feeder WA, which is shown in Fig. 5.5 with placed voltage measurements and branch flow measurements. The location of the voltage regulator is shown as "VR" in the figure and is a standard 32 tap voltage regulator with a max boost of 10%. Additionally the presence of a capacitor downstream of the voltage regulator is indicated in the figure as "cap." Again, it is common to place a voltage measurement on the receiving node of the voltage regulator to be used for automated control, however this will be falsified in the data falsification attack, thus we neglect it from the analysis.

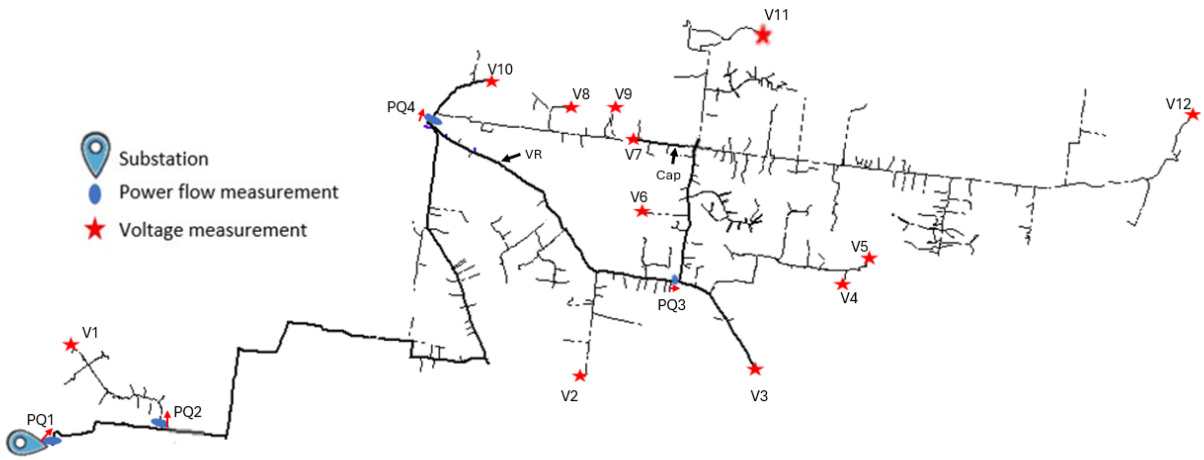

Figure 5.5 Realistic Feeder WA One-Line Diagram

The topology error in the voltage regulator as a result of a data falsification attack is simulated on phase A of the voltage regulator with the minimum desired tap change from the cyber adversary of 8. The resulting normalized residuals for voltage measurements are first shown in Table 5.3

Table 5.3 Feeder WA Voltage Measurement Normalized Residual w/ Tap Error 8 on Phase A

Meas. ID	r_a^N	r_b^N	r_c^N
V1	0.711	-	-
V2	5.590	-	-
V3	3.934	0.836	1.061
V4	-	-	1.834
V5	5.891	-	-
V6	-	-	0.349
V7	4.107	0.239	0.969
V8	-	-	0.647
V9	-	0.594	-
V10	0.192	0.842	0.285
V11	6.186	-	-
V12	-	0.056	-

From Table 5.3, it is clear that there are large normalized residuals (larger than standard threshold of 3) in voltage measurements, especially on phase A. Also, it is clear that all large voltage measurement normalized residuals can be localized downstream of the voltage regulator and not confined to a single measurement island. This is indicates the signature of voltage regulator topology error.

For completeness, we investigate the normalized residuals for real and reactive power branch flow measurements. Table 5.4 provides the normalized residuals for these branch flow measurements.

