DATA-ENHANCED APPLICATIONS FOR POWER SYSTEMS ANALYSIS

BY

KATHERINE MARGARET ROGERS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Thomas J. Overbye, Chair
Professor Peter W. Sauer
Assistant Professor Alejandro Domínguez-García
Professor David M. Nicol
Dr. James D. Weber, PowerWorld Corporation

Abstract

As power system data becomes more abundant, there is an increasing need for applications which can make appropriate use of this data to perform power system analysis tasks. Data provides opportunities for performing event detection, diagnostics, and forensics on the system in a way which has not previously been possible. The integration of data and information systems remains a key challenge in taking full advantage of the smart grid. There is a need to "tap into the data" to discover relationships and to draw attention to patterns. Opportunities are provided by the data to develop and perform diagnostics. New functionality and applications in the smart grid can best be enabled by taking full advantage of available data. If not resolved, the lack of suitable advanced data analysis capability threatens to become a barrier to developing new intelligent grid operation and control paradigms. This need motivates the work in this thesis. There remains a great deal of opportunity to advance the state of the art, especially for developing suitable techniques to perform automated analysis and event detection based on the data. The work is presented with the anticipation of encouraging future extensions, especially with respect to the identification and classification of patterns in power system data.

This thesis examines data mining and advanced data analysis techniques in the context of a number of specific power system applications. In particular, these applications concern the use of model data (sensitivities) to identify relationships, the data-enhanced estimation of network models, event identification from oscillation monitoring data, and dealing with the challenges of real-world data and data quality. Several important contributions of this work are the following. Analysis and results show that sensitivity and model data can be leveraged via correlation and cluster analysis to gain information about the expected or model-supported relationships and patterns in power systems. In particular, results exemplify these benefits for the areas of market power potential identification, coordinated control of resources, and in the creation of structure-preserving network reductions. Results also show that a space of network reductions which satisfy power flow solution equivalence exists and can be further explored by choice of desirable constraints. Another key contribution is the realization that the expected or model-supported patterns can be ascertained from actual data, which is done in the form of estimating equivalent models and network model parameters. Solutions are presented which highlight the benefit of data-over-time for estimating the structure of a reduced network model in the form of a matrix

representing the system structure, and at a lower-level, for estimating parameters on individual transmission lines from historical data. The realization is that while data facilitates these advanced data-centric applications, there are also barriers to the progress of such applications. These limitations are a function of the data itself, and arise both with respect to the noise and error qualities of the data as well as from a lack of adequate representation of true characteristics present in the data, such as load fluctuations. Thus, it is imperative to understand, improve, and enhance the quality of real data.

It is concluded that the ability to use and analyze real data is critical to implementing and advancing cutting-edge data mining solutions in power systems. The second half of the work in this thesis focuses on addressing these issues. Real data is shown to be useful for diagnosing problems. An oscillation monitoring event detection framework is presented, where results indicate that with information from nominal modes, it may be possible to correctly classify ringdowns to an originating event. The data quality (availability and integrity) issues are found to be critical to this line of work, and it is concluded that this is probably the most pressing area for immediate attention. Without trusting the data, one cannot trust decisions based upon the data. A key contribution is to relate the data and cyber dependencies of power systems and demonstrate how it is possible to take advantage of available information (including from cyber network monitors) to flag suspicious data sensors and measurement values and then take appropriate action. Results show that bad data from malicious sources can be detected without relying upon traditional residual-based schemes. The analysis is performed before such data is available to corrupt applications such as state estimation. In summary, this work presents a number of contributions for the enhancement of power system applications through the use of data and data mining techniques.

Acknowledgments

I would like to thank all of those whom I have had the opportunity to collaborate with and all of those who have taught or mentored me. First, I would like to thank my advisor, Professor Tom Overbye, for his continual guidance and support throughout my time at the University of Illinois. Secondly, thanks to all of my committee members for their interest, involvement, and encouragement of my work over the past years. Additionally, there have been quite a few other people I have worked with, both directly and indirectly, who have inspired me, both at the University of Illinois and in other places. I hope you all know who you are. Specifically, I would like to extend my sincere and continual gratitude to Dr. W. Mack Grady, Professor at the University of Texas at Austin, for encouraging me from the beginning and giving me numerous opportunities to learn, expand, and demonstrate my capabilities.

I would like to acknowledge the Trustworthy Cyber Infrastructure for the Power Grid (TCIPG, formerly TCIP) and all the interactions and collaborations it supports. Thanks in particular to Professor Bill Sanders, Rakesh Bobba, Erich Heine, Kate Morrow, Saman Zonouz, Ray Klump, Angel Aquino-Lugo, Sudipta Dutta, and Tim Yardley. I would also like to thank the Consortium for Electric Reliability Technology Solutions (CERTS) for its support. Thanks to additional financial support from the University of Illinois ECE Distinguished Fellowship, the Grainger Power Engineering Award, the Sargent and Lundy Fellowship, and the Ernest A. Reid Fellowsip. Finally, thanks and good luck to all other Team Overbye students, both past and current, and to my ECE 333 students.

Finally, I would like to thank all of my family, from whom I am incredibly fortunate to receive such ongoing love and support. My grandparents Margaret and Ames Smith are, and have always been, a tremendous source of inspiration to me. I am especially grateful to my immediate family: my parents Lucy and Bob Rogers, my brother Robert, my sister Melanie, and my brother-in-law Max. Last but not least, thanks to my fiancé Charles Matthew Davis (in the future, my name will be Katherine Davis).

Table of Contents

1. Introduction

The system of systems referred to as the "smart grid" is a modernization of current electricity delivery systems that will monitor, protect, and optimize the operation of its elements. A topic at the forefront of the electric power industry is the need for advanced data analysis techniques for power system data. The integration of data and information systems is considered one of the key challenges in taking advantage of what the smart grid offers [1]. High-resolution time-synchronized data has the potential to revolutionize the way the power grid is controlled and operated. Satisfying the requirements for handling the increasing amounts of data from measurement sources is a critical issue. A number of new sensors have been introduced in the last 10-15 years, and it is expected that sensors will continue to increase in both type and quantity [2]. Data from sensing and measurement technologies, enabled by an interoperable communications and information infrastructure, can support a variety of important applications. The balance between factors such as cost, performance, and risk can be optimized by taking advantage of integrated data from sensor, communication, and information management systems.

The idea of considering the data acquisition, data transmitting, and data processing together in the analysis of a wide area measurement system (WAMS) is important. Emphasis must be placed on the WAMS as a means for efficient usage of data and dataflow to achieve a more secure and better strategy for the flow of electrical energy. Data is a fundamental requirement for performing WAMS functions. Consider the data flow from measurements to applications. The data flow begins with the devices: RTUs, SCADA, PMUs, etc. Phasor measurement units (PMUs) measure and record synchronized phasor, or *synchrophasor* data, made possible by a high-accuracy clock, commonly from a Global Positioning System (GPS) receiver [3]. The 1980 paper [4] explains the use of GPS to provide a precise time source, and the 1983 paper [5] shows its use for accurately recording phase angles in power systems. One of the first recognized applications is to use precise angle measurements to improve state estimation [6]. PMUs can exist in a variety of devices found throughout the power system, including protective relays [7]. With the growing deployment of PMUs, a recurring theme is how to proficiently extract "information" and "learn" from data that PMUs provide [8].

In the data flow process, measured data must undergo processing and communication in order to be used. The end point-of-use must successfully receive the data, either raw or filtered. After data is transmitted and received, it must be filtered in a data processing layer before more advanced data analysis, data mining, and future power systems applications can be performed. The data processing level currently includes energy management systems (EMS) where operation and control of the power grid is carried out. Here optimization is performed to operate the system in a state which satisfies objectives such as minimizing cost or losses. Current applications at the EMS level include state estimation, power flow, optimal power flow, and oscillation monitoring. Applications require credible data. State estimation receives significant focus because it often plays the role of a data filter, taking the raw measurements and producing data which can then be used by other applications. Most of this thesis is concerned with the data processing layer. The current grid, as well as the anticipated more intelligent grid, relies on this data. Data integrity is important. Thus, communication properties are critical. In fact, the communication infrastructure of the power grid is in some ways equally important to the electrical infrastructure.

One theme commonly expressed among professionals in the power industry is the need for effective ways to discover key relationships from the data and to emphasize important *patterns*. This leads to the question: How should a pattern be defined? A pattern can be defined as the "opposite of chaos; it is an entity, vaguely defined, that could be given a name" [9]. Patterns can be many things including a face, a word, a speech signal, etc. The goal of pattern recognition is to utilize available sensors, processers, and domain knowledge in the best possible way so that decisions can be made automatically [9]. Doctors diagnosing disease, pilots landing airplanes, and banks scoring credit applications are just a few instances where pattern recognition is valuable. When problems occur in the grid, it is important that the patterns representing the problems be identified as quickly as possible. Then, it is important to provide feedback to system operators so that they can improve their control strategies. From the perspective of improving grid resiliency, if it is possible to effectively identify and classify patterns in the data, it should also be possible to better recognize what is occurring, and act appropriately. Specifically, there is a need to "tap into the data" [2] for diagnostics, performance assessment, etc. There is also a need to develop and enhance computational capabilities to deal with large

volumes of data and algorithms to interpret the data and convert it into information that can be acted upon. The smart grid can best enable new functionality and advanced applications if we take advantage of a wide range of systems which can provide valuable information and then manage the data effectively.

The term "data mining" (like "smart grid") encompasses an entire genre of definitions, and hence can be a dangerously ambiguous problem. There are many challenges: What are the "patterns" we seek, how do we find them, how do we analyze them, and what do they tell us? Certainly, many different disciplines have a need to solve some variants of the problem of extracting useful information from data. Independent efforts in numerous disciplines, unfortunately, tend to be highly specific to a particular problem. It has since been recognized that the need to extract useful information from raw data is a commonality bridging many scientific and engineering fields. Only relatively recently (~1990s) has data mining gained significant recognition as its own field, so a potentially great number of pattern-finding concepts remain to be adopted from other disciplines. For example, a classification approach is applied in [10] to nuclear plant sensors. Similarly, operational hydrology and turbulence provide a surprisingly relevant perspective on the nature of severe disturbances as observed from the time-series data streams [11], [12]. Previous work in the power systems area on cascading disturbances is also relevant [13], [14], [15] in the sense that it involves looking for patterns in data over time. The great challenge with respect to data mining and the power industry is to develop, adapt, and improve techniques for identifying representative patterns specific to power systems.

For power systems, in contrast to general data analysis problems, the underlying physics of the system and devices are mostly known. Also, years of experience, intuition, and understanding about power systems are available to build upon. To most effectively incorporate collected data with this knowledge, we must know exactly which data is relevant and to what purpose. Skill is needed to recognize and obtain the minimum amount of required data for a particular application. Important practical issues arise in actually obtaining the required data and then in securely and reliably transporting the appropriate subset of information to the applications that will ultimately use it.

Currently, operators have insights and rules-of-thumb that come from watching the system over a long period of time and learning about its behavior. Over a long enough time, a watcher of the system will start to see patterns in its behavior. This observer will start to notice what is normal, what is abnormal, what events usually precede other events, etc. Then, after an event occurs, such a knowledgeable observer may be able to look through data and reconstruct a sequence of events. The drawback is that this reconstruction is generally only available after time-consuming manual effort is dedicated by experienced engineers, sifting through measurement data. Manual analysis cannot be the norm. More effective ways to automatically process and analyze both historical and real-time data are needed. Improved methodology is needed to bring important patterns in the system to the forefront, both to engineers and to operators.

Integrating data and data mining theory into power systems analysis is the main purpose of this work. This thesis examines problems and applications for which data mining can be an important and useful part of the solution. Rather than taking a low-level approach to the storage, handling, and passing of data, we approach the problems from an application-centric point of view: What do we really want to accomplish, and how does the data help achieve that goal? The definition of "knowledge" that we seek to obtain is inherently non-concrete. Therefore, we find it necessary to investigate useful data mining concepts by presenting them with respect to particular concrete applications.

Essentially, the applications in this thesis can be classified along two main routes. One route seeks patterns in *measurement* data to ascertain information about the system. The other route seeks patterns in *calculated* data, which includes data produced by a model and may also include measurement data that has undergone transformations. Calculated data can allow benefits such as aggregation for system reduction. In the chapters to follow, Chapter 2 provides an overview of key background concepts related to data mining which are important to our work. The subject of modeling is an important component of our work, relevant for measurement as well as calculated data types. Essentially, we can use the data to learn about the system creating the data, while considering different levels of modeling assumptions. A common purpose of aggregation is to make an "equivalent" or a reduced model of the system, and clustering provides such an aggregation tool. Clustering power system objects, such as generators, can reveal insight about

the way certain objects behave together. The subsequent chapters each detail the use of data mining concepts in different types of power system applications. The chapters and applications presented are outlined below.

First, Chapter 3 focuses on identifying relationships from sensitivities, and two specific applications are presented. These involve the aggregation and control of reactive power devices and the identification of generator groups with the potential ability to exhibit local market power. The work in this chapter exhibits the usefulness of data mining techniques to identify relationships from model data (sensitivities). The analysis can reveal and quantify previously unrecognized relationships which are supported by the model. Sensitivities reveal the propensity of certain components to operate together. In summary, it is important to characterize and exploit "model" relationships along with those observed from measurement data. Relationships discovered in the model provide a reference for comparison of what is expected to be observed in the measured data. Monitoring the extent to which relationships in the measurement data mirror those in the model data helps to assess the quality of both the model and the data.

While Chapter 3 recognizes the usefulness of model data for identifying important relationships between system components, Chapter 4 deals with the problem of identifying the model itself. A network model can be considered a particular type of pattern, and Chapter 4 deals with its estimation in three particular applications. The first application concerns the estimation of an equivalent network connecting PMU buses based on PMU data. The benefit of data over time is exploited by considering multiple snapshots of data to obtain information about the system. A reduced model connecting PMU buses is created from data. The second application concerns the creation of structure-preserving equivalents. Finally, the third application in Chapter 4 utilizes the data-over-time concept to estimate parameters of individual transmission lines rather than an equivalent model.

Then, Chapter 5 utilizes time-series data analysis techniques which are useful when dealing with real data. When estimating models from data, one must deal with the properties of real measurement data rather than idealized or simulation data. In such cases, the data has random and possibly systemic errors, and the model of the system may be unknown or incorrect. Real

data and its uncertainty present a number of challenges to the development of robust algorithms which need to deal with such data and still be able to correctly extract information. In Chapter 5, we study the associations of variables in time series data.

Another important component of our work is the ability to quickly reveal information about what occurred in the system based on measurement data. In Chapter 6, the use of oscillation monitoring data for event pattern recognition is presented. To obtain more value from data collected over time, it is useful to investigate the information contained in the dynamics of the system rather than only steady-state patterns. The goal is to find and exploit associations between the model content of oscillations which occur as the system changes state (ringdowns) and, in particular, to identify information about the causing event. Estimated modal content is used to classify events and estimate where in the system the event may be occurring.

In developing the applications described above, a key challenge has emerged which we have identified as critical to progressing new data-enhanced applications to the point of deployment in real world systems. This issue, in summary, is data quality. The availability of information regarding the quality of the data, as well as availability of the data itself, is a significant challenge. Chapter 7 deals with different aspects of the data quality problem. It is important to consider the impact of false data and make algorithms robust with respect to such data. Chapter 7 presents applications for the protection of data and applications and deals with how to find false data. Finally, conclusions of the thesis and suggestions for moving forward in this area are presented in Chapter 8. Many power system applications are in a position to benefit from such techniques.

2. Data Mining Background

Terms such as "data explosion" are being used more frequently to describe the situation of power systems with respect to data. The work in this thesis spans several multi-disciplinary research areas, all of which can be said to be related to the broad category of data mining. Data mining is indeed a broad subject, so it is important to provide some clarification. We introduce in this chapter some background on the topics this work covers, as well as background on what "conventional" data mining tools can already provide. Existing off-the-shelf solutions can provide subsequent value to the work described in this thesis; these tools are likely to be useful from a practical standpoint but are not the focus of this research.

Our focus with respect to data mining is to look for relationships or "patterns" which we can exploit for power systems analysis. An important point is that intractably many rules or patterns often exist in a dataset. Thus, results may contain patterns which are not what we intend to discover—such patterns may happen by chance or by overfitting of the model. Therefore, it is important to define what patterns we are interested in capturing and to focus on searching for "interesting patterns," rather than for "patterns" in general. The purpose of this chapter is to provide the necessary background and tools to accomplish this task.

In this chapter, we review several key background concepts in data mining necessary for our work. First, Section 2.1 describes common ways to represent the discovered "knowledge." Section 2.2 describes cluster analysis. Section 2.3 provides some essential background on the statistical analysis of data. Section 2.4 describes system identification methods which seek a model of the system from the data. Then, an overview of singular value decomposition (SVD), a multifaceted tool, is provided in Section 2.5.