Table 5.4 Feeder WA Branch Flow Measurement Normalized Residual w/ Tap Error 8 on Phase A

	Real Power			Reactive Power		
Meas. ID	r_a^N	r_b^N	r_c^N	r_a^N	r_b^N	r_c^N
PQ1	2.247	0.875	1.678	1.391	0.323	0.336
PQ2	0.474	-	-	1.175	-	-
PQ3	0.650	0.101	0.355	8.604	0.051	0.800
PQ4	0.967	0.062	0.247	0.776	0.560	0.126

From Table 5.4, there is one reactive power branch flow measurement with a large normalized residual for reactive power on phase A of PQ3. This measurement is the head of a measurement island that is downstream of the voltage regulator and contained within the measurement island. This is a indication of a capacitor error. The pseudo-measurement normalized residuals are observed in the heat map in Fig. 5.7 and Fig. 5.7

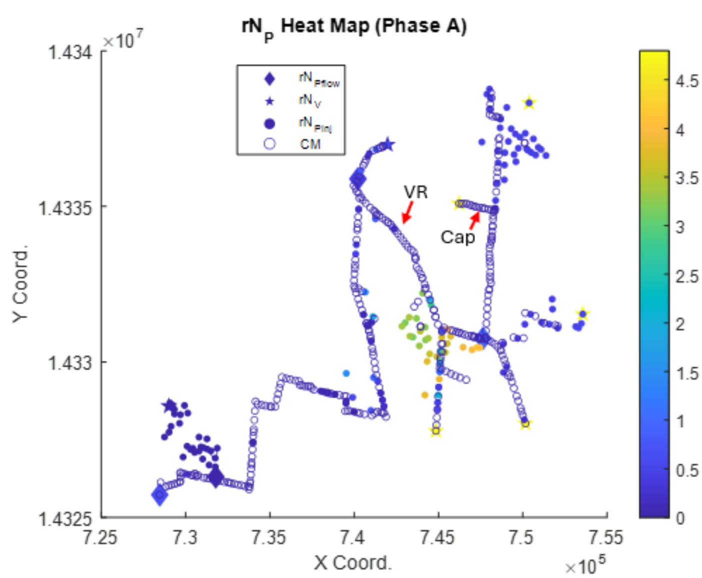

Figure 5.6 Feeder WA Real Power Normalized Residual Heat Map for Voltage Regulator Error

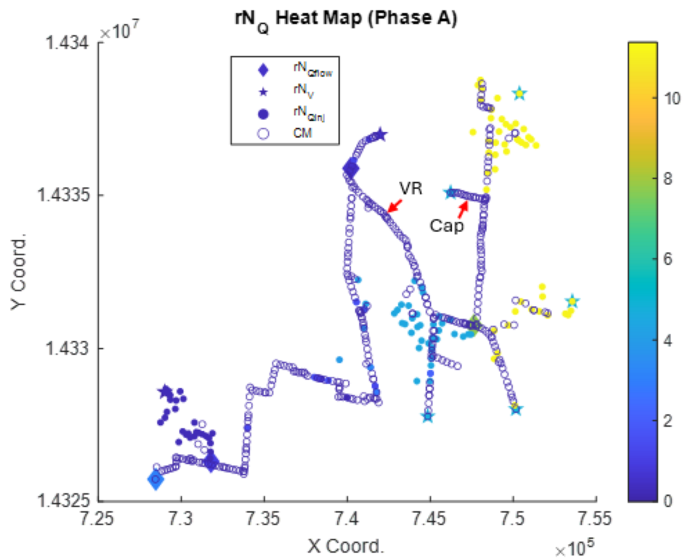

Figure 5.7 Feeder WA Reactive Power Normalized Residual Heat Map for Voltage Regulator Error

From these figures, it is clear that there is large normalized residuals for the load pseudo-measurements for reactive power. Additionally, the largest are localized within the measurement island that contains a capacitor. These are all indications that we have seen for capacitor failure signatures previously. A key distinction however comes with the signatures that are present in voltage measurements. Previously, in capacitor failure signatures, the voltage measurement normalized residuals were never indicated as large, whereas here there is clearly large normalized residual in the voltage measurements and a pattern that they all exist downstream of the voltage regulator.