2.1. Knowledge Representation

A *knowledge representation style* is the form in which results of a data mining algorithm or application are presented. Understanding the possible ways to represent knowledge—what type of solution to seek for a particular problem—is essential to understanding how to develop data mining applications for power systems. Specifying the style of the sought solution is important to decompose the problem into manageable pieces. A fundamental concept in data mining which

is often encountered in machine learning is the distinction between *supervised* and *unsupervised* learning. Classification is called supervised learning because it uses training samples with known outcomes or "classes." Three types of data are used for machine learning systems: training examples, test data, and real data. For both training data and test data, the outcomes are known. The training data is used to come up with the model for learning or classification, and the test data is used to validate the model. Classification techniques from the data mining field which are useful for pattern recognition are reviewed in [16]. Then there is real data for which the classification is not known. In unsupervised learning, or clustering, groups are formed without first knowing the classifications of the data. Cluster analysis is addressed in Section 2.4.

Supervised learning and common knowledge representation styles are introduced in this section to illustrate the benefits that existing data mining tools can already provide and how they can offer subsequent value to this work. As introduced in Chapter 1, the research in this thesis is not on applying off-the-shelf data mining techniques but on identifying key areas and applications in power systems where data-enhanced solutions are immediately apparent and tangible. It is expected that as work in this direction continues, more insight from the pure data mining field can be increasingly incorporated into power systems solutions.

Objects are described in a data mining context by their *attributes*. A set of attributes which occur together is called an "item." In market basket analysis, items are the articles being purchased, and the goal is to look for associations among the purchases. Combinations of items meeting a pre-specified value for minimum coverage are called "item sets."

For supervised learning, two common knowledge representation styles are decision trees and rules. A decision tree comprises choices at each node and consequences in each branch. Decision trees can be constructed using scatter plots to show which two attributes do a "good job" splitting the classes at a particular level. Decision trees can be automatically generated for classification purposes by applying the following steps in a recursive manner: select an attribute to place at the root node and make one branch for each possible value. To construct *decision trees*, a measure of "purity" of the nodes denoted *information* is used, as in signal processing [17]. Top-down decision tree algorithms seek, at each stage, an attribute to split which best

separates the classes. The decision to split is based on the resulting *information gain*. Top-down induction of decision trees may also be called a "divide and conquer approach." The ID3 and C4.5 systems are well-known decision tree induction systems [18]. The "practical and influential" decision tree algorithm called C4.5 is an improvement over ID3 [19]. Ideally, the procedure ends with the objects in each leaf node belonging to the same classification, when the data cannot be split any further. Interpretation of the results simply requires following choices starting at the root node. Top-down decision tree induction is one of the most extensively researched topics with respect to data mining, and many variations on basic methods exist. However, such variations rarely produce much improvement in accuracy over diverse datasets [18].

In contrast to decision trees, *rules* are modular; each represents an independent "nugget" of knowledge. It is easier to add new rules than to modify a decision tree. While it is possible to generate rules from decision trees, the conversion is not trivial. Proper interpretation of the rules, including the order in which they are executed, is important because it can change how objects are classified. The problem is that as rules become more accurate, they also become more complex. Order-independent rules are advantageous and represent independent modules of knowledge. Furthermore, there is an issue of what to do when different rules lead to different conclusions for the same instance. Ordered decision lists prevent this ambiguity since execution would stop once a single rule is applicable to the instance. When rules result in inconsistent classification results, one solution is to simply give no conclusion. It is also possible to have an instance that cannot be classified by the rules at all. Comparatively, a decision tree prevents ambiguity in interpretation. A downfall of decision trees is that they cannot easily express *disjunction* that may be implied among the rules that govern a data set. Since a node in a decision tree performs a single test, rules can be much more compact.

Types of rules can be further labeled. *Association rules* "predict" attributes from other attributes. These may be *propositional rules*, which compare attributes with a constant. *Relational* rules compare attributes with each other; this often involves inductive logic programs. We are most interested in relational rules, but propositional rules can be useful if we need to compare power system values to predefined limits.

A simplicity-first methodology should be applied in analyzing practical datasets. *Covering algorithms* seek to generate rules which cover all instances of each class in turn. *Coverage*, or *support*, expresses the number of instances correctly predicted by a rule. *Confidence*, or *accuracy*, given by *p/t*, is the number of instances *p* that a rule correctly predicts as a proportion of all of the instances *t* to which the rule applies.

The *1R* algorithm can be used for association analysis and classification. The *1R* algorithm makes a rule for each value of each attribute based on the most frequent class, and the rules with the smallest error rates are chosen. The rule construction covering algorithm *PRISM* generates only correct or "perfect" rules by maximizing *p/t*. Any rule with accuracy less than 100% is "incorrect." Algorithms such as *PRISM* are outlined in [18]. The *apriori algorithm* finds frequent item sets in the first step, while the second step generates rules for the frequent item sets [18]. These algorithms are useful for determining which combinations of items are often encountered together.

Another common algorithm is *näive Bayes* which calculates the probability of an instance belonging to a class, given the evidence. Näive Bayes, based on Bayes' rule, "näively" assumes independence of the attributes, so the resulting probabilities are only valid if that assumption is true. Prior probabilities are required for each attribute in the evidence. Estimates of the prior probabilities are made from the training data. When a new instance is encountered, the probability of it belonging to each class is evaluated, and the instance is assigned to the class with the highest probability.

Rote learning is a form of *instance-based learning*, which works directly from examples. There is much work in the artificial intelligence community on this topic. The training instances are "memorized." As new instances are encountered, memory is searched for the instance that "most strongly resembles" the new one. Instance-based learning is "lazy" because it puts off the work until it is time to classify a new instance. Conversely, "eager" methods produce a generalization as soon as possible. "Nearest-neighbor" and "*k*-nearest neighbor" classifications are two basic instance-based learning schemes. The *k*-nearest neighbor method relies on the majority class of

the *k* nearest neighbors. Attributes may be weighted if some are more important than others. However, determining which attributes are the most important and deriving weights is itself a key problem. Also, it may not be possible to store all of the training instances, so another key problem is determining what to save. Regions in attribute space which are "stable" require only a few examples to clearly define the classification. One major drawback of instance-based learning is that it does not make explicit the structures that are learned, so it does not really provide a description of the patterns in the data.

The algorithms listed above are more favorable for use on data sets with nominal attributes, but can be extended to numerical attributes. In power systems, we are typically concerned with numerical attributes and models, where data is generated by a system with a known type of structure. When knowledge needs to be represented by a structure, *linear models* and *regression* can often satisfy this goal. Regression methods are concerned with deriving an expression to predict a numeric quantity. *Support vector learning* [20] seeks such models. Linear regression expresses the class as a linear combination of attributes with predetermined weights,

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k \tag{1}$$

where x is the class, $a_1 \ldots a_k$ are attributes, and $w_0 \ldots w_k$ are weights. The weights are chosen to best fit the training data attributes to their class. If the class label is a nominal value, one may find a function for each class where $x = 1$ for an object in that class and $x = -1$ otherwise. Data is *linearly separable* if it is possible to separate instances which belong to different classes by a hyperplane. The *perceptron learning rule* [18] is a very simple algorithm for finding such a separating hyperplane, and *Winnow* [18] is similar. Both methods are "mistake-driven," updating the weights only when a misclassified instance is encountered. Winnow is reportedly an "attribute-efficient learner," effective at homing in on the relevant features in a dataset. This is advantageous if the data set has numerous attributes where most are irrelevant.

Association analysis as described in this section is computationally expensive. Classification is a special case of association analysis, so algorithms which are sufficient for classification are often prohibitive for association analysis. For classification algorithms, classes must be assigned to

the training instances by the system designer. If the system designer is unable to label classes for the training data, then designating the classes is itself a part of the problem. In such a case, clustering may be a better approach, at least initially. Cluster analysis is considered in the next section. Effectively, the clusters found are the classes.

2.2. Cluster Analysis

Cluster analysis is another major part of data mining, often referred to as "unsupervised learning," since it does not require class labels to be known ahead of time. The goal of clustering algorithms is to group objects based on the information that describes them. Clusters can be considered classes, to which the clustering process automatically assigns objects. These algorithms are also encountered in the literature under the names of "numerical taxonomy," "vector quantization," and "learning by observation [21]."

The first important point to convey about clustering is that there is no "perfect" clustering algorithm. To better understand this, consider that the same set of points is easily clustered in different ways, as illustrated in Figure 1. The initial set of points is readily perceived as two, four, or six clusters.

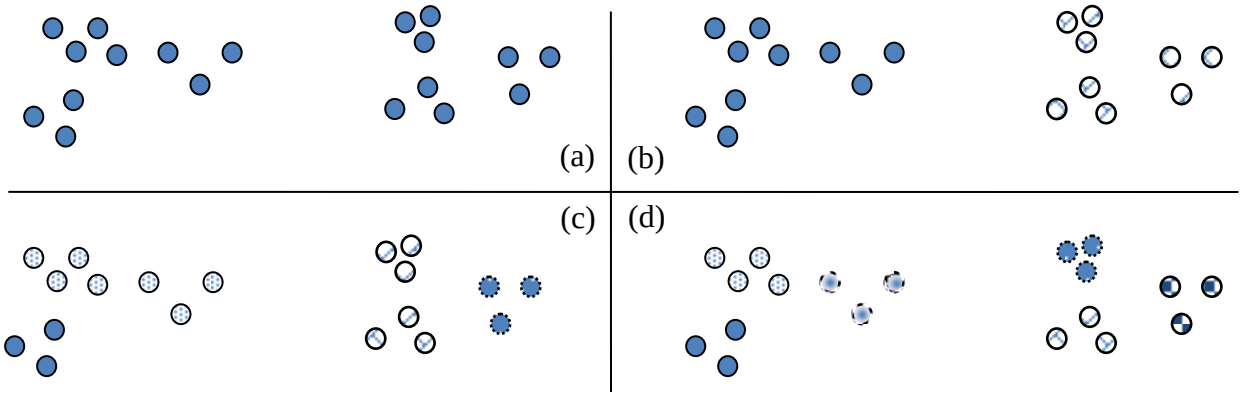

Figure 1. (a) Set of points, (b) Two clusters, (c) Four clusters, (d) Six clusters

A fundamental question arises: Which clustering is "correct?" The concept of a cluster is often difficult to define. Interpretations are made by the human visual system, but a precise definition

of a cluster is lacking. How data should be clustered often depends on the data itself and what information or results are desired.

There are different types of clusters. In *overlapping* or *non-exclusive* clusters, objects may be in more than one cluster. Some clusters are represented by a *prototype*, where members of the cluster are more similar to the prototype of their cluster than the prototype of another cluster. Clusters can also be defined as connected components in graphs, where they are connected to each other but not to objects in other clusters. In *well-separated* clusters, an object is more "similar" to every other object in its cluster than to objects in other clusters; this is satisfied when distinct groups of points are located far from each other. *Partitional* clustering is a division of the objects into non-overlapping subsets, so any object may only be in one cluster. In incomplete or *partial* clusterings, not all objects may be assigned to a cluster; unassigned objects may be regarded as noise. These are some of the most important concepts and definitions with respect to clustering for this thesis.

Since judging clustering results can be subjective, it is useful to develop clustering metrics to better quantify similarity. One common metric is the *sum of the squared error (SSE)*, also referred to as *scatter*, which is defined as

$$SSE = \sum_{i=1}^{M} \sum_{x \in C_i} dist\left(c_i, x\right)^2 \tag{2}$$

where M is the number of clusters, C_i is the i-th cluster, and c_i is the centroid of the i-th cluster [22]. Typically, the centroid is the mean of the points in the cluster. The *SSE* is related to the *radius* of a cluster i,

$$R_i = \frac{\sum_{x \in C_i} dist(c_i, x)}{K} \tag{3}$$

a measure of cluster tightness, where K is the number of elements in the cluster. Another measure of cluster "quality" is called *category utility* [18]. Unless otherwise specified, our algorithms use the Euclidean distance,

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{k=1}^{n}\left(x_k - y_k\right)} \tag{4}$$

but any distance measure may be used and the algorithms may be adapted accordingly.

The *SSE* in (2) can be considered an objective function, and the process of clustering can be considered an optimization problem. The objective is then to minimize the sum of the squared distance of each object to its centroid. According to [18], the following problem is nonconvex, discrete, and has an abundance of local minima: "Given the desired number of clusters K and a dataset of N points, and a distance-based measurement function, find a partition of the dataset that minimizes the value of the measurement function." There is typically no way to find a global minimal solution without trying all possible partitions, an approach called *exhaustive enumeration*. With approximately $K^N/K!$ ways of partitioning the data, exhaustive enumeration is infeasible unless N and K are quite small. Another computationally intractable method called *iterative optimization* specifies an initial partition, then tries all possible swapping of points from one group to another to see if the objective function improves. Usually, even if an optimal solution is found, it is still only locally optimal. Thus, in attempt to find solutions to this problem, numerous clustering algorithms exist; only a few appear in this document. Clustering is used in various capacities throughout this work. Several algorithms are examined; other algorithms can be easily understood from these. In particular, Chapters 3 and 4 illustrate the use of cluster analysis to find and exploit relationships in power systems.

2.3. Statistical Analysis of Data

Statistical analysis provides another perspective of data mining. Since statistical properties of the data can be employed for the purpose of extracting information about the system, it is sensible to consider statistical techniques as data mining techniques. As will become clear, many seemingly different data mining methods are in fact quite similar and complimentary. This is yet

another perspective with which to approach the same problems. Statistical methods of pattern recognition are reviewed in [9]. As with the other types analysis in this chapter, statistical analysis is important for a number of applications throughout this thesis, particularly in Chapters 5, 6, and 7.

Here we provide some necessary background from statistical analysis concerning averages and variances. When dealing with data, the average of a series of measurements is an expected value of that quantity as a random variable. The expected value of a random variable x is

$$E(x) = \sum_i x_i p_i \tag{5}$$

where p_i is the probability that the random variable x takes on each value x_i. In addition to its average value, another important aspect of a distribution is its dispersion or how spread out the data is around the average value. This is the expected value of the square of the deviation of x from its average value. Dispersion relates how far, on average, each data point deviates from the mean. The variable to reflect this dispersion is the variance:

$$V(x) = E\left(x - E(x)\right)^2 \tag{6}$$

The standard deviation is the square root of the variance and has the same units as the measurement units of x.

$$\sigma_x = \sqrt{V(x)} \tag{7}$$

The variance is also the second moment, and the average is the first moment. In general, other moments of the data can be found. The third moment is related to the "skewness" of the distribution. In theory, if all of the moments of a distribution are calculated, they provide a complete characterization.

It is important to note that each sample is merely a *sample* of the population. For an infinite population, without an infinite number of samples (all items in the population), one cannot expect a completely accurate representation of the population. Histograms provide a *sample-image* of the shape of a distribution curve from its samples. While a histogram provides some insight about the general shape of a frequency distribution curve, one must be cautious about making conclusions based on few samples. Thus, the way the samples are selected from a population is important and is itself a science [23]. The most basic method is random sampling. The key point is that from samples we can merely achieve *estimates*.

Thus, estimates of the mean μ and variance σ^2 may be computed from the data using the following relationships:

$$\hat{\mu} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{\sum_i x_i}{N} \tag{8}$$

$$\hat{\sigma}^2 = \frac{\sum_i \left(x_i - \hat{\mu}\right)^2}{N-1} \tag{9}$$

The estimate of the mean (8) is simply the arithmetic average $\bar{x}$. The estimate of the variance (9) contains an *N-1* term in the denominator which denotes the number of degrees of freedom in the estimation. The reason there are not *N* degrees of freedom is because any one of the residuals, x_i - $\bar{x}$ in (9), can always be exactly calculated from the others since the value of $\bar{x}$ is defined by (8).

The normal or Gaussian distribution, prominent in statistical theory and in practice, is completely characterized by its mean and variance.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{10}$$

The standard deviation σ is the distance from the location of the mean μ to the point of inflection on the distribution curve. The normal distribution is especially important as a result of the

central limit theorem which states that under certain conditions, as sample size increases, the sample mean becomes normally distributed about the population mean.

2.4. System Identification

Although perhaps not widely recognized, system identification is an important type of data mining. The fundamental purpose of system identification is to infer models from observations by studying the properties of observations. Models are an attempt to link observations together in order to describe a "pattern" [24]. Thus, system identification tools are useful to approach certain problems in data mining. Using data mining methods for finding linear models and regressions, as discussed in Section 2.1, is the same problem as solving a system identification problem. Models constructed by system identification are intended to serve certain purposes and must be constructed accordingly. The techniques used depend on the type of model that is sought.

In power systems, system identification helps to solve problems such as state estimation, model parameter validation, and identification of topology errors. System identification concepts are valuable throughout this work, especially for the applications described in Chapter 4 and Chapter 6. In power systems, system identification has historically played, and continues to play, a significant role. It often seems to be the area which makes the most sense to use. Background on least squares estimation, a general tool, and power system state estimation are provided in Section 2.4.1 and Section 2.4.2 respectively.

2.4.1. Least Squares Estimation

Least squares estimation is a common technique in many areas of science and engineering. It is a practical tool with numerous potential uses in power systems. The method is simple, and provides a natural starting point for many types of analysis. Even system identification methods which appear more complicated can typically be interpreted in terms of extensions to the simple concept of least squares estimation. Hence, it is worth describing the method briefly, for a general case.

Rather than just mentioning the matrix manipulations needed to arrive at the solution, which is trivial, here we instead focus on where that procedure comes from. Further discussion on this topic will be withheld until Chapter 5, where time-series data is considered in more detail. The key point here is to give a warning that one must be careful when applying least-squares methodologies. Assumptions of the method, as with any method, make it well-suited for some situations but not for others. If applied to the wrong kinds of problems, it may still produce a numerical answer, but it will not mean what is intended.

Mandel [23] describes least squares estimation in the context of the statistical analysis of data, an especially appropriate viewpoint for this work. The problem of system identification and least squares estimation can be considered an "adjustment of observations," where the goal is to *deduce a set of well-defined quantities in some optimum fashion from a collection of experimental values* [23]. The simplest case, where the value in terms of improved precision is obvious, is the case where several replicate experimental measurements are averaged to estimate the value of some quantity. Thus, least squares estimation lies at the intersection of statistical analysis of data, described in the previous section, and of system identification, discussed here. Once values for the desired quantities are determined, it is an additional challenge to assess the quality of such estimates. In the case of averages, the problem is simple to conceptualize; each point has an impact on the value of the estimate, such that if one point is far off, it will cause the average to be a poor representation of the true quantity. These issues can result in quality problems associated with derived data and estimates, discussed more in Chapters 5 and 7. Quality problems often arise due to inconsistencies in the measuring processes and due to difficulty in determining the "best" way to combine the values obtained into one final estimate.

The arithmetic average is not necessarily the best value. Weighted least squares (WLS) is a weighted average where the weights are determined based on measurement variances. The problem is to select values of coefficients such that the estimated quantity has the highest precision. This means the variance of the estimated quantity must be minimized. The following three conditions characterize WLS estimation: (1) the best value is assumed to be a linear combination of all three measurements, (2) the best value is assumed to be free of systematic errors, so it is an unbiased estimate (the expected value is equal to the true value) and (3) the best

value has minimum variance. Thus, taking the average of measurements is a case of linear unbiased minimum-variance estimation, as described in [23].

To illustrate the concept of least squares estimation, consider the following matrix equation:

$$\mathbf{Ax} = \mathbf{b} \tag{11}$$

$$\mathbf{r} = \mathbf{Ax} - \mathbf{b} \tag{12}$$

for which least squares estimation minimizes the sum of the squared residuals in (12) and results in the following estimate (109):

$$\mathbf{x} = \left[\mathbf{A}^{\mathrm{T}}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \tag{13}$$

where **x** denotes the vector of estimated quantities.

In (12), the residuals **r** are estimates of the errors. When measurements have different weights, the weighted sum of the squares of residuals is minimized instead. To derive the solution, one forms an objective function and takes the partial derivative with respect to each variable to be estimated:

$$S = \sum_{i=1}^{n} w_{ii} r_i^2 \tag{14}$$

If there are *k* variables to be estimated, this yields *k* equations, the solution of which yields the *k* unknowns. In matrix form, the result is

$$\mathbf{x} = \left[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{b} \tag{15}$$

An interesting and useful byproduct of least squares estimation is that it also provides an estimate of the *variances* of the unknown variables. This is because the estimate **x** is a linear function of the measurements **b**. Assume we know the weights or variances of the

measurements, or at least their relative proportions. The variances of all of the measurements can be represented in terms of σ^2, and the weights w_i are the reciprocal of the variances. Then, the law of propagation of errors provides a means to estimate the variance of each element of **x** [23, p. 145]:

$$\hat{\sigma}^2 = \frac{\sum_i w_i r_i^2}{N - p} \tag{16}$$

where N is the number of observations (measurements) and p is the number of estimated parameters. One can also estimate the confidence intervals for the parameter estimates obtained [23].

2.4.2. Power System State Estimation

The concept of using data in an intelligent manner to learn about the power system has its roots primarily in state estimation. A brief overview of the state estimation problem for power systems is provided here. This section provides background for other parts of this work. Concerns about false data in power systems, as discussed in Chapter 7, often arise with respect to its impact on the state estimation solution. Additionally, building equivalent models, as in Chapter 4, is closely related to the state estimation problem. State estimation is a well-studied field of its own, and an attempt is not made here to give a full coverage of the subject.

The power system state estimation problem is to estimate the current conditions of the network based on snapshots of real-time measurements. Estimated quantities principally include bus voltage magnitude **V** and angle **θ**, the power flow state variables. If it is possible to estimate **V** and **θ** throughout the network based on measurements, the network is said to be observable [25]. Once these state variables are known, all other quantities such as currents and real and reactive line flows can be computed. Many resources in the literature describe power system state estimation. Detailed coverage can be found in [26] and more in [27] and [28]. Topology error detection for state estimation is considered in [29], [30] and [31].

State estimation provides operators with useful information about the system, thus increasing situational awareness. The state estimator runs approximately every five minutes. The estimate is computed using known model equations which relate the power system measurement quantities to the unknown states which are to be determined. The model equations contain static network data including line and bus configuration (status data) as well as physical network parameters such as line impedances. A topology processing phase precedes state estimation and reconciles status data with a more detailed node-breaker model to obtain the bus-branch model which is ultimately used by the state estimator. Once the network topology is known, conventional state estimation assumes it to be correct.

In general, the power system state estimation problem is described by (19) as follows [26]. State estimation seeks to minimize the difference between the actual measurements **z** and the computed values of measurements **h**(**x**), where **h**(**x**) is a function relating measurements to state variables. Typical power system measurement quantities include real and reactive line flows, $\mathbf{P}_{ij}$ and $\mathbf{Q}_{ij}$ respectively,

$$\mathbf{P}_{ij} = \mathbf{V}_i^2\left[-\mathbf{G}_{ij}\right] + \mathbf{V}_i\mathbf{V}_j\left[\mathbf{G}_{ij}\cos\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right) + \mathbf{B}_{ij}\sin\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right)\right] \tag{17}$$

$$\mathbf{Q}_{ij} = -\mathbf{V}_i^2\left[-\mathbf{B}_{ij} + b/2\right] + \mathbf{V}_i\mathbf{V}_j\left[\mathbf{G}_{ij}\sin\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right) - \mathbf{B}_{ij}\cos\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\right)\right] \tag{18}$$

where $\mathbf{G}_{ij}$ and $\mathbf{B}_{ij}$ denote the admittance elements connecting buses i and j, from $\mathbf{Y}_{\mathbf{bus}} = \mathbf{G} + j\mathbf{B}$, where $\mathbf{I} = \mathbf{Y}_{\mathbf{bus}}\mathbf{V}$. The constraints **g**(**x**) and **c**(**x**) represent the fact that the state estimates must also be a valid power flow solution.

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{z} - \mathbf{h}(\mathbf{x})) \\ &\text{such that } \mathbf{g}(\mathbf{x}) = 0 \\ &\qquad\qquad \mathbf{c}(\mathbf{x}) \leq 0 \end{aligned} \tag{19}$$

The equality constraints **g**(**x**) represent power balance; they enforce the fact that power generation must always equal the load plus the losses. The inequality constraints **c**(**x**) represent

physical limits such as the reactive power limits of generators which are also handled within the power flow problem.

Power system state estimation is typically an overdetermined problem since there are more measurements available than are needed to solve for the unknown voltage magnitudes and angles. The following quadratic objective function is used in the weighted least-squares (WLS) approach to calculate the best estimate of the power system state:

$$f = (\mathbf{z} - \mathbf{h}(\mathbf{x}))^T \mathbf{W} (\mathbf{z} - \mathbf{h}(\mathbf{x})) \tag{20}$$

The measurement variance inverses are typically used as the weights in matrix **W**. The Gauss Newton method provides an iterative state estimation procedure, [26]:

$$\left(\mathbf{H}^{\mathbf{T}} \mathbf{W}^{-1} \mathbf{H}\right) \Delta \mathbf{x}^{\nu} = \mathbf{H}^{\mathbf{T}} \mathbf{W}^{-1} \mathbf{r} \tag{21}$$

$$\mathbf{x}^{\nu+1} = \mathbf{x}^{\nu} + \Delta \mathbf{x}^{\nu} \tag{22}$$

where the residual $\mathbf{r} = \mathbf{z} - \mathbf{h(x)}$, and the Jacobian matrix $\mathbf{H}(\mathbf{x})$ is associated with the measurement equations of a WLS state estimator based on an AC (nonlinear) network model. A detailed account of the contribution of different types of power system measurements to the Jacobian $\mathbf{H}(\mathbf{x})$ is given in [26].

The model described above is the non-linear AC system model. Comparatively, a simplified DC model (23) assumes linear functions linking measurement to states,

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e} \tag{23}$$

which is commonly used in papers addressing false data injection attacks on state estimation [32].

2.5. Singular Value Decomposition

Singular value decomposition (SVD) is a matrix decomposition technique which is also a useful data mining tool in many respects. In this work, SVD appears in Chapters 3 and 5, so an overview of SVD is provided. The motivation, the outcome, and the manner of applying SVD can vary considerably depending on the application. The fact that SVD is such a multi-faceted tool is one of its most attractive features. In fact, SVD fits several different interpretations, each of which lends value in a different way to the present work. Several useful interpretations of SVD are now briefly summarized.

First, in a geometrical interpretation, SVD represents a unit sphere in the input space and an ellipse in the output space [33]. The SVD expresses a matrix as $\mathbf{A} = \mathbf{USV}^{\mathbf{T}}$, where **S** is the diagonal matrix of singular values of **A**. According to [34], computation is of order $O(nm^2)$, so ideally m should be small. Consider matrices that are 2x2, so they may be visualized in two dimensions. Then, **V** is an orthonormal basis in 2D, and those vectors define a circle. Since **U** and **V** are orthogonal matrices, their inverse is their transpose, so multiplying both sides by $\mathbf{V}^{\mathbf{T}}$ gives $\mathbf{AV} = \mathbf{US}$. This multiplication has the interpretation that **A** maps a circle **V** to the ellipse given by **US**, as illustrated well by Figure 2 from [33].

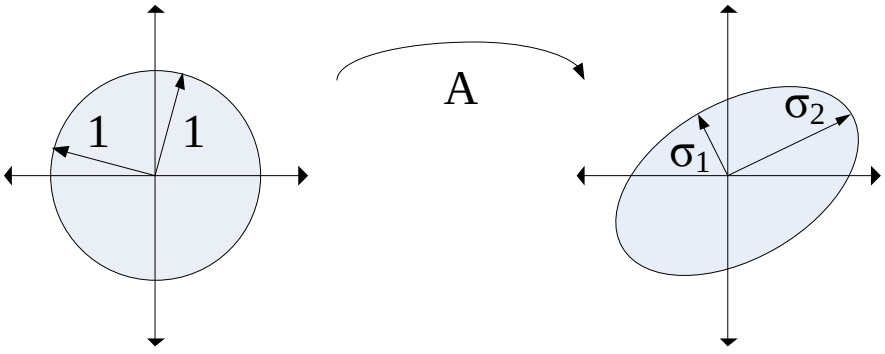

Figure 2. Geometric view of SVD [33]

Since **U** is also an orthogonal matrix, it can be interpreted that the columns of **U** are the normalized basis vectors for an ellipse (right side of Figure 2), whereas the axes lengths are given by the singular values or the diagonal entries of **S**. In higher dimensions, **A** maps a hypersphere **V** into a hyperellipse **US**. There are many references about this interpretation of

SVD, especially in the computer graphics area. In computer graphics, the SVD $\mathbf{A} = \mathbf{USV}^{\mathbf{T}}$ can be thought of as a rotation matrix × a scale matrix × a rotation matrix [33]. In SVD, **V** captures patterns in the attributes (columns) of the data (assuming each row in **A** is an object) and **U** captures the patterns in the objects (rows), and a new data matrix can be obtained from **A'**=**A***[columns of **V** corresponding to highest singular values]. More mathematical details and explanations are found in [35].

Properties of SVD make it a useful tool for solving linear systems. In particular, SVD exhibits a useful relationship for finding a matrix inverse; since $\mathbf{A} = \mathbf{USV}^{\mathbf{T}}$ then $\mathbf{A}^{-1} = (\mathbf{USV}^{\mathbf{T}})^{-1} = \mathbf{VS}^{-1}\mathbf{U}^{\mathbf{T}}$. SVD is also used to calculate rank, which is the number of non-zero singular values of a matrix. When solving equations of the form $\mathbf{Ax}=\mathbf{0}$ for **x**, SVD is especially valuable. This interpretation of SVD is significant for the market power potential identification algorithm of Chapter 3. Any such non-zero **x** belongs to the null space of **A**. Thus, **x** is a right singular vector of **A** corresponding to a singular value of zero in **A**. If **A** has several singular values which are zero, the corresponding right singular vectors, which correspond to columns in **V**, are a basis for the null space of **A**, and any linear combination of these basis vectors gives an **x** which satisfies $\mathbf{Ax}=\mathbf{0}$. If the rank of matrix **A** is *r*, the first *r* columns of **U** are an orthonormal basis for the range space of **A** and the last *(n-r)* columns of **V** are an orthonormal basis of its null space [36], [34]. It is also possible to use SVD to find projections [34], such as onto the null space or the range space of A. Additionally, SVD can be used to solve least squares problems. Consider that min $\|\mathbf{Ax}-\mathbf{b}\|^2$ is solved by $\mathbf{x} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{b}$. Rewriting this in terms of the SVD, the resulting matrix is the pseudoinverse of **A**.

In data mining literature, SVD is often used in a form referred to as principal component analysis (PCA). The relationship between SVD and PCA is briefly summarized. In PCA, the mean of the data is removed, whereas in SVD it is not. PCA performs eigen-analysis on a covariance matrix, formed by calculating the covariance between the columns of the data matrix. Covariance calculations are useful for revealing relationships between power system variables, discussed in Chapter 5. PCA finds the orthogonal basis that best represents the given data set. If *d* is the dimension of the original data set, PCA finds a best approximating plane of dimension $d'<d$ by taking the first *d'* eigenvectors and examining the subspace they span. In PCA, the

strong *components* are simply eigenvectors that correspond to large eigenvalues. This indicates that the data exhibits a clear preferable direction. The goal is to find new dimensions that better capture the variability of the data. The first dimension captures the most variability, and the second dimension is orthogonal, and it captures the next most variability, etc. The decomposition tends to identify the strongest patterns in the data. Also, most of the variability of data can often be captured in only a few dimensions. The decomposition can be used to eliminate much of the noise in the data as well as for dimensionality reduction. PCA can be used for facial recognition to match faces in a database from their PCA basis and in animation compression [33]. In [37], PCA is applied to study power flow data.

3. Identification of Patterns from Sensitivities

This chapter describes the use of data from a model, in the form of sensitivities, to identify patterns or discernible relationships between certain system components. Sensitivities are the linearized effect of one variable to another variable, and they describe the impact that a small perturbation of a variable has on the rest of the system [38]. Studying sensitivities reveals a considerable amount of information about expected patterns. Sensitivities indicate the space of patterns or relationships which may occur in the real system by revealing how variables are coupled to each other. There is considerable insight to be gained from sensitivities, which is valuable throughout this thesis. Under certain conditions, the sensitivities determined from empirical and analytical approaches, both reviewed here, are the same.

Sensitivities are analytically obtained based on the equations of the power network model. For example, in (24) the sensitivity S_α can be found by calculating the partial derivatives of P with respect to q_α.

$$S_\alpha = \frac{\partial P}{\partial q_\alpha} \tag{24}$$

Sensitivities originate from nonlinear equations which are linearized around an operating point. If the derivatives of sensitivities (24) are evaluated at an operating point x_0, the approximation may only be valid near that point. That is, if Δq_α is too large, (25) may not hold.

$$\Delta P = \left[S_\alpha \right] \Delta q_\alpha \tag{25}$$

The exception is the simplest case, where the system is linear, and then (25) is exact. Thus, analytical computation of sensitivities requires knowledge of the model and also requires knowledge of the system operating point. A more detailed description of sensitivities in power systems is found in [39].