Another distinction can be made in the magnitude of normalized residuals seen in the reactive load pseudo-measurements. The largest normalized residual seen for reactive pseudo-measurements seen in Fig. 5.7 is roughly 10-11. The heat map shown in Fig. 5.8

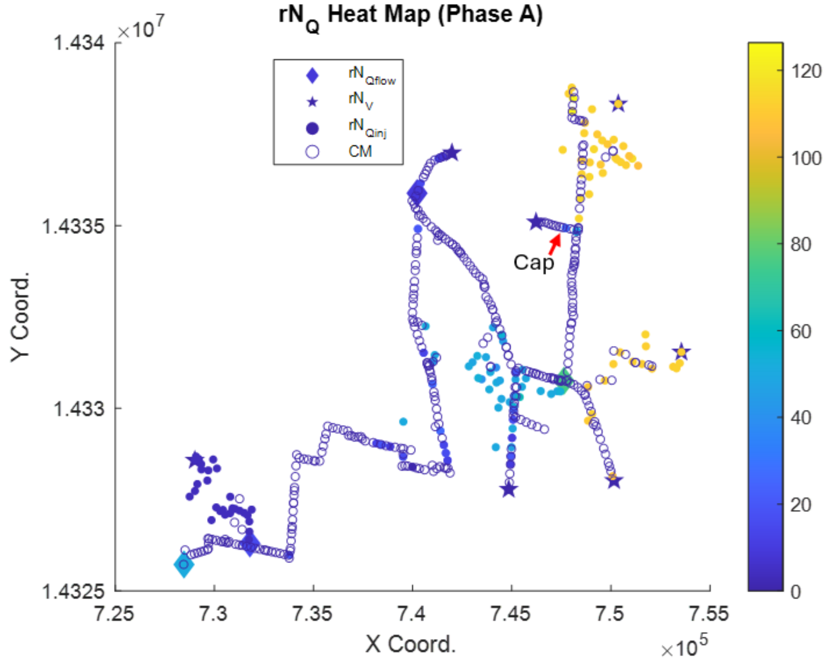

Figure 5.8 Feeder WA Reactive Power Normalized Residual Heat Map for Capacitor Failure

From Fig. 5.8, the largest normalized residuals for reactive power are localized in the measurement island and at a magnitude of around 120. This is a huge discrepancy between the 10-11 in reactive load pseudo-measurement normalized residual when the voltage regulator is simulated with a topology error. This correlates to the large difference between the capacitor being completely disconnected and the estimated injection being adjusted slightly by the new estimated voltage level. This insight will be used as elimination of the signature indicating capacitor failure and verification that voltage regulator topology error is the main signature identified.

5.4 Conclusions

This chapter shows the robustness of the signature-based approach to be used for topology error monitoring of new topology errors. In this chapter, the new topology error addressed is as a result of action from a malicious cyber actor to cause a topology error in voltage regulator tap position in order to raise the voltage to dangerous levels that will cause permanent damage to the system. Such an attack is modeled through the falsification of data that is used in automated Volt-VAR control to adjust the voltage regulator to maintain a regulated voltage within a desired setting. The signature for such a topology error is defined and case studies validate the signature of voltage regulator topology errors.

The main contribution of this chapter is the extension of the signature-based analysis to consider such a topology error threat of incorrect voltage regulator tap modeling as a result of cyber interaction. This new insight shows the ability of the proposed approach to consider new topology error threats that have not been considered in original development

CHAPTER

6

CONCLUSION AND FUTURE WORK

6.1 Conclusion and Contributions

This dissertation focused on addressing the challenges associated with using real-time monitoring methods in distribution systems as it relates to challenging and common topology errors. For effective and reliable distribution system operation and control, accurate real-time monitoring is critical to provide quality situational awareness. Such real-time monitoring methods are severely limited in the presence of topology errors because these methods assume a correct system model and correct system topology.

This work first conducts an assessment of branch current based distribution system state estimation (C-DSSE) and bad data detection (BDD) performance in the presence of common challenging topology errors in chapter 2. Previous works had considered the use of traditional BDD methods for topology error monitoring, however did not consider the most common and challenging topology errors, capacitor failures and load transfers, thus an assessment of BDD capabilities to be used for topology error monitoring is needed. The main contributions in chapter 2 are:

- A new approach to extending traditional BDD methods to incremental analysis for enhanced performance needed for topology error monitoring.
- An assessment of the performance of C-DSSE and BDD methods in the presence of common and challenging topology errors and the minimum measurement placement scheme to enhance BDD capabilities is identified from common practical placement schemes.
- Preliminary signatures are observed that will be used for topology error detection.
- Limitations are identified in the identification of these topology errors that will encourage further development in the next chapter.

In chapter 3, this work then addresses the observations and limitations identified from the assessment. In this chapter, focus is first made to improve the overall quality and accuracy of C-DSSE estimates through more accurate modeling of in an effort to improve the overall capability of BDD for topology error monitoring. Then, residual-spread analysis is used as verification of normalized residual signatures as a result of the more capable BDD. Finally, a new method using the signatures is proposed to accurate detect and identify the common challenging topology errors. The main contributions of chapter 3 are:

- Proposal of a new approach to modeling of measurement uncertainty meant to improve the overall quality/accuracy of C-DSSE estimates.
- Verification of normalized residual signatures through numerical analysis of the normalized residual spread.
- Proposal of a three step method that uses the signatures in normalized residuals from BDD to 1) Detect the presence and type of topology error, 2) Localize the detected topology error to a

specific smaller portion of the system called measurement islands, and 3) Identify the specific source of the detected topology error.

In practice, utilities will use a variety of different similar real-time monitoring methods. In chapter 4, a practical application of the signature-based analysis proposed in the previous chapter is examined on a different, yet highly comparable, practical real-time monitoring method. A method is introduced to efficiently identify unique features present in distribution systems that have consistent challenges with large mismatch between voltage estimates and measured voltages. These features are meant to provide insight into the possible cause, however a more direct signature-based approach is defined to specifically diagnose and pinpoint suspected cause(s) for the large voltage mismatch. This work is meant to assist utility engineers efficient diagnose issues in the real-time monitoring to streamline correction and improve performance. The main contributions of chapter 4 are:

- A new approach to help identify unique features that contribute to poor real-time monitoring performance on distribution systems.
- A new practical application that uses the signature-based analysis in a guideline based process to help streamline the process for utility engineers to diagnose the cause(s) for poorly performing real-time monitoring.

Challenging topology errors can also come from malicious cyber actor sources. In chapter 5, an attack model is examined that is aimed at maliciously changing voltage regulator settings that will raise the system voltages to dangerous levels and cause significant permanent damage. Considering this attack model, the proposed signature-based analysis is extended to consider this new topology error of incorrectly modeled voltage regulator tap. Thus, the main contributions of chapter 5 include the extension of the extension of the signature-based analysis to consider such a topology error threat of incorrect voltage regulator tap modeling as a result of cyber interaction. This new insight shows the ability of the proposed approach to consider new topology error threats that have not been considered in original development

6.2 Future Work

Potential areas for further development of topology error monitoring in distribution systems are identified. These potential research topics are as follows:

6.2.1 Computational Cost of Topology Monitoring Method to a New Feeder in Practice

In chapter 3, test results effectively show the scalability of the method to realistically sized distribution models. This applicability in practice can be more definitively defined. Specifically, future work should focus on the overall methodology that a utility can use to apply this method to a new feeder in practice and ultimately quantitatively present the computational effort required.

6.2.2 More Accurate Load Estimation Techniques

In chapter 3, more accurate measurement uncertainty modeling is proposed. This modeling assumes a basic load estimation technique for pseudo-measurements, which is highly varying and less accurate. In this, certain assumptions are made about the load accuracy, thus future work should focus on a more accurate depiction of load variation. As a result, focus can be placed on further improvement by developing more accurate load estimation techniques, especially based on Advanced Metering Infrastructure (AMI).

6.2.3 Impact of Artificial Intelligence in Topology Monitoring

In this work, methods for topology error monitoring are proposed. The basis of the method is through a signature-based analysis which requires manual development. Across various similar fields of research, artificial intelligence is being utilized to enhance such analysis for large complex systems, thus can be an area of future focus to enhance topology error monitoring in distribution systems.

6.2.4 Detection of Cyber-Attacks Against DER Control

In chapter 5, a cyber-attack model is discussed that targets the system topology through malicious operation of voltage regulators meant to cause dangerous voltage levels and cause permanent damage. Other threats to system topology should be considered as well, such as cyber threats to DER control. Attacks can target smart inverters and other DER controllers to create maloperations that manipulate the output of the DER or even cause the device to incorrectly trip [71]. As seen in the practical application of the signature based analysis in chapter 4, incorrect modeling of DER penetration can impact estimation, thus such an attack against DER must be carefully monitored.

BIBLIOGRAPHY

[1] Abur, A. & Exposito, A. G. *Power System State Estimation Theory and Implementation*. New York, NY: Marcel Dekker, Inc, 2004.