Empirically, sensitivities may be obtained by considering the change of a variable, q_α, by a small amount Δq_α, and then observing the change in the output variable P. The resulting change in

output is obtained by measurement or by running a simulation so that it is possible to observe the change in P. If the change is given by ΔP, then sensitivity S_α may be calculated as follows:

$$S_\alpha = \frac{\Delta P}{\Delta q_\alpha} \tag{26}$$

Thus, sensitivities can be verified by simulation or by measurement in the real system, provided that the input q_α is possible to change and that the output P is possible to observe.

In this chapter, applications are presented which find and utilize the relationships described by the sensitivities. The concept behind using sensitivities in this manner is first briefly explained. Then, two applications are presented. The first application uses sensitivities to aggregate and control reactive support devices. The second application aggregates and identifies generators with the potential for local market power.

3.1. Clustering Sensitivities for Control

Since both applications in this chapter involve clustering of sensitivities, this section provides an explanation of the concept. The principle is to use sensitivities to identify groups of devices. Specifically, clustering may be used to group elements, as introduced in Chapter 2. Clustering sensitivities can allow groups of smart grid devices to be formed which have the ability to respond together in predetermined ways when various scenarios occur. This involves studying the impact of devices on a control objective. The clustering of devices for control is possible based on the coupling of variables, as determined from the corresponding sensitivities. There is no limit to the types of devices which may be controlled in this manner.

Notable examples to which this theory applies include controllable loads and parameters. An observation is made in [39] that some real power line flows have the ability to be independently controlled, while others do not. Lines with highly coupled flows cannot be controlled independently; controlling one will always control the other. Exploring this observation further, the ability of certain lines to exhibit this *independently controllable* property is discernible from the relationships in the sensitivities. For example, computing the sensitivities of real power line

flows to a change in control parameters and examining the couplings between sensitivities will reveal this property.

3.1.1. Coupling Index

Specifically, the coupling between variables, as determined from sensitivities, can be found using the coupling index (CI). The CI is also called cosine similarity and is given in this application by the cosine of the angle between two row vectors $\mathbf{v_1}$ and $\mathbf{v_2}$ of the sensitivity matrix:

$$\cos\theta_{v_1 v_2} = \frac{\mathbf{v_1} \cdot \mathbf{v_2}}{\|\mathbf{v_1}\|\|\mathbf{v_2}\|} \tag{27}$$

The CI has values between -1 and 1. When the CI has an absolute value of 1, the angle between the vectors is zero, and there is complete correlation, either positive or negative, between the vectors. When the CI is zero, the row vectors are orthogonal, and the vectors behave independently. This coupling index is one way to identify patterns in sensitivities which denote the ability to control certain sets of devices together. More generally, for example, if we consider all line flows, then clustering can find all sets of independently controllable line flows. It is necessary to control coupled components together since they cannot be controlled independently. Couplings in any set of sensitivities may be discovered in this way. In the first application in this chapter, we use this analysis to identify voltages with the ability to be controlled independently via reactive power injection. Examining the couplings allows us to determine how independently controllable certain bus voltages are with respect to other bus voltages. Finding these relationships allows for simplification of control.

3.1.2. Simplified System Models for Dynamic Exactness

Simplifying system models and reducing computation are often important for control problems and applications. Two classes of methods, aggregation and perturbation, can address the model simplification problem [40]. Aggregate models simplify a mathematical model of a system by introducing a coarser representation to retain some key qualitative properties of the system. In particular, aggregation is useful for implementing decentralized control methods in large-scale

systems [40]. While centralized control is common, the survey paper [40] responds to the "failure of centrality" due either to lack of centralized information or lack of centralized computing capabilities. The full system S_1 can be represented by

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) + \mathbf{B} \cdot \mathbf{u}(t) \tag{28}$$

and the aggregated system S_2 is

$$\dot{\mathbf{z}}(t) = \mathbf{F} \cdot \mathbf{z}(t) + \mathbf{G} \cdot \mathbf{u}(t) \tag{29}$$

where it is required that the following equation is satisfied for all time:

$$\mathbf{z}(t) = \mathbf{C} \cdot \mathbf{x}(t) \tag{30}$$

For a linear system, "dynamic exactness" is obtained when certain conditions are satisfied [40]. The aggregate state vector $\mathbf{z}(t)$ must be a linear combination of certain modes of $\mathbf{x}(t)$. The eigenvalues of **F** will be the eigenvalues of **A** corresponding to the retained modes. In fact, matrix **C** must be chosen such that it creates zeros in the input-output relationship between $\mathbf{u}(t)$ and $\mathbf{z}(t)$ to cancel the poles in the relationship that should not be retained. Further information about obtaining a minimal realization that satisfies the definition of dynamic exactness (ways to choose **C**) is found in [40]. This is a way to create equivalents which preserve certain modal responses.

Coupling methods are discussed in the context of perturbation methods for model simplification [40]. Weak coupling methods are nonsingular perturbations, where perturbations are on the right-hand side of the differential equations, as in the following:

$$\begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \varepsilon \mathbf{A}_{12} \\ \varepsilon \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1(t) \\ \mathbf{u}_2(t) \end{bmatrix} \tag{31}$$

where ε represents a small positive parameter. This is essentially the cosine similarity (27). If ε were exactly zero, the system would be completely decoupled and many computations would be reduced. Singular perturbations represent strong coupling. If the system has slow and fast components,

$$\dot{\mathbf{x}}_1(t) = \mathbf{A}_{11} \cdot \mathbf{x}_1(t) + \mathbf{A}_{12} \cdot \mathbf{x}_2(t) \tag{32}$$

$$\varepsilon \dot{\mathbf{x}}_2(t) = \mathbf{A}_{21} \cdot \mathbf{x}_1(t) + \mathbf{A}_{22} \cdot \mathbf{x}_2(t) \tag{33}$$

and ε is taken to be zero, the second equation becomes an algebraic equation for $\mathbf{x}_2(t)$. This yields the "strong coupling" approximation,

$$\dot{\mathbf{x}}_1(t) = \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)\mathbf{x}_1(t) \tag{34}$$

$$\dot{\mathbf{x}}_2(t) = \left(-\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)\mathbf{x}_1(t) \tag{35}$$

The fast modes associated with $\mathbf{A}_{22}$ are discarded. The other modes are approximated by the modes of the system matrix of (34). We are often interested in poorly damped, low-frequency modes, and since we deal with large scale power systems, singular perturbation methods are useful for providing a means to reduce computation. These algebraic variables $\mathbf{x}_2$ can be representative of the power flow solution while the differential equation variables $\mathbf{x}_1$ can be representative of the modeled machine dynamics.

3.2. Aggregation and Control of Reactive Power Resources

The selective grouping of reactive support devices for providing wide-area voltage support, developed in conjunction with this thesis, is outlined in [41] and [42]. This application is fundamentally built upon extracting information from sensitivities as well as from measurement data. The presented framework encompasses devices at the residential level with the ability to provide reactive support. Reactive power may be supplied by converter-connected devices [43] [44], which includes many types of distributed resources, as well as by machines. The power buffer concept [43], [45], [46] allows the power electronics to present a desired behavior to the grid. Such a capability is not currently utilized. Integrated end-user reactive-power-capable

devices can be called upon to provide voltage support via a secure communications infrastructure outlined in Chapter 7.

Sensitivities are used to determine effective locations in the transmission system and to control reactive power resources at those locations, hereafter denoted *Q-Controlled (Q-C) buses*. Reactive support groups are found from cluster analysis of sensitivities which parallel the secure communications architecture, described in Chapter 7. The proposed framework is envisioned to coordinate distributed devices to correct voltage violations in their local area, using secure, authenticated messaging to achieve the control. Details and results are reported in [41] and [42].

3.2.1. The Voltage Control Problem

In this section, we briefly describe the voltage control problem addressed in this application. To maintain *operational reliability* [47], the system should be able to withstand the loss of any one credible contingency [48]. However, as power systems become more heavily loaded, the number of limit violations and unsolvable contingencies may increase. The effects of an outage may perhaps lead to a voltage collapse, whereby voltages progressively decline until it is no longer possible to maintain stable operating voltages [49]. It is well known that available reactive power resources can be used to control voltages and thus make the system less vulnerable to voltage instability.

The August 14, 2003, blackout establishes the importance of this work. The report [49] states that inadequate reactive power supply was a factor in most previous major North American blackouts, and that prior to the 2003 blackout, a lack of adequate dynamic reactive reserves combined with a lack of knowledge about critical voltages and maximum import capability left the system in a vulnerable state. In [41], the reconstructed 2003 blackout case just prior to the blackout is compared to the case after correcting the power factor to unity at just five buses, and the voltages are seen to considerably improve. The purpose of this work is to facilitate corrective control, which directs the system to a new stable equilibrium point shortly following a severe disturbance [50], [51], as well as preventative control, which is carried out before any instability occurs. Available corrective control actions are not always practical and may require the involvement of a large number of buses in the system [52]. Generator re-dispatch is time-

constrained by ramp limits, and load shedding is often costly and used as a last resort. The most effective and the least costly solutions [53] can be determined and should be enacted within an allotted time frame to restore the system to a secure system state. Example reactive power controls include transmission line switching [54], [55], [56], flexible AC transmission system devices (FACTS) [57], [58], [59], and synchronous condensers. Currently, control actions mostly occur at the substation level, whereas this application facilitates a comprehensive form of reactive power control that goes all the way to the end-user.

The sensitivities of voltages to reactive power injections are fundamental for determining appropriate locations and setting reactive power outputs of Q-C buses. The negative inverse of the power flow Jacobian **J** describes the way the state variables $\mathbf{s}_{(\boldsymbol{\theta},\ \mathbf{V})} = [\boldsymbol{\theta},\ \mathbf{V}]^{\mathrm{T}}$ change in a solution of the power flow due to bus power injection mismatch.

$$\boldsymbol{\Delta s}_{(\boldsymbol{\theta},\mathbf{V})} = \left[\mathbf{-J}\right]^{-1} \cdot \mathbf{f}_{(\mathbf{p},\mathbf{q})} \tag{36}$$

In (36), vector $\mathbf{f}_{(\mathbf{p},\mathbf{q})} = [\boldsymbol{\Delta p},\ \boldsymbol{\Delta q}]^{\mathrm{T}}$ expresses power balance which must equal zero at solution. Elements of **Δp** and **Δq** respectively are $\Delta p_i = P_{i,calc}\text{-}(P_{i,gen}\text{-}P_{i,load})$ and $\Delta q_i = Q_{i,calc}\text{-}(Q_{i,gen}\text{-}Q_{i,load})$ where the calculated power injections,

$$P_{i,calc} = V_i \sum_{j=1}^{n} V_j \left[G_{ij} \cos\left(\theta_{ij}\right) + B_{ij} \sin\left(\theta_{ij}\right) \right] \tag{37}$$

$$Q_{i,calc} = V_i \sum_{j=1}^{n} V_j \left[G_{ij} \sin\left(\theta_{ij}\right) - B_{ij} \cos\left(\theta_{ij}\right) \right] \tag{38}$$

should equal the specified net power injections. Let $\mathbf{Q_s}$ be the vector of specified net reactive power injections at each bus *i*, so $Q_{s,i} = Q_{i,gen}\text{-}Q_{i,load}$. Then, the sensitivity of voltage magnitude **V** to specified reactive power $\mathbf{Q_s}$ is given by the block matrix $\boldsymbol{\Lambda}_{\mathbf{VQ}}$ of $\mathbf{J}^{-1}$:

$$\left[\mathbf{J}\right]^{-1} = \begin{bmatrix} \dfrac{\partial \boldsymbol{\theta}}{\partial \mathbf{P_s}} & \dfrac{\partial \boldsymbol{\theta}}{\partial \mathbf{Q_s}} \\ \dfrac{\partial \mathbf{V}}{\partial \mathbf{P_s}} & \dfrac{\partial \mathbf{V}}{\partial \mathbf{Q_s}} \end{bmatrix} = \begin{bmatrix} \Lambda_{\boldsymbol{\theta P}} & \Lambda_{\boldsymbol{\theta Q}} \\ \Lambda_{\mathbf{VP}} & \Lambda_{\mathbf{VQ}} \end{bmatrix} \tag{39}$$

The sensitivities $\boldsymbol{\Lambda}_{\mathbf{VQ}}$ describe how voltage magnitude state variables change in a solution of the power flow due to a small change in specified reactive power injection at a bus. Effective locations of Q-C buses and control settings are determined based on the sensitivities (39).

The problem of controlling the voltages at specific buses is also an optimization problem, where the objective function f_1 is the sum of the differences of the bus voltages from their specified values, and M is the number of bus voltages targeted for control:

$$f_1 = \sum_{i=1}^{M} \left[\mathbf{V} - \mathbf{V}_{\mathbf{spec}}\right]_i^2 = \left[\boldsymbol{\eta}_i\right]^2 \tag{40}$$

The solution is to minimize f_1 subject to the power flow constraints and limits on equipment. The result gives the required reactive power injections at the Q-C buses such that the voltage profile $\mathbf{V}$ of the system equals the specified voltage profile $\mathbf{V}_{\mathbf{spec}}$ within some tolerance. The desired voltage profile may include any number of buses, even all buses, making this a global rather than a local optimization problem. The values of $\mathbf{Q_s}$ are constrained based on the individual loads, so the controllability of the reactive component of loads should be classified and incorporated into the Q-C bus selection algorithm. As the system changes, these classifications should be updated.

3.2.2. Selection and Aggregation of Q-C Buses

Sensitivities may be used to identify buses whose reactive power injections have a high impact on voltages of interest. Buses with higher sensitivities are able to provide more control, and are candidates for selection as Q-C buses, whereas buses with sensitivities of zero have no impact on the control objective. The sensitivities of f_1 to reactive power injection are given by the following vector:

$$\nabla f_1 = 2\boldsymbol{\eta}\boldsymbol{\Lambda}_{\mathbf{VQ}} \tag{41}$$

The *k* most effective locations correspond to the *k* elements of the sensitivity vector (41) which are furthest from zero.

From optimization theory, there are a number of ways to solve this problem [60]. A comparison of optimization methods is not a subject of this thesis. Here we use the steepest descent approach, and the $\mathbf{Q_s}$ values to minimize (40) are solved for using the sensitivities in (41). The use of the system sensitivities of f_1 (41) to solve the problem (40) requires a centralized control model.

To control voltage by adjusting loads and sources requires communicating control commands efficiently and securely. The corresponding required changes in $\boldsymbol{Q_s}$ must be communicated to all of the end-user devices. Requirements of this communication are described in context of a hierarchical control framework in Chapter 7, where the response is obtained by communicating the composite request $\boldsymbol{Q_s}$ to the next lowest level, which is then responsible for achieving the response amount from within its members.

The task of ensuring that communications are efficient and secure is simplified when the control effort can be focused on a subset of the controllers. Thus, for voltage control purposes as well as for communication purposes, buses are aggregated into *reactive support groups*. Reactive support groups are a key enabler of the framework. Each bus is associated with a reactive support group that consists of the devices determined *a priori* from a recent system model to have the greatest potential to control the affected node's voltage. Thus, only certain devices need to be considered as potential destinations for each control message.

Buses in a reactive support group are intended to help each other correct voltage problems. Sensitivities allow supporter buses to be identified for each voltage. To do this, each row of $\boldsymbol{\Lambda}_{\mathbf{VQ}}$ corresponding to each voltage is examined and the *l* highest values in the row are determined. Since each row is a bus voltage and each column is a reactive power injection, the *l* highest-

magnitude columns for a row give the *l* best Q-C supporters for that bus's voltage. This method provides redundancy since if there are *m* voltages of interest, there will be *m* reactive support groups where each region has *l* elements. Ensuring that each voltage has *l* supporters may be beneficial assuming we want reactive control groups to be at least some minimal size. Similar work has been done [61], [62], [63], [64], [65] to help gauge a system's proximity to voltage collapse. The makeup of the reactive support group for a particular node is determined based on analysis and clustering of sensitivities. Clustering groups the rows of $\mathbf{\Lambda_{VQ}}$ to identify voltages which are affected similarly by reactive power injections at Q-C buses. These clustering algorithms and results of the approach are presented in the following sections. To describe the algorithms in the most general context, the rows of $\mathbf{\Lambda_{VQ}}$ which represent buses to be clustered are referred to as *objects*.