[2] Ghosh, A. et al. “Load modeling for distribution circuit state estimation”. *IEEE Transactions on Power Delivery* **12**.2 (1997), pp. 999–1005.

[3] Kersting, W. H. & Phillips, W. H. “Load allocation based upon automatic meter readings”. *2008 IEEE/PES Transmission and Distribution Conference and Exposition*. 2008, pp. 1–7.

[4] Baran, M. “Challenges in state estimation on distribution systems”. *2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.01CH37262)*. Vol. 1. 2001, 429–433 vol.1.

[5] Lu, C. et al. “Distribution system state estimation”. *IEEE Transactions on Power Systems* **10**.1 (1995), pp. 229–240.

[6] Donmez, B. & Abur, A. “Static State Estimation Without Topology Processor”. *IEEE Transactions on Power Systems* **36**.5 (2021), pp. 4712–4722.

[7] Zamani, V. & Baran, M. “Topology processing in distribution systems by Branch Current based state estimation”. *2015 North American Power Symposium (NAPS)*. 2015, pp. 1–5.

[8] Jouybari-Moghaddam, H. & Sidhu, T. S. “A study of capacitor element failures in high voltage Shunt Capacitor Banks”. *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. 2017, pp. 1–4.

[9] Moghaddam, H. “Fault Location in High Voltage Shunt Capacitor Banks”. PhD thesis. London, Ontario, Canada: University of Western Ontario, 2017.

[10] Baran, M. & Wu, F. “Network reconfiguration in distribution systems for loss reduction and load balancing”. *IEEE Transactions on Power Delivery* **4**.2 (1989), pp. 1401–1407.

[11] Haq, S. U. et al. “Automatic three phase load balancing system by using fast switching relay in three phase distribution system”. *2018 1st International Conference on Power, Energy and Smart Grid (ICPESG)*. 2018, pp. 1–6.

[12] Park, D. & Walstrom, M. *Cyberattack on Critical Infrastructure: Russia and the Ukrainian Power Grid Attacks*. URL: `https://jsis.washington.edu/news/cyberattack-critical-infrastructure-russia-ukrainian-power-grid-attacks/`. (accessed: 08.03.2022).

[13] Liu, Y. et al. “False Data Injection Attacks against State Estimation in Electric Power Grids”. *ACM Trans. Inf. Syst. Secur.* **14**.1 (2011).

[14] Ashok, A. & Govindarasu, M. “Cyber attacks on power system state estimation through topology errors”. *2012 IEEE Power and Energy Society General Meeting*. 2012, pp. 1–8.

[15] Faquir, D. et al. “Cybersecurity in smart grids, challenges and solutions”. *AIMS Electronics and Electrical Engineering* **5**.1 (2021), pp. 24–37.

[16] Anwar, A. et al. “False Data Injection Attack Targeting the LTC Transformers to Disrupt Smart Grid Operation”. *International Conference on Security and Privacy in Communication Networks*. Ed. by Tian, J. et al. Cham: Springer International Publishing, 2015, pp. 252–266.

[17] Zamani, V. “State Estimation for Volt/VAR Control on Active Power Distribution Systems”. PhD thesis. Raleigh, NC, USA: NC State University, 2014.

[18] Janssen, P. et al. “Meter placement impact on distribution system state estimation”. *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*. 2013, pp. 1–4.

[19] Crawford, K. & Baran, M. E. “Topology Error Monitoring Using Bad Data Detection Methods”. *IEEE Transactions on Industry Applications* **60**.1 (2024), pp. 1476–1483.

[20] Pau, M. et al. “Uncertainty sources affecting voltage profile in Distribution System State Estimation”. *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. 2015, pp. 109–114.

[21] Santos, R. Z. S. & Orillaza, J. R. C. “Distribution System State Estimator Using SCADA and PMU Measurements”. *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*. 2018, pp. 558–562.

[22] Almutairi, S. et al. “Performance of Branch-Current Based Distribution System State Estimation”. *2018 North American Power Symposium (NAPS)*. 2018, pp. 1–6.

[23] Crawford, K. & Baran, M. E. “More Accurate Measurement Modeling to Improve the Performance of Distribution System State Estimation”. *IEEE Access* **12** (2024), pp. 124527–124536.