3.2.3. Hierarchical Clustering Algorithm

Hierarchical clustering is one approach [22] that complements well the nature of the communications and security aspects of the problem since the granularity of control ranges from the transmission system to the end-users on the distribution system. In hierarchical clustering, the process may begin either with all items as individual clusters which are successively merged clusters or with all items in one cluster which are successively divided. The former type is referred to as agglomerative clustering and the latter is divisive. Hierarchical methods are flexible and non-parametric. Each level is defined by a distance threshold, where all items which are closer than this threshold to other items have been clustered together. One of the goals of hierarchical clustering is to be interactive and flexible, allowing the user to select a particular clustering for further examination. Instead of only one clustering result, we obtain many clusterings which form a hierarchy. Viewing this hierarchy can allow relationships in the system to be more easily understood.

Agglomerative hierarchical clustering schemes begin at the lowest level, with each element as a single cluster, and at each increasing level the closest clusters are merged. We can slice the hierarchy at different levels depending on how coarse or fine we want the clusters to be. The higher the level, the coarser the grouping becomes. At the highest level, all elements are in one cluster.

Agglomerative clustering relies on the use of distances between objects, which may be represented by a distance matrix $\boldsymbol{D}$. Elements D_{ij} give the distance between row i and j of $\boldsymbol{\Lambda_{VQ}}$. Any measure of distance may be used; here we use Euclidean distance. At each level, we find the most similar pair of clusters (r) and (s) by finding the minimum value D_{rs} in the current distance matrix. Then, clusters (r) and (s) are merged into one cluster. The rows and columns for (r) and (s) in the distance matrix are deleted and a new row is added for the new cluster (r,s). Any metric can be used to determine the distance of the new cluster (r,s) to each other cluster l, but typically the metric is $D_{(r,s)l} = \min(D_{rl}, D_{sl})$. That is, the distance between any two clusters is given by the minimum distance between any two elements of the clusters.

3.2.4. Quality-Threshold (QT) Clustering Algorithm

Another clustering algorithm which may be used to cluster the rows of $\boldsymbol{\Lambda_{VQ}}$ is the *quality threshold (QT) algorithm*, originally developed to cluster genes [66]. This method also uses the distance matrix $\boldsymbol{D}$. A threshold and a maximum cluster size are specified initially. For each row of $\boldsymbol{\Lambda_{VQ}}$, we build a candidate cluster that contains all other rows of $\boldsymbol{\Lambda_{VQ}}$ which are closer in distance than the threshold. The candidate cluster with the most elements becomes a true cluster. All the points in the true cluster are removed from further consideration. The process then iterates until all points belong to a true cluster.

The criterion which motivated the design of the QT algorithm is the need to form an unknown number of potentially large clusters which satisfy a "quality guarantee," meaning that the cluster diameters should not exceed a certain threshold. The distance of each object to every other object is required. A threshold and, optionally, a maximum cluster size are specified initially. For each object, a candidate cluster is built which contains all other objects closer than the threshold. An object not within the threshold of any other object forms its own cluster. The candidate cluster with the most elements becomes a true cluster. Once points are assigned to a cluster, they are removed from further consideration, so the process stops after all points are assigned. The QT algorithm is computationally intensive, $O(n^2)$ [67], as it requires a distance metric to be computed between all points.

3.2.5. Coupling Index (CI) Clustering Algorithm

An additional clustering method has been developed, particularly for this application, which we call the *coupling index* (CI) algorithm. The CI algorithm clusters objects based on their cosine similarity (27). The resulting clusters are approximately orthogonal to each other. While we developed the CI algorithm specifically for this voltage control application, it is implemented in a general way and may be applied to any foreseeable application. Its use in other applications appears later in this thesis. Here the purpose of clustering is to group buses based on the correlations of sensitivities of voltage with respect to reactive power injections. The CI accomplishes this by grouping together buses which are highly coupled to other buses in the same group. In general, the CI algorithm establishes what we henceforth denote as *simple clusters* and *reciprocal clusters* and (referred to in [41] as weak and strong clusters, respectively).

For each object, the algorithm identifies all other objects that are coupled by a CI with magnitude greater than a threshold. For an item i, these items form the simple cluster, A_i. The second stage identifies clusters where the items are extremely well matched to the other items in the cluster. At this stage, for each simple cluster A_i, the algorithm goes through the elements of A_i. For each element in A_i, it identifies items coupled by a CI greater than the threshold and places them in another group, A_j. If all the elements in A_i and A_j match, the set of matching elements is identified as a *reciprocal cluster*. Psuedocode is given in [41]. It should be noted that in the current algorithm, any strong cluster is also a weak cluster and points can belong to multiple clusters. Points that do not belong to a cluster are outliers and form their own cluster. The interpretation is that strong clusters are good candidates for the support groups, but weak clusters may need to be used to ensure that all objects are included in a group. There is room for refinement of the CI algorithm implementation.

3.2.6. Example Reactive Support Case

An example is provided in this section to illustrate the potential to control distributed reactive power devices to improve the voltage profile, also documented in [41], [42]. The study system considered is the IEEE 24-Bus Reliability Test System (RTS) [68] (Figure 3), which has low voltages around 0.95 per unit. The lowest voltages in the system (highlighted on the one-line

diagram) are at buses 3, 4, 8, 9, and 24. The voltage control problem is solved for this case in a centralized fashion using the methodology presented.

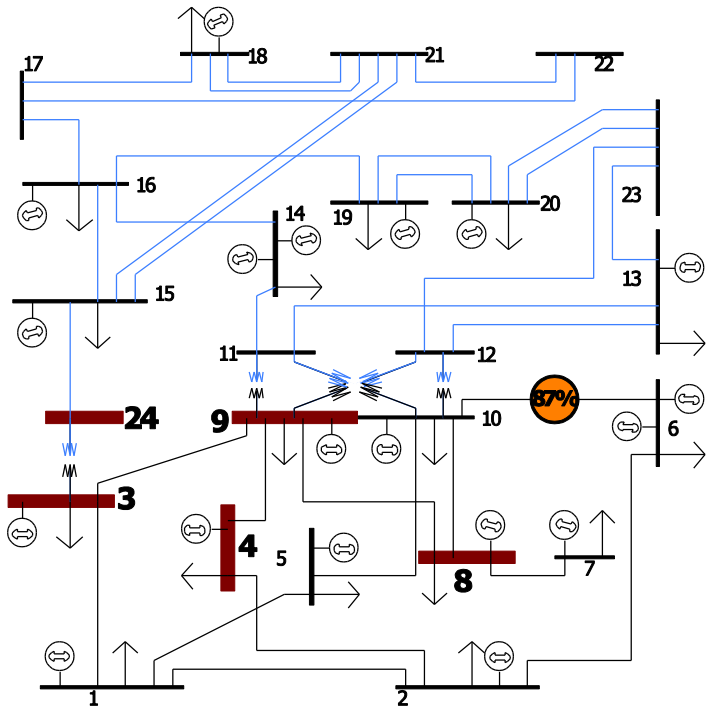

Figure 3. IEEE 24-Bus RTS, reactive power support example

The four buses with the greatest impact on the control objective are set as Q-C buses, and their settings needed to raise the five lowest voltages to a voltage profile of 1 per unit are determined. Columns 2 and 3 of Table 1 show the required MVAr outputs (Q_{net}) to achieve this control.

Table 1. RTS voltage improvement

Bus #	Initial Q_{net}	Final Q_{net}	Initial voltage	Final voltage
3	-37 MVAr	37 MVAr	0.9469	1.0057
4	-15 MVAr	15 MVAr	0.9598	1.0022
8	-35 MVAr	35 MVAr	0.9593	0.9975
9	-36 MVAr	36 MVAr	0.9603	1.0050
24			0.9594	0.9852

Negative Q_{net} indicates a net load while positive Q_{net} indicates a net source. When the new settings are applied to achieve the voltage correction, Table 1 indicates that the power factors at all Q-C buses became leading instead of lagging. Columns 4 and 5 of Table 1 list the voltages at the low-voltage buses. The voltages at all system buses are shown in Figure 4.

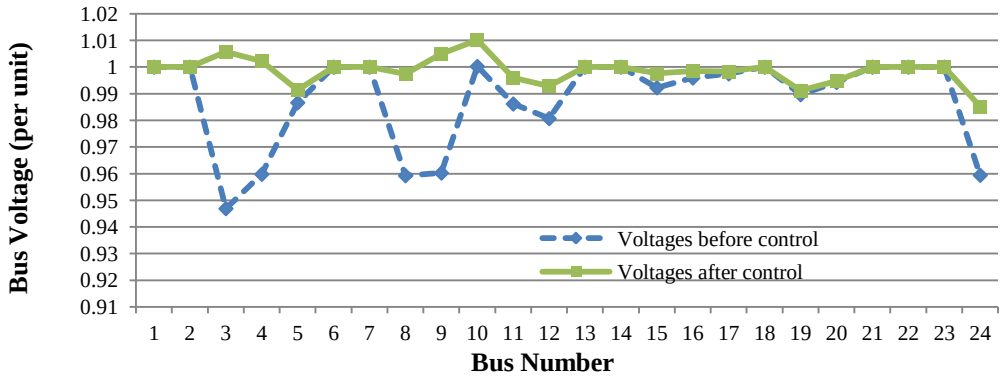

Figure 4. RTS voltage profiles

Reactive power adjustment at a small number of buses can cause a substantial improvement of the overall voltage profile. The use of reactive-only controls as opposed to other forms of corrective control has the advantage that such controllers are already available in the system but are not being utilized, and more may be added. Also, the need to use coarser controls such as shedding load or changing generation output may be prevented.

In the scenario above, we solved a nonlinear problem for the Q_{net} values needed to achieve the desired voltage profile within a small tolerance. Alternatively, we can directly use the sensitivities to approximate the control needed, which requires no iteration. The accuracy of this *linear estimate of control* is dependent on the linearity of the relationship between the reactive power at the Q-C buses and the voltages. In Table 2, these linear estimates of the controls are compared to the actual controls needed. The linear estimates of Q_{net} at buses 3 and 4 are close to the actual values, but buses 8 and 9 are not.

Table 2. RTS linear control estimate

Bus #	Q_{net} Needed	Q_{net} Estimated	% Error	V_F	V_{ESTQ}
3	37 MVAr	37.26 MVAr	0.69 %	1.0057	1.00294
4	15 MVAr	14.52 MVAr	3.21 %	1.0022	0.9982
8	35 MVAr	41.69 MVAr	19.11%	0.9975	0.99813
9	36 MVAr	13.01 MVAr	63.86%	1.0050	0.99826
24				0.9852	0.98396

Using the Q_{net} values from the linear estimate, the voltages at the low-voltage buses V_{ESTQ} are given in column 6 of Table 2. The final voltages V_F from Column 5 of Table 1 are also shown.

With both linear and nonlinear controls, the voltage profile is substantially improved, but the improvement with the linear control is not as great. However, the linear approximation may be calculated a lot faster, so the discrepancies in the final voltage values may be tolerable. Linear versus nonlinear control is investigated further for the probing application in Section 7.4.

To find the reactive support groups for the RTS, the clustering algorithms are applied to the rows of sensitivity matrix $\mathbf{\Lambda_{VQ}}$. For the agglomerative (hierarchical) scheme, the clusters form as shown in Figure 5. From Figure 5, clusters initially form containing (11,12) and (17,20). Then, a large cluster forms and begins to grow. At a higher level in the hierarchy, a cluster forms containing (3,24) which is then absorbed into the large cluster. The last buses to join any cluster are buses 3, 4, 5, 8, 9, and 24.

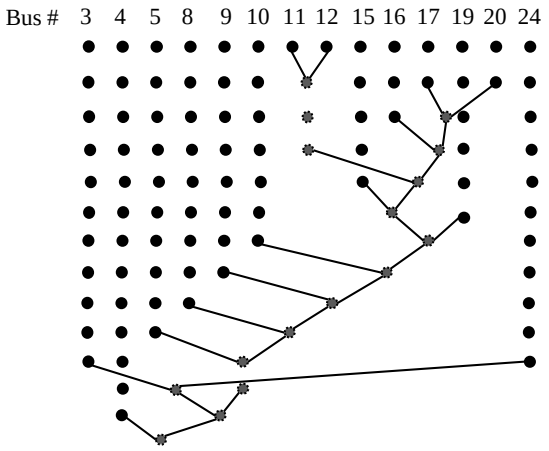

Figure 5. Progression of hierarchical clustering for voltage control example

Results of the QT clustering algorithm, with a threshold of 0.04 and a maximum cluster size of five, the clusters are given by column 1 of Table 3. The QT algorithm shows similar results to the hierarchical algorithm. Buses which were the last to join a cluster in the hierarchical method are shown by the QT algorithm to be their own cluster. The rest of the buses are in two groups. If the maximum cluster size is increased to eight instead of five, the buses in the second cluster become part of the large group, which coincides with the large group we saw form in the hierarchical method.

Results of the CI algorithm, both the reciprocal and simple clusters, are identified in columns 2 and 3 of Table 3. The same bus may belong to multiple reciprocal clusters, so some regions overlap. Buses with voltages not coupled to any other voltages form their own simple cluster.

Table 3. RTS clusters, each shown in brackets

QT	CI - Reciprocal	CI - Simple
[3], [4],[5],[8],[9],[24]	[9,11,12]	[4],[5],[8]
[10, 12, 11, 17, 20]	[15,16,17]	[3,24],[3,15,24]
[15, 16, 19]	[16,17,19]	[9,11,12],[10,11,12]
	[19,20]	[9,10,11,12]
	[3, 24]	[15,16,17,24]
		[15,16,17,19], [19,20]

Based on the reciprocal clusters of the CI algorithm and the additional buses 4, 5, 8, and 10 we identify for each group l supportive buses; let l=5. Each set of buses appears as a group in the first column of Table 4. Thus, each region's five most supporting buses are given by subsequent columns of Table 4, in order of decreasing support effectiveness. Supporting buses may only be CAT1 buses. The supportive buses are listed based on the maximum of the sensitivities of the voltages of the buses in column 1 to reactive power injections. The sensitivity to each supporter bus's reactive power injection is given beneath the bus number.

Table 4. RTS reactive support groups

Voltage-Coupled Groups	5 Supporter Buses and their Corresponding Sensitivities				
4	4	9	3	8	24
	0.071	0.019	0.009	0.005	0.004
5	5	10	8	9	4
	0.050	0.012	0.003	0.002	0.001
10	10	5	8	9	4
	0.024	0.012	0.006	0.005	0.003
9,11,12	9	4	3	8	24
	0.034	0.019	0.015	0.008	0.007
15,16,17	15	24	16	19	3
	0.013	0.011	0.011	0.006	0.006
16,17,19	19	16	15	20	24
	0.017	0.011	0.006	0.006	0.005
19,20	19	20	16	15	24
	0.017	0.009	0.006	0.004	0.003
3,24	3	24	9	4	15
	0.063	0.028	0.015	0.008	0.006

The hybrid approach of Table 4 uses information we know about how voltages can be controlled with respect to other voltages and also ensures that each bus voltage will have at least *l* supporters.

3.3. Local Market Power Potential Identification

In this section, sensitivities and clustering are shown to provide valuable insight into the problem of market power potential identification. The sensitivities of power flows on constrained lines to generator output reveal generators which form load pockets under certain binding transmission system constraints. Details and results are reported in [69] and [70]. Our approach uses this sensitivity information to assess the ability of certain generators to exhibit local market power, a major concern to ISOs.

Market power is defined as the ability of a market participant to profitably maintain prices above a competitive level for a significant period of time [71] and can distort competition in electric markets [72]. In an electricity market, the substitutability of suppliers depends on the structure and capacity of the electric power grid. When the market is competitive, if a generator unilaterally raises its price above marginal cost, it loses customers [73]. However, under transmission congestion, suppliers may possess the ability to manipulate the market to their advantage when their product is no longer substitutable. The ability of non-substitutable generators to raise prices without impacting dispatch is hereafter referred to as "load pocket potential" (LPP). Identification of generators with LPP is achieved using sensitivity information available from the linear programming optimal power flow (LP OPF). The LP OPF is a well-known method for solving optimal redispatch problems [74],[75],[76]. The impact of network constraints on admissible price perturbations in the form of a sensitivity matrix is considered to group generators with the potential to exhibit local market power [69].

The pattern of possible prices is entirely governed by the network characteristics and constraints. Each binding network constraint introduces a degree of freedom in price perturbations, as explained in [69]. The difference between possible or "admissible" price patterns and realized prices is treated in [77]. The theory builds on prior work [73, 78, 79, 77] in which the sensitivity of dispatch and revenues to price is analyzed, accounting for the impact of transmission system

constraints, and admissible LMPs. In this application, it is found that the pattern of admissible LMPs is spanned by vectors associated with constrained line sensitivities, so the basis for admissible price perturbations can be calculated from these sensitivities. Also, the effect of any number of network constraints may be considered simultaneously.

In identifying LPP, we seek instances of price perturbation vectors with a relatively small number of concentrated, non-zero entries. Associated with these entries, a small number of suppliers may jointly share the ability to manipulate prices. In economic experiments such as those done at Cornell University [77], pairs of suppliers with joint market power potential always discover this ability without direct collusion. However, buyers and sellers who are less favored in terms of market power cannot overcome this disadvantage through learning [80].

A basis for these price perturbation vectors is the null space of a price dispatch sensitivity matrix which can be derived from the first order conditions of optimality [77, 81]. The same information can be obtained from transmission constraint sensitivities. The basis is the same as that of the augmented constrained line dispatch sensitivities (generation shift factors), easily determined as a by-product of the LP OPF. Thus, constrained line sensitivities efficiently determine a basis for admissible price perturbation vectors. This application analyzes sensitivities and highlights certain vectors which show concentrated price variations in the fewest entries.

For simplicity, a DC OPF model with only generator MW outputs as controls is considered. The concepts can be sufficiently explained using this model, and extension of this model is possible but is a lengthy topic. In the model,

$$\min_{P_g} \sum_i C_i\left(P_{gi}, w_i\right)$$

such that

$$\mathbf{A}\,\mathbf{diag}(\mathbf{b})\mathbf{A}^{\mathrm{T}}\boldsymbol{\theta} - \begin{bmatrix}\mathbf{P}_g \\ \mathbf{P}_d\end{bmatrix} = 0 \qquad (42)$$

$$\mathbf{diag}(\mathbf{b}_f)\mathbf{A}_f^{\mathrm{T}}\boldsymbol{\theta} - \mathbf{P}_{flow} = 0$$

$C_i(\mathbf{P}_{gi})$ is the generator cost function that depends on the dispatch $\mathbf{P}_{gi}$. The first constraint represents the relation of power injections to bus angle θ, and the second constraint imposes the binding line flow $\mathbf{P}_{flow}$ constraints. Branch susceptances are specified in **b**. Subscript *f* denotes quantities associated with the constrained lines, which are specified in this model. Unconstrained lines are below their limits and are not shown. Matrix **A** is a branch-node incidence matrix that describes the topology of the network. From this model, we show that a basis for possible price perturbation vectors can be determined from sensitivities. For more analysis, refer to [69].

3.3.1. Generator Clustering and Price Perturbations

The LPP algorithm comprises two steps, an initial generator clustering step followed by a step to find admissible price perturbation vectors. The results reveal groups of generators with the potential to exercise market power under the conditions. Each step is briefly explained below. Then, overviews of clustering algorithms implemented for this work which are not already introduced in Section 3.2 are provided in the sections to follow. These include the K-means algorithm and an adaptation of the agglomerative algorithm which is special because of its integration with efficient data structure tools. In particular, tools called *k*-d trees and minimal spanning trees are implemented to enable hierarchical clustering to be performed quickly, even for the large systems studied in this work. Otherwise, the use of hierarchical clustering for large systems is not generally feasible. The QT algorithm and the CI algorithm are also implemented for this work, introduced in Section 3.2.

The first stage of the LPP algorithm is the clustering stage, where sets of candidate generator groupings are formed. We start with the augmented sensitivities **S** from the LP OPF and compute an orthonormal basis **B** using singular value decomposition (SVD; SVD is explained in Chapter 2). The computation order of SVD is $O(nm^2)$. In this application, *m* is the number of constraints, which is small, so this process is not a computational burden. Rows of **B** are then clustered to provide us with the following block matrix structure:

$$\begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} \mathbf{x} = \begin{bmatrix} \Delta\mathbf{y}_1 \\ \Delta\mathbf{y}_2 \\ \vdots \\ \Delta\mathbf{y}_k \end{bmatrix} \tag{43}$$

where k is the number of clusters, $\Delta\mathbf{y}$ is a particular network-admissible price perturbation vector, and $\mathbf{x}$ is a weighting vector (to be determined). In (43), the rows of $\mathbf{B}$ are shown grouped together with other generators in the same cluster. Various clustering algorithms have been implemented and evaluated for LPP identification; extensive details are found in [70].

Clustering reflects the fact that generators with similar impact on the constrained lines may be able to jointly raise their prices without affecting dispatch. To see this, consider the network topology and constraints shown in Figure 6.

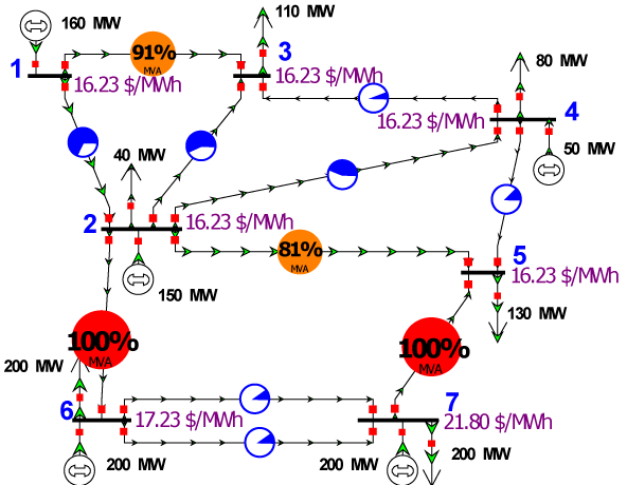

Figure 6. Seven-bus system - initial LMPs and generator dispatches are shown

A load pocket situation exists for bus 6 and bus 7. Incrementally, the load at these buses can only be supplied by the generators at these buses due to the constrained line between buses 2 and 6. Also, both Gen. 6 and Gen. 7 can independently exercise market power. To see why this is true, look at the one-line diagram and suppose Gen. 6 increases its price by 1 $/MWh. The load will accept this price in order to satisfy power balance and in order not to further overload the constrained lines. The same situation is true for Gen. 7. Thus, we expect that Gen. 6 alone, Gen. 7 alone, and Gen. 1, 2, and 4 combined will have some amount of market power potential. The

algorithm results support this intuitive example. Our approach simply extends this notion to identify groups that are not as obvious.

The second stage of the algorithm determines a price perturbation vector $\mathbf{\Delta y}$ in the space of network-admissible vectors to maximize price perturbations for the candidate generators while minimizing price perturbations for the other generators. It is suitable to consider only one cluster at a time and attempt to find a price perturbation vector that will allow those generators to experience LPP. For each cluster *i*, we partition **B** into a matrix $\mathbf{B_i}$ consisting of the rows of **B** which are in *i*, and a matrix $\mathbf{B_{-i}}$ consisting of all other rows. The subscript *–i* is used to denote generators which are not in group *i*. This gives us *k* problems of the following form:

$$\begin{bmatrix} \mathbf{B_i} \\ \mathbf{B_{-i}} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{\Delta y_i} \\ \mathbf{\Delta y_{-i}} \end{bmatrix} \tag{44}$$

where the elements in $\mathbf{\Delta y_i}$ should be much larger than those of $\mathbf{\Delta y_{-i}}$. That is, for each cluster *i*, we want to achieve an **x** so that $\mathbf{\Delta y}$ is of the desired form. The objective function may be rewritten as

$$\begin{aligned} &\text{Max} \quad \mathbf{x^T}\left(\mathbf{B_i}'\right)\mathbf{x} \qquad \text{where} \quad \mathbf{B_i}' = \mathbf{B_i}^{\mathbf{T}}\mathbf{B_i} - \mathbf{B_{-i}}^{\mathbf{T}}\mathbf{B_{-i}} \\ &\text{st} \quad \|\mathbf{x}\| = 1 \end{aligned} \tag{45}$$

The maximum is obtained by choosing **x** to be the eigenvector corresponding to the largest eigenvalue λ_{max} of $\mathbf{B_i}$**'** for each cluster *i*. The proof is given in [69].

3.3.2. K-Means Clustering Algorithm

The *K-means* technique is one clustering method which has shown to be satisfactory for identifying LPP groups. It is also one of the oldest and most widely used clustering algorithms in practice. Details are found in [22]. The algorithm is simple and fast (polylogarithmic) in practice [82], although in theory it has polynomial smoothed running time [83]. K-means is a prototype-based partitional clustering scheme which represents *K* clusters by their centroids. The centroid is typically the mean of the points in the cluster. The K-means algorithm proceeds

as follows: First, *K* points are chosen as the initial cluster centroids. Then, each point is assigned to the closest centroid. Each collection of points is a cluster. Then, centroids for each cluster are recomputed. The process ends when points stop changing clusters, equivalent to when the centroids stop changing.

The K-means algorithm is locally optimal since, by assigning points to their closest centroid, it minimizes the *SSE* in (2) for a given set of centroids. However, K-means may not result in a globally optimal clustering since each iteration considers specific centroids, and the algorithm considers a specific number of clusters. In fact, different runs of K-means for different initializations will generally result in a different *SSE*. To avoid this problem, K-means can be run multiple times, where the results with the lowest *SSE* are used. A lower *SSE* means that the centroids or prototypes of the given clustering are a more accurate representation of the groups in the data.

While results are sensitive to choice of initial centroids, no initialization methods appear to be without problems. For example, consider random initialization, which is a commonly attempted remedy [22]. The first initial centroid may be selected at random; then for each additional initial centroid, the point that is the furthest away from the current centroids may be selected. The problem is that this method can select outliers in sparse regions of the space and is also computationally intensive [22]. Outliers cause clusters to be less well represented by a centroid and thus make the *SSE* larger.

A variant called the *bisecting K-means* algorithm may be used to apply post-processing. In post-processing, a cluster can be split or a new cluster centroid may be added, thereby increasing the number of clusters. To decrease the number of clusters, a cluster can be "dispersed" by removing the centroid and reassigning points in that cluster to the remaining clusters. Clusters can be merged based on the proximity of their centroids to each other or based on whichever merger would result in the smallest increase to the *SSE*. Bisecting K-means successively splits points into two clusters and keeps splitting clusters until *K* clusters are obtained. The bisecting K-means algorithm can be considered a type of hierarchical clustering algorithm if the results at each stage are recorded.

A limitation of K-means is the difficulty in detecting clusters with non-spherical shapes, since clusters are formed based on the distance of points to centroids, and the results can be misleading when the "natural" clusters have different densities [22]. As an example, Figure 7 shows one large, sparse natural cluster and two small, dense natural clusters. The clusters found by K-means, with centroids indicated by crosses, may represent the large cluster as two while representing the two dense clusters as one. These results can be deceptive, especially if the small dense clusters are of more interest than the sparse clusters. In our market power work, for example, this problem may be of concern.

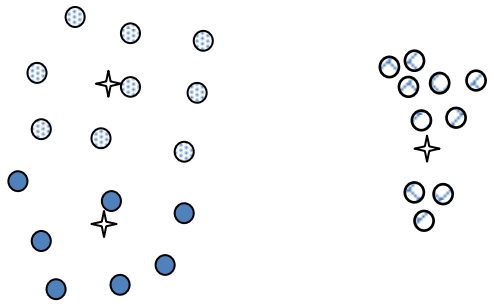

Figure 7. K-means problem with clusters of different density

This is also a suitable place to mention a challenge of high dimensional data. Figure 7 shows data points in 2D, but the data we deal with is in *k* dimensions. It is challenging to understand and visualize clusters of data that exists in *k* dimensions, when only 2D or 3D display is possible.

3.3.3. Minimal Spanning Tree Hierarchical Clustering Algorithm

In the LPP work, it is useful to look at how the groups form and to validate results against other clustering methods. Basic hierarchical clustering has been described in Section 3.2.3. This section focuses on the data structure tools which make agglomerative hierarchical clustering computationally efficient for large systems. A byproduct of these tools is that results can be displayed in a clear, visual way.

Before outlining the data structure tools used, it is important to understand the properties of the algorithm we wish to exploit. In particular, the merging of clusters in agglomerative clustering can be done using *single-linkage*, in which the distance between two clusters is computed as the

minimum distance between its elements. We use agglomerative, single-linkage clustering analysis (SLCA) in this work. A known issue with SLCA is referred to as the *chaining effect*, which means that, due to the way clusters are formed, it is possible for two items in the same cluster to be farther apart than two items in different clusters.

Making hierarchical clustering efficient is a key issue for practical implementation. A poor implementation can have computational complexity of $O(n^3)$. Generally, however, hierarchical clustering has complexity $O(n^2)$. A main factor in the computational cost is the need to compute distances between all points. Thus, nearest-neighbor searching is an important sub-problem that must be carefully handled.

Since the underlying issue which increases computational complexity has to do with nearest-neighbor searches, much effort has been devoted to making these searches as fast as possible and reducing the number of searches needed. We utilize a data structure called a *k-d tree* to partition the items in the *k*-dimensional coordinate space. *K*-d trees were introduced in [84], and [85] describes how to efficiently use them to find the *m* nearest neighbors of a given query point. A *k*-d tree is a binary search tree in multiple dimensions. A simple 2-D kd-tree for points (2,3), (5,4), (9,6), (4,7), (8,1), and (7,2) is shown in Figure 8.

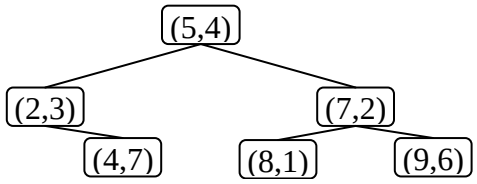

Figure 8. Example 2-D kd-tree

The levels of the tree cycle through each of the *k* axes of the data. The tree is formed recursively by cycling through the axes and splitting the data at the median point for each axis. Since the median value is used to split the data, the tree is balanced by design. The *k*-d tree can be built in $O(n\log n)$ time. Construction of *k*-d trees and different types of queries on the data using *k*-d trees are discussed in [84]. In this context, the type of query we are most interested in is a nearest-neighbor query. That is, given any point, we want to search the tree and return the point with the smallest distance to the query point. Nearest-neighbor searches using *k*-d trees are

discussed in detail in [85], where *k*-d trees are shown to be efficiently used for searching for the *m* nearest neighbors of a query point. There are several variants on *k*-d trees and the associated algorithms, and details of our implementation are found in [70].

It is worth mentioning that *k*-d trees can be used to improve implementations of other clustering algorithms, as well. For example, density of points in space is the basis for clusters in the *DBSCAN* algorithm. Regions of high density are located and separated from regions of low density. Points are either (1) inside of a dense region (a core point), (2) on the edge of a dense region (a border point), or (3) in a sparse region (a noise or background point), defined in [22]. Density is estimated by counting the number of points within a radius, which must be specified, of the current point. The highest possible density includes all points and the lowest possible density is 1. This algorithm time complexity is typically $O(n^2)$, but can be improved using *k*-d trees [22]. In DBSCAN, as in hierarchical methods, there is a need to do expensive nearest-neighbor computations, but *k*-d trees can expedite this process. An advantage of DBSCAN is its ability to identify clusters of unusual shapes, especially when there is a noisy background environment. For applications where outliers are not important, DBSCAN is useful. This type of clustering has proven useful in extending the non-intrusive load monitoring work in [86] on identifying signatures of household loads.

Minimal spanning trees have an important relationship to SLCA. The relationship between minimal spanning trees and SLCA is proven in [87] and is also recognized in [88] and [89]. Consider a set of nodes and a set of edges or branches connecting the nodes. The minimal spanning tree (MST) is a subset of the branches of a graph which has the minimum total distance while providing a route between every pair of nodes [90]. Fast algorithms for constructing minimal spanning trees in coordinate spaces are discussed in [90]. We implement an approximate minimal spanning tree algorithm which uses the concept of *k*-d trees, introduced above, to perform efficient nearest-neighbor searches. The fundamentals of using minimal spanning trees for SLCA are given in [91]. Understanding this connection requires introducing the concepts of a fragment and a nearest-neighbor chain. Simply stated, a *fragment* is a part of the final MST, and a *nearest-neighbor chain* is a particular way to form a fragment, building it towards increasing density. More formal explanations are given in [70] and [91].

For the simple one-dimensional example in Figure 9, we can form a nearest-neighbor chain from Point A consisting of A→B→C→D and a nearest-neighbor chain from Point H consisting of H→G→F→E. Then, the final MST is formed by linking together the existing nearest-neighbor chains, in this case by connecting D and E. From the MST, a dendogram is easily constructed to represent the SLCA hierarchy, as shown in Figure 10. The conversion of the MST to obtain the hierarchy is discussed in [89].

Figure 9. Set of points (left) connected into two fragments (right)

Figure 10. Connecting the MST (left) and dendogram of the hierarchy (right)

The multiple-fragment algorithm, Algorithm 6 in [91], is based on the minimal spanning tree algorithm (MST) and approximate minimal spanning tree algorithm (AMST) of [90]. As the multi-fragment algorithm builds its fragments, the fragments obtained are essentially clusters. The links used to connect these fragments are unnecessary if they are going to be deleted anyway as we aggregate the clusters. In hierarchical clustering, it is practical to stop after clusters have attained some desired size, and not to continue the process until all items are in one cluster. Thus, we may have no need to connect the fragments at all. For details about these structural tools, refer to [70].