[24] Göl, M. & Abur, A. “A modified Chi-Squares test for improved bad data detection”. *2015 IEEE Eindhoven PowerTech*. 2015, pp. 1–5.

[25] Tarali, A. “Bad Data Detection in Two Stage Estimation Using Phasor Measurements”. PhD thesis. Boston, MA, USA: Northeastern University, 2012.

[26] Crawford, K. & Baran, M. “Improving Topology Error Detection with Distribution System State Estimation”. *2022 IEEE Texas Power and Energy Conference (TPEC)*. 2022, pp. 1–6.

[27] IEEE PES Test Feeder Resources. Accessed July 5, 2021. IEEE PES. URL: `https://cmte.ieee.org/pes-testfeeders/resources/`.

[28] Clanents, K. A. et al. "Power System State Estimation with Measurement Deficiency: an Observability/Measurement Placement Algorithm". *IEEE Transactions on Power Apparatus and Systems* **PAS-102**.7 (1983), pp. 2012–2020.

[29] Giordani, A. & Mari, L. "Measurement, Models, and Uncertainty". *IEEE Transactions on Instrumentation and Measurement* **61**.8 (2012), pp. 2144–2152.

[30] Wang, H. et al. "A Robust Measurement Placement Method for Active Distribution System State Estimation Considering Network Reconfiguration". *IEEE Transactions on Smart Grid* **9**.3 (2018), pp. 2108–2117.

[31] Muscas, C. et al. "Impact of Different Uncertainty Sources on a Three-Phase State Estimator for Distribution Networks". *IEEE Transactions on Instrumentation and Measurement* **63**.9 (2014), pp. 2200–2209.

[32] Mann, B. & Heydt, G. "Noncollocated Voltage and Current Measurements Used to Obtain Power". *IEEE Transactions on Power Systems* **21**.4 (2006), pp. 1972–1973.

[33] Primadianto, A. et al. "Requirements of state estimation in smart distribution grid". *Proc. 23rd Int. Conf. Elect. Distrib.* 2015, pp. 1–5.

[34] Steele, J. S. *The Variance Sum Law*. URL: `https://web.pdx.edu/~joel8/resources/ConceptualPresentationResources/VarianceSumLaw.pdf`. Last visited on 2023/02/22. Portland State University.

[35] Singh, R et al. "Choice of estimator for distribution system state estimation". *IET generation, transmission & distribution* **3**.7 (2009), p. 666.

[36] Schoene, J. et al. "Quantifying Performance of Distribution System State Estimators in Supporting Advanced Applications". *IEEE Open Access Journal of Power and Energy* **7** (2020), pp. 151–162.

[37] Mooney, C. *Monte Carlo Simulation*. Monte Carlo Simulation no. 116. SAGE Publications, 1997.

[38] Devore, J. L. *Probability and statistics for engineering and the Sciences*. 9th. Cengage Learning, 2016.

[39] Zamani, V. & Baran, M. E. "Meter Placement for Conservation Voltage Reduction in Distribution Systems". *IEEE Transactions on Power Systems* **33**.2 (2018), pp. 2109–2116.

[40] Fotopoulou, M. et al. "A Review on Distribution System State Estimation Algorithms". *Applied Sciences* **12**.21 (2022).

[41] Boos, D. D. & Stefanski, L. A. *Essential statistical inference: theory and methods*. Vol. 591. Springer, 2013.

[42] Clements, K. A. et al. "Power System State Estimation Residual Analysis: An Algorithm Using Network Topology". *IEEE Transactions on Power Apparatus and Systems* **PAS-100**.4 (1981), pp. 1779–1787.

[43] Hearst, M. A. et al. "Support vector machines". *IEEE Intelligent Systems and their applications* **13**.4 (1998), pp. 18–28.

[44] Weston, J. & Watkins, C. *Multi-class support vector machines*. Tech. rep. Department of Computer Science, Royal Holloway University of London, 1998.

[45] Ukil, A. "Support vector machine". *Intelligent systems and signal processing in power engineering* (2007), pp. 161–226.

[46] Scholkopf, B. et al. "Comparing support vector machines with Gaussian kernels to radial basis function classifiers". *IEEE Transactions on Signal Processing* **45**.11 (1997), pp. 2758–2765.