Special data structures, such as minimal spanning trees and *k*-d trees, are themselves valuable tools to efficiently search for relationships in the data. Minimal spanning trees and *k*-d trees also provide a way of mapping multi-attribute data into two dimensions. These tools lend themselves to the subfield of data mining called *knowledge discovery in data sets (KDD)*, which has the goal of aiding the user in learning about the data: the user applies KDD techniques which reveal some new insight to the user; then the user applies this new insight to develop the next stage of the

data mining or learning process. Many problems, including finding the minimum spanning tree and triangulation, rely on the proximities of N points. In 2D or 3D, the problem is simplified, and is often encountered when connecting terminals or routing cables using minimal wire length [92],[93]. Closest-point problems in 2D or 3D can be solved in O($n$log$n$) time using a Voronoi diagram [94].

3.3.4. Application to Large Systems

Results of this work for various study scenarios of different sizes, both academic test systems and large utility test systems with real constraints, are presented in [69] and [70]. Briefly summarized in this section are results for two large-case examples in the Eastern Interconnect. Details are available in [70]. A total of 43,311 buses are present in the Eastern Interconnect.

For LPP identification, a cluster quality metric is introduced in [70] for the purpose of comparing results from different clustering methods. For a given scenario, LPP is said to exist if a clustering of generators is found, where for at least one cluster i, the objective function evaluated at the optimal value of $\mathbf{x}$ produces *well-separated* price perturbation vectors. To measure this quality of the clusters, we define f_{LPP},

$$f_{LPP} = \frac{\mathbf{\Delta y_i}^{\mathbf{T}} \mathbf{\Delta y_i}}{Ni} - \frac{\mathbf{\Delta y_{-i}}^{\mathbf{T}} \mathbf{\Delta y_{-i}}}{N_{-i}} \tag{46}$$

where N_i denotes the number of generators in LPP group i and N_{-i} denotes all other generators. The values of f_{LPP} are between -1 and 1. When LPP exists, the values are near 1. Conversely, f_{LPP} near 0 indicates that the price perturbation vector is not well separated, where generators outside of the LPP group may have substantial non-zero price perturbations. For each group in a clustering, a price perturbation vector $\mathbf{\Delta y}$ is found, and (46) is evaluated. The f_{LPP} provides information about the extent to which an algorithm has succeeded in identifying load pockets, which is important for our application. The need for developing metrics, exhibited here, is also seen throughout this thesis.

All price perturbation vectors found from the algorithm, by definition, result in no change in dispatch [69]. Generators outside of the LPP cluster are assumed not to change their prices, which is why we search for vectors with entries in $\mathbf{\Delta y_{-i}}$ which are zero. However, they are not always exactly zero. Experimentally, we have found that even when these entries are not exactly zero, we can often approximate them as zero by implementing the price perturbations only for the generators in the LPP cluster, and the result still produces no change in dispatch. Thus, rather than requiring f_{LPP} exactly equal to 1 for LPP to exist, values slightly less than 1 may also be acceptable. For the test cases, results are shown where $f_{LPP} > 0.9$. Presumably, there is some threshold where, if entries in $\mathbf{\Delta y_{-i}}$ become too large, they are no longer negligible in this manner. However, more research is needed to identify this threshold by examining the range over which price perturbation vectors of the desired form exist.

The first area of interest, which we refer to as Area A, has 6,453 buses and 638 generators, and six binding constraints are considered. Results are shown in Table 5.

Table 5. Area A clustering results

Algorithm	No. Total Clusters	No. of Clusters with f_{LPP}>0.9	f_{LPP} of Clusters > 0.9	Gens in Clusters
QT, 0.25	6	5	0.999	37550
			0.989	5067
			0.927	9380, 9124, 9123, 9078
			0.998	37658, 37657, 37656, 37655
			1	32,32
QT, 0.1	10	4	0.999	37550
			0.989	5067
			0.998	37658, 37657, 37656, 37655
			1	32,32
K-Means, 6	6	4	0.999	37550
			0.998	37658, 37657, 37656, 37655
			0.998	32
			0.998	32
K-Means, 7	7	4	0.999	37550
			0.998	37658, 37657, 37656, 37655
			0.998	32
			0.998	32
CI	6	1	1	32,32
Agglo., 7	7	3	0.989	5067
			0.999	37550
			0.998	37658, 37657, 37656, 37655

A second large example case is referred to as Area B. A list of binding constraints is available on the MISO website [95]. For each day, a file records the binding constraints in the system every five minutes. A constraint which was biding 119 times for an arbitrary day is given in Table 6.

Table 6. Example MISO binding constraints

Constraint Name	Branch Name	Contingency Description	No. Times Binding
'AMI13148_OTTO WATP_IP-1516_2'	'OTTOWATP IP-1516 2 (LN/AMIL/AMIL)'	'HENNEPIN - HALLOCK 138 (1512)'	119

Figure 11 shows several buses in the area of interest. The contingency is represented by opening the line between "4RICHLND JCT" and "4TLV STL ETP." The thick, dark line is assumed to be binding. Results in Table 7 indicate that two well-separated clusters are found. The small cluster contains two generators at the same bus.

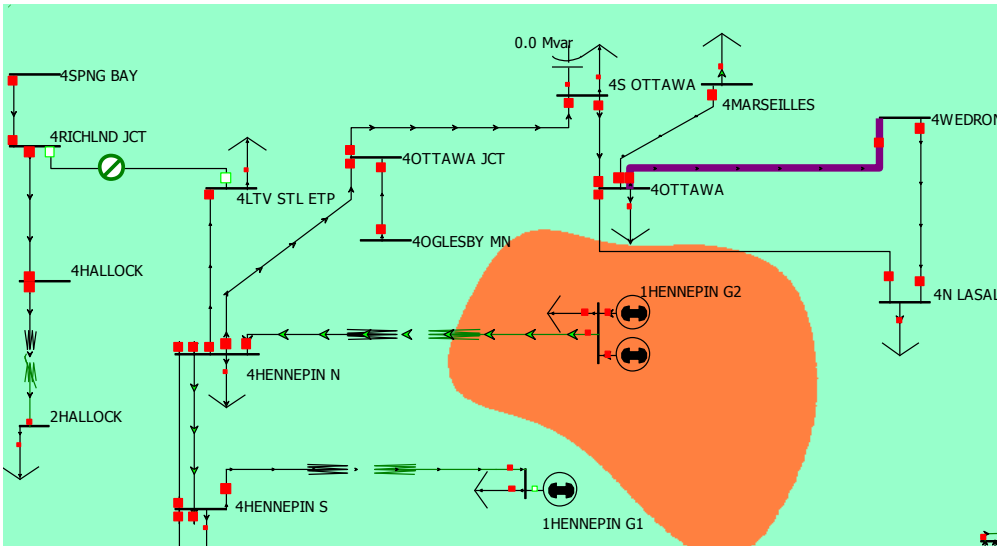

Figure 11. An LPP group for the Area B scenario

Table 7. Area B clustering results

Algorithm	No. Total Clusters	No. of Clusters with f_{LPP}>0.9	f_{LPP} of Clusters > 0.9	Gens in f_{LPP}>0.9 Clusters
QT, 0.25	2	1	0.999	349107, 349107
QT, 0.1	4	1	0.999	349107, 349107
K-Means, 2	2	1	0.999	349107, 349107
K-Means, 3	3	1	0.999	349107, 349107
CI	2	0	-	-
Agglo., 7	2	1	0.999	349107

To further evaluate our clustering results, the f_{LPP} of randomly generated clusters is examined. A random number of clusters are created, and generators are assigned at random to these clusters. Results of five random clusterings on three different systems are shown in Table 8. As the larger study systems are examined, randomly selected clusters are less likely to select "true" clusters. The random-clustering results are poor for both large systems, as expected; out of five clusterings, there are no clusters with f_{LPP}>0.9.

Table 8. f_{LPP} for random cluster assignments

Case	No. of Clusters	No. Clusters with f_{LPP}>0.9	Max f_{LPP} of any Cluster
IEEE 118	6	0	0.653
	10	1	0.956
	7	0	0.653
	12	1	0.986
	3	0	0.312
Area A	223	0	0.842
	195	0	0.482
	170	0	0.493
	157	0	0.412
	3	0	0.183
Area B	27	0	0.125
	31	0	0.059
	118	0	0.497
	49	0	0.064
	126	0	0.247

3.4. Conclusions

It is clear from the work in this chapter that considerable insight can gained from sensitivities. Clustering sensitivities allows for the selection and aggregation of network elements with the ability to "help each other" solve problems. Several clustering algorithms are found useful, including an algorithm we created based on the coupling index. It is also shown that appropriate data structure tools, *k*-d trees and minimal spanning trees in particular, are indispensible for achieving the computational efficiency needed for studying large systems, particularly for hierarchical clustering. Applications throughout this thesis employ sensitivities to characterize relationships; that insight has been revealed in this chapter.

Sensitivities and clustering are shown to facilitate the coordination of multiple devices to achieve a control objective in Section 3.2. A specific case of system voltage control via distributed

reactive power sources is presented. Centralized control is assumed; there remains opportunity for work on the decentralized problem. For example, one may be able to design a control scheme in terms of *constrained information patterns*, which considers the amount of information available to each controller [96], [40].

A sensitivity-based grouping approach for LPP identification is presented in Section 3.3. Results reveal that the identified LPP groups are strikingly consistent, regardless of the choice of clustering method and parameters. There are a few differences, which are examined in [70]. This method-independence of the results is a key strength. Also, although it has not been emphasized, the entire MST and all clusterings in the agglomerative hierarchy are available. This additional information shows exactly how the clusters merge with each other as the size of clusters increases, which can make the discovered relationships easier to understand. There is a need, evident from this chapter, for more quantitative metrics to define acceptable grouping results.

The key conclusion is that sensitivities provide a repository of patterns useful for power systems analysis which may be exploited by myriad machine learning, data mining, and pattern recognition tools. The concepts, data structures, and algorithmic tools developed in Chapter 3 are general, and the framework is in place for their wide-ranging application. We utilize the observations, results, and tools from Chapter 3 in subsequent chapters, Chapter 4 in particular. The work in this chapter also underscores the importance of secure communications and control, considered in Chapter 7.

4. Estimation of Network Models

The validity of power network models is important; the model needs to be consistent with the data which is obtained. The critical role of network modeling is obvious; results of power systems analysis are only as valid as the models they are based upon. The goal of the applications presented in this chapter is to gain insight into the characteristics of the network, in the form of a model, through the use of data mining techniques. Section 4.1 examines how model validity in power systems can be assessed and improved with the aid of measurement data. The first application, in Section 4.2, presents the estimation of a static equivalent model from data. The second application, in Section 4.3, examines the development of structure-preserving reduced network models. The third application, in Section 4.4, is similar to the first application except that rather than estimate an equivalent model, the parameters on specific lines are estimated using historical data.

4.1. Data Capabilities for Network Modeling

This section describes how utilizing data can enable improved network modeling. The availability of PMU data motivates the first application. High-precision time-stamped data are obtained from PMUs at typical rates of 30 samples per second (1 sample per 2 cycles for a 60 Hz system). A single PMU's analog measurement quantities may include multiple three-phase voltages and currents and their positive sequence quantities, frequency, and the rate of change of frequency [7]. There are many opportunities for this data to improve the reliable operation of power systems, and the North American Synchrophasor Initiative (NASPI) exists to help advance these efforts [97]. PMU data can be shared over a communications network in real time and collected at a centralized device, called a phasor data concentrator (PDC) [98]. NASPInet will enable PMU data sharing so that utilities can use the data for situational awareness, event analysis for operations and planning, model validation, real-time control and protection, system restoration, and more [99]. There are many Recovery Act synchrophasor projects in various stages of development [97], and NASPInet is likely to be an important factor in achieving long-term satisfaction from the projects.

The model identification problem considered in this application is motivated in part by the fact that a spreadsheet of data is often obtained with no accompanying information for mapping the

data points to the real-world system. PMU data is being made available to researchers in limited forms through non-disclosure agreements (NDAs) [97], but often underlying models of the system are unavailable or incomplete. In fact, the underlying system model may be unknown, even to the owners of the PMUs. Still, there is a need to know what events are actually occurring even when one does not have detailed knowledge about the system. Thus, a natural question is whether it is possible to develop such models from the data itself. An equivalent model of the network estimated such that it is matched to the PMU buses may be helpful in analyzing the system. In [100], we explore the use of PMU data to gain information about the system for this purpose.

Fundamental to these applications is exploiting the benefit of data-over-time. A small example can be used to show the concept of using data from multiple time points to obtain more information about the system. To examine how measurements over time can improve an estimate, consider the three-bus example in Figure 12. Let P_{12}, P_{32}, θ_1, and θ_3 denote measurements at one instance in time. The unknown variables are denoted by a hat.

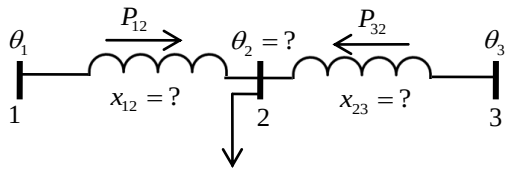

Figure 12. Two-impedance example circuit

$$P_{12} = \left(\theta_1 - \hat{\theta}_2\right) / \hat{x}_{12} \tag{47}$$

$$P_{32} = \left(\theta_3 - \hat{\theta}_2\right) / \hat{x}_{23} \tag{48}$$

At first glance, it appears that there is not enough information to solve for the unknown impedances, since θ_2 is also unknown. However, if we observe the operating point of the system at multiple snapshots, it becomes apparent that we actually do have enough information. If new measurements P_{12}, P_{32}, θ_1, and θ_3 are taken at time instance *b*, the value of the unmeasured θ_2, will also be different (presumably). This adds two equations, but only one new unknown.

$$P_{12}{}^{b} = \left(\theta_1^{b} - \hat{\theta}_2^{b}\right) / \hat{x}_{12} \tag{49}$$

$$P_{32}{}^{b} = \left(\theta_3^{b} - \hat{\theta}_2^{b}\right) / \hat{x}_{23} \tag{50}$$

With four equations and four unknowns, we can accurately solve for the values of x_{12} and x_{23}. As long as enough sufficiently different operating points are obtained, it is possible to improve estimates by taking samples over time. These are fundamental concepts which facilitate the use of data to enhance power system models.

4.2. Computation of Equivalent at PMU Buses Based on Data

Data mining in the form of system identification techniques can be used for constructing PMU-based equivalents. This application computes an equivalent as seen from the PMU buses. Details and results are reported in [100]. The ultimate question is to determine how far down we can see into the system and what we are able to observe. Initially, a worst-case, "model-free," point of view is adopted, where it is assumed that there is no knowledge of the underlying system.

From the discussion of sensitivities in Chapter 3 and the analysis in [100], it is apparent that information about the power flow Jacobian **J** and its inverse can be obtained from PMU data. The Jacobian reveals information about the structure and the behavior of the underlying network. Consider that analytical sensitivities in the Jacobian can be verified empirically by changing the load at each bus by a small amount and then solving a power flow to observe the resulting angle changes. The empirical computation of sensitivities is fundamentally the same concept which allows **B** to be estimated from PMU data. For simplicity, we denote the network model **B**, and the application is described in terms of the basic DC power flow assumptions [101],[102]. The estimate recovered is actually the negative of the real power vs. angle portion of **J**. Details are explained in [100]. It is clear that computation of **B** requires *n* linearly independent sets of injections and their corresponding angle changes. Ideally, a small load perturbation should originate from each bus, one at a time. In reality, the changes could be anything.