[47] Crawford, K. & Baran, M. "Capacitor Bank Failure Detection in Distribution Systems Using State Estimation". *2023 North American Power Symposium (NAPS)*. 2023, pp. 1–5.

[48] Schoene, J. et al. "Historical AMI Data for Distribution System State Estimator Deployment". *SoutheastCon 2021*. 2021, pp. 1–5.

[49] Schoene, J. et al. "Quantifying Performance of Distribution System State Estimators in Supporting Advanced Applications". *IEEE Open Access Journal of Power and Energy* **7** (2020), pp. 151–162.

[50] Adeosun, O. et al. "An Analysis of DMS Power Flow Performance". *2023 IEEE Texas Power and Energy Conference (TPEC)*. 2023, pp. 1–5.

[51] Cakir, G. et al. "Identifying Factors Contributing to Poor Performance of Near-Real-Time Power Flow". *2023 IEEE Texas Power and Energy Conference (TPEC)*. 2023, pp. 1–6.

[52] Wang, D. et al. "A survey on bad data injection attack in smart grid". *2013 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. 2013, pp. 1–6.

[53] Pau, M. et al. "Analysis of Bad Data Detection Capabilities Through Smart Meter Based State Estimation". *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe)*. 2018, pp. 1–6.

[54] Primadianto, A. & Lu, C.-N. "A Review on Distribution System State Estimation". *IEEE Transactions on Power Systems* **32**.5 (2017), pp. 3875–3883.

[55] Crawford, K. et al. “Identifying Features Correlating to Poor Performance of Distribution System Near-Real-Time Power Flow”. *2023 North American Power Symposium (NAPS)*. 2023, pp. 1–6.

[56] Alstom. *Distribution Network Analysis Functions (DNAF) Standard Applications Analyst’s Guide*. April 11, 2016.

[57] Ko, C.-Y. & Leu, F.-Y. “Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques”. *IEEE Transactions on Education* **64**.1 (2021), pp. 50–57.

[58] Nainggolan, R. et al. “Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the elbow method”. *Journal of Physics: Conference Series* **1361**.1 (2019), p. 012015.

[59] Shahapure, K. R. & Nicholas, C. “Cluster Quality Analysis Using Silhouette Score”. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. 2020, pp. 747–748.

[60] Zou, X. et al. “Logistic Regression Model Optimization and Case Analysis”. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*. 2019, pp. 135–139.

[61] James, G. et al. *An introduction to statistical learning: With applications in R*. Second Edition. Springer, 2021.

[62] Akaike, H. “A new look at the statistical model identification”. *IEEE Transactions on Automatic Control* **19**.6 (1974), pp. 716–723.

[63] Wesolowska, M. “A study on feature selection based on AICc and its application to microarray data” (2009).

[64] Paul, S. et al. “On Vulnerability and Resilience of Cyber-Physical Power Systems: A Review”. *IEEE Systems Journal* **16**.2 (2022), pp. 2367–2378.

[65] Edomah, N. “Effects of voltage sags, swell and other disturbances on electrical equipment and their economic implications”. IET, 2009, pp. 1–4.

[66] Schainker, R. *Effects of Temporary Overvoltage on Residential Products*. Tech. rep. EPRI Solutions Inc., 2005.

[67] *Cyber Threat and Vulnerability Analysis of the U.S. Electric Sector*. Mission Support Center, Idaho National Laboratory, 2016.

[68] Wang, W. & Lu, Z. “Cyber security in the Smart Grid: Survey and challenges”. *Computer Networks* **57**.5 (2013), pp. 1344–1371.

[69] Parthasarathy, S. & Kundur, D. “Bloom filter based intrusion detection for smart grid SCADA”. *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. 2012, pp. 1–6.

[70] Bhusal, N. et al. “Detection of Cyber Attacks on Voltage Regulation in Distribution Systems Using Machine Learning”. *IEEE Access* **9** (2021), pp. 40402–40416.

[71] Qi, J. et al. “Cybersecurity for distributed energy resources and smart inverters”. *IET Cyber-Physical Systems: Theory & Applications* **1**.1 (2016), pp. 28–39. eprint: `https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-cps.2016.0018`.