4.2.1. PMU-Based Equivalent Estimation

In [103], a method is presented for estimating an external equivalent based on the measurements at the boundary buses. An adaptation of this method applies to PMU buses. Generally, each PMU bus is connected to at least one non-PMU bus. With PMUs, angles can be measured directly, whereas in [103] they must come from a state estimator. At PMU buses, denoted by subscript *P*, real powers and angles are measured. Subscript *E* denotes all other (external) buses, where we assume no data is available. Thus, the system can be partitioned as

$$\begin{bmatrix} \Delta \mathbf{P_P} \\ \Delta \mathbf{P_E} \end{bmatrix} = \begin{bmatrix} \mathbf{B_{PP}} & \mathbf{B_{PE}} \\ \mathbf{B_{EP}} & \mathbf{B_{EE}} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{\theta}_\mathbf{P} \\ \Delta \boldsymbol{\theta}_\mathbf{E} \end{bmatrix} \tag{51}$$

where the second equation of (51) can be written as

$$\Delta \boldsymbol{\theta}_\mathbf{E} = \mathbf{B_{EE}}^{-1} \left(\Delta \mathbf{P_E} - \mathbf{B_{EP}} \Delta \boldsymbol{\theta}_\mathbf{P} \right) \tag{52}$$

Substituting (52) into the first equation of (51) reduces to

$$\Delta \mathbf{P_P} = \mathbf{B_{PP}} \Delta \boldsymbol{\theta}_\mathbf{P} + \mathbf{B_{PE}} \left[\mathbf{B_{EE}}^{-1} \left(\Delta \mathbf{P_E} - \mathbf{B_{EP}} \Delta \boldsymbol{\theta}_\mathbf{P} \right) \right] \tag{53}$$

$$\Delta \mathbf{P_P} = \left(\mathbf{B_{PP}} - \mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \mathbf{B_{EP}} \right) \Delta \boldsymbol{\theta}_\mathbf{P} + \left(\mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \right) \Delta \mathbf{P_E} \tag{54}$$

If $\Delta \mathbf{P_E}$ in (54) is zero, the relationship between $\Delta \mathbf{P_P}$ and $\Delta \boldsymbol{\theta}_\mathbf{P}$ is completely characterized by $\mathbf{B_{PP}}$-$\mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \mathbf{B_{EP}}$. We can define an equivalent injection associated with the angle changes at the PMU buses and rewrite (54):

$$\Delta \mathbf{P_{eq}} = \Delta \mathbf{P_P} - \left(\mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \right) \Delta \mathbf{P_E} \tag{55}$$

$$\Delta \mathbf{P_{eq}} = \left(\mathbf{B_{PP}} - \mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \mathbf{B_{EP}} \right) \Delta \boldsymbol{\theta}_\mathbf{P} \tag{56}$$

Thus, the correct estimation of $\mathbf{B_{PP}}$-$\mathbf{B_{PE}} \mathbf{B_{EE}}^{-1} \mathbf{B_{EP}}$ from PMU measurements requires an adjustment of the power injection measurement values at the PMU buses by first subtracting

$\mathbf{B_{PE}B_{EE}^{-1}}\Delta\mathbf{P_E}$ from the true injections. That is, we must subtract an amount from each PMU bus injection equal to

$$\Delta\mathbf{P_{adjust}} = \left(\mathbf{B_{PE}B_{EE}^{-1}}\right)\Delta\mathbf{P_E} \tag{57}$$

However, $\mathbf{B_{PE}B_{EE}^{-1}}\Delta\mathbf{P_E}$ is unknown, and even the dimension of $\Delta\mathbf{P_E}$ may be unknown. Such an "adjusted estimate" results in the true equivalent at the PMU buses. If $\Delta\mathbf{P_E}$ is zero or if the entire expression in (57) is zero, no adjustment is needed, and the estimate obtained will be the true equivalent. We are interested in knowing under what conditions this is zero and the quality of an estimate we can obtain by assuming it to be zero, due to lack of better information. Then, perhaps we can characterize what the adjustment value needs to be in order to obtain a more accurate estimate. All external buses may be reduced to one bus, and then we only need to consider scalar values of $\mathbf{P_E}$ and $\mathbf{B_{EE}}$. The vector $\mathbf{B_{PE}}$ may be known in reality if the utility knows what is connected to each of the PMU buses. For the following discussion, let us assume that the adjustment term is zero (or that it is known, effectively making it zero).

The estimate of $\mathbf{B_{PP}}$-$\mathbf{B_{PE}B_{EE}^{-1}B_{EP}}$ is computed using vectors $\Delta\mathbf{P_P}$ and $\Delta\boldsymbol{\theta}_\mathbf{P}$ at r time points, which can be combined into the matrices $\mathbf{P_r}$ and $\boldsymbol{\theta}_\mathbf{r}$, respectively:

$$\begin{aligned}\mathbf{P_r} &= \left[\Delta\mathbf{P_P}(1),\cdots,\Delta\mathbf{P_P}(r)\right]\\ \boldsymbol{\theta}_\mathbf{r} &= \left[\Delta\boldsymbol{\theta}_\mathbf{P}(1),\cdots,\Delta\boldsymbol{\theta}_\mathbf{P}(r)\right]\end{aligned} \tag{58}$$

The solution obtained is equivalent to minimizing the objective function J,

$$J = \sum_{i=1}^{r}\left[\Delta\mathbf{P_P}(i) - \mathbf{B_{est}}\Delta\boldsymbol{\theta}_\mathbf{P}(i)\right]\left[\Delta\mathbf{P_P}(i) - \mathbf{B_{est}}\Delta\boldsymbol{\theta}_\mathbf{P}(i)\right] \tag{59}$$

so that an estimate for $\mathbf{B}$ is obtained from the least-squares estimate:

$$\mathbf{B_{est}} = \mathbf{P_r}\boldsymbol{\theta}_\mathbf{r}^{\mathbf{T}}\left(\boldsymbol{\theta}_\mathbf{r}\boldsymbol{\theta}_\mathbf{r}^{\mathbf{T}}\right)^{-1} \tag{60}$$

When this estimate is obtained, the residual term

$$\mathbf{\Delta P_P - B_{est} \Delta \theta_P} \tag{61}$$

is as small as possible for all given $\mathbf{\Delta P_P}$ and $\mathbf{\Delta \theta_P}$.

An advantage of this approach is that it can readily handle situations where two buses are not connected to each other. Conversely, with other estimation approaches, numerical problems are encountered when attempting to estimate the value of **B** (or line impedance **z**) for lines which are actually out of service.

4.2.2. Incorporating Additional Measurements

The above estimate is computed assuming that for each PMU bus, both power injection and angle measurements are known. The formulation can be extended to include buses where only partial data, either the real power injection or the angle, is known. The approach for incorporating additional measurements into the estimate is described in [100].

To model additional measurements, we first divide the system into PMU buses and the external system. Then, the system is further divided to distinguish buses which may have either known power injection measurements $\mathbf{\Delta P_K}$ or known angle measurements $\mathbf{\Delta \theta_K}$. The unknown external angles are $\mathbf{\Delta \theta_U}$ and the unknown external injections are $\mathbf{\Delta P_U}$.

$$\begin{bmatrix} \mathbf{\Delta P_P} \\ \mathbf{\Delta P_K} \\ \mathbf{\Delta P_U} \end{bmatrix} = \begin{bmatrix} \mathbf{B_{PP}} & \mathbf{B_{PU}} & \mathbf{B_{PK}} \\ \mathbf{B_{KP}} & \mathbf{B_{KU}} & \mathbf{B_{KK}} \\ \mathbf{B_{UP}} & \mathbf{B_{UU}} & \mathbf{B_{UK}} \end{bmatrix} \begin{bmatrix} \mathbf{\Delta \theta_P} \\ \mathbf{\Delta \theta_U} \\ \mathbf{\Delta \theta_K} \end{bmatrix} \tag{62}$$

Solving the second equation in (62) for $\mathbf{\Delta \theta_U}$,

$$\mathbf{\Delta \theta_U = B_{KU}^{-1} \left(\Delta P_K - B_{KP} \Delta \theta_P - B_{KK} \Delta \theta_K \right)} \tag{63}$$

and substituting (63) into the first equation in (62), we obtain the following after simplification:

$$\begin{aligned}\Delta \mathbf{P_P} &= \left(\mathbf{B_{PP}} - \mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\mathbf{B_{KP}}\right)\Delta\boldsymbol{\theta}_\mathbf{P} \\ &+ \left(\mathbf{B_{PK}} - \mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\mathbf{B_{KK}}\right)\Delta\boldsymbol{\theta}_\mathbf{K} \\ &+ \left(\mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\right)\Delta\mathbf{P_K}\end{aligned} \tag{64}$$

which, in matrix form, can be written as

$$\Delta \mathbf{P_P} = \left[\mathbf{A'} \quad \mathbf{A''} \quad \mathbf{A'''}\right]\left[\Delta\boldsymbol{\theta}_\mathbf{P}{}^{\mathbf{T}}, \quad \Delta\boldsymbol{\theta}_\mathbf{K}{}^{\mathbf{T}}, \quad \Delta\mathbf{P_K}{}^{\mathbf{T}}\right]^{\mathbf{T}} \tag{65}$$

where the matrix elements are given by

$$\begin{aligned}\mathbf{A'} &= \mathbf{B_{PP}} - \mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\mathbf{B_{KP}} \\ \mathbf{A''} &= \mathbf{B_{PK}} - \mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\mathbf{B_{KK}} \\ \mathbf{A'''} &= \mathbf{B_{PU}}\mathbf{B_{KU}}^{-1}\end{aligned} \tag{66}$$

and may be estimated as in (60). Instead of the matrix $\boldsymbol{\theta}_\mathbf{R}$ consisting of $\Delta\boldsymbol{\theta}_\mathbf{P}$, it now consists of concatenated vector $[\Delta\boldsymbol{\theta}_\mathbf{P}{}^{T}, \Delta\boldsymbol{\theta}_\mathbf{K}{}^{T}, \Delta\mathbf{P_K}{}^{T}]^T$ at r time points. Then, the estimate gives us an equivalent network connecting PMU buses with PMU buses, PMU buses with known angle buses, and PMU buses with known power injection buses. It should be noted that the formulation above does not connect the buses where only theta is known or the buses where only the power injection is known with each other.

4.2.3. Mapping System Events

One desired application of the estimate is to relate measurements seen at the PMU buses to events in the true system. This same goal is also motivation for the oscillation monitoring based event detection work in Chapter 6. At the most basic level, if large changes in the system occur, the estimate will show a change. The question is whether or not the cause of the change can be determined. Similarly to monitoring boundary line flows in [104], a large change can serve as an indicator that a major event may have occurred in the system:

$$\mathbf{\Delta P} = \mathbf{B}_{\mathbf{EQ,normal}} \cdot \mathbf{\Delta\theta} \tag{67}$$

$$\frac{\|\mathbf{\Delta P}\|}{\|\mathbf{\Delta\theta}\|} \geq \text{tol} \tag{68}$$

However, if more detail is to be obtained about the event, other than an indicator that "something" has occurred, it is necessary to know more about the model. Thus far, the application has taken a completely model-free approach. Now we relax that assumption. If the *base-case* or "normal" system is known, a topology change in the full system can be mapped to a change in the equivalent system or to a change in equivalent line flow. Then, it is possible to determine the impact of a line outage on the line flows in the equivalent system, as described in [100].

Let the base-case system be $\mathbf{B}_{\mathbf{base}}$. A line outage in the system has an impact on the matrix of the following form:

$$\mathbf{B}_{\mathbf{base}} = \mathbf{B}_{\alpha} + \mathbf{E}_{\alpha} \tag{69}$$

where $\mathbf{E}_{\alpha}$ is the topology error matrix for the outage of line α.

$$\mathbf{E}_{\alpha} = \begin{bmatrix} & \vdots & & \vdots & \\ \cdots & b & \cdots & -b & \cdots \\ & \vdots & & \vdots & \\ \cdots & -b & \cdots & b & \cdots \\ & \vdots & & \vdots & \end{bmatrix} \tag{70}$$

and b is the inverse of the line impedance on line α. To find the expected equivalent $\mathbf{B}_{\mathbf{EQ},\alpha}$ at the PMU buses after outage of line α, $\mathbf{E}_{\alpha}$ is partitioned and subtracted from the partitioned matrices of the **B** matrix:

$$\mathbf{B}_{\alpha,\mathbf{PP}} = \left(\mathbf{B}_{\mathbf{PP,Base}} - \mathbf{E}_{\mathbf{PP},\alpha}\right), \quad \mathbf{B}_{\alpha,\mathbf{EE}} = \left(\mathbf{B}_{\mathbf{EE,Base}} - \mathbf{E}_{\mathbf{PP},\alpha}\right)$$
$$\mathbf{B}_{\alpha,\mathbf{PE}} = \left(\mathbf{B}_{\mathbf{PE,Base}} - \mathbf{E}_{\mathbf{PE},\alpha}\right), \quad \mathbf{B}_{\alpha,\mathbf{EP}} = \left(\mathbf{B}_{\mathbf{EP,Base}} - \mathbf{E}_{\mathbf{EP},\alpha}\right) \tag{71}$$

$$\mathbf{B}_{\mathbf{EQ},\alpha} = \mathbf{B}_{\alpha,\mathbf{PP}} - \mathbf{B}_{\alpha,\mathbf{PE}}\mathbf{B}_{\alpha,\mathbf{EE}}^{-1}\mathbf{B}_{\alpha,\mathbf{EP}} \tag{72}$$

The expected line flows in the equivalent ($\mathbf{B}_{\mathbf{EQ},\alpha}$) are then easily calculated from the equivalent B matrix and the true angles at the PMU buses. The angles found from the DC power flow solution, using the estimated equivalent,

$$\boldsymbol{\theta}_{\mathbf{P}} = \mathbf{B}_{\mathbf{EQ}}^{-1} \cdot \mathbf{P}_{\mathbf{EQ}} \tag{73}$$

should always exactly equal the measured angles at the PMU buses. That is, the equivalencing procedure necessarily preserves the system angles as a result of the way the equivalent is made. This measured information about the equivalent can then be further analyzed.

4.2.4. Example Equivalent Estimation Cases

In this section, results of the estimation approach for the seven-bus test system are presented. The system is shown in Figure 6 of Chapter 3. For more details, see [100]. The full **B** matrix for this system is given in Table 9:

Table 9. Actual full system matrix

B =	-18.75	15	3.75	0	0	0	0
	15	-47.5	5	5	7.5	15	0
	3.75	5	-38.75	30	0	0	0
	0	5	30	-38.75	3.75	0	0
	0	7.5	0	3.75	-26.25	0	15
	0	15	0	0	0	-22.5	7.5
	0	0	0	0	15	7.5	-22.5

In the most ideal situation, every bus would provide information which could be used to estimate **B**. For such a case, we simulate data by perturbing the load at each bus, except for the slack bus, bus 7. Although bus 1 does not ordinarily have a load, we assign it a load of 0 MW. As expected, using the estimation method with simulated angle and power measurement data, the original **B** is recovered as the estimate $\mathbf{B}_{\mathbf{est}}$ in Table 10.

Table 10. Full system estimated matrix

$\mathbf{B}_{est}$ =

-18.75	15	3.75	0	0	0
15	-47.50	5.00	5.00	7.50	15.00
3.75	5	-38.75	30	0	0
0	5.00	30.00	-38.75	3.75	0
0	7.50	0	3.75	-26.25	0
0	15	0	0	0	-22.50

The exception to the perfect recovery in this example is that no information can be obtained about the slack bus. In a real system, these perturbation sets would be obtained from the load profile of the system over the course of time, where the perturbations must be diverse in the sense that they should include as many buses as possible.

An example of the estimation approach is presented assuming that the PMUs are located at buses 1, 5, and 6. Of interest is whether outages of lines not connected to PMU buses manifest themselves in a noticeable way. The Ward equivalent calculated at the PMU buses is denoted the *actual equivalent* in Table 11 and has the line flows given in Table 12.

Table 11. Actual Ward equivalent at PMU buses 1,5, and 6

$\mathbf{B}_{EQ}$ =

-10.811	4.6573	6.1538
4.6573	-23.119	3.4615
6.1538	3.4615	-17.115

Table 12. Line flows in actual equivalent of Table 11

EQ flows, MW =

67.86
41.33
-66.45

For this case, the estimate $\mathbf{B}_{est}$ in Table 13 incorporates injection changes which occur over time at all buses in the full system. The injection sets for this example include equally significant perturbations at all the buses, including the external buses for which no data is available to compute the estimate.

Table 13. Estimate with external injections

$\mathbf{B}_{est}$ =

-15.158	16.388	11.512
1.9809	-14.095	6.7097
2.9222	10.264	-12.882

As expected, the estimate shows that the presence of unmeasured load perturbations in the external system can have a significant effect on its accuracy. However, if the external injection term is truly zero, or equivalently if it is possible to compute the adjusted injections in (55), the estimate is $\mathbf{B}_{\mathbf{est,0}}$ in Table 14.

Table 14. Estimate without external injections

$\mathbf{B}_{\mathbf{est,0}}$ =

-10.811	4.6574	6.1538
4.6573	-23.119	3.4615
6.1536	3.4615	-17.115

Some observations can be summarized about these estimates:

1. The equivalent estimated when no injection changes occur at non-PMU buses accurately captures **B**.
2. The adjusted equivalent also accurately captures **B**.
3. The equivalent network is generally a full network with lines connecting all possible pairs of buses.
4. The estimate when injection changes are equally present at both the PMU buses and the non-PMU buses does not accurately capture **B**, and it is not symmetric.

For the same example case, we now examine the impact of line outages. Single-line outages of line (2,4) and line (3,4) result in actual equivalents given by the matrices in Table 15.

Table 15. Actual equivalents with line outages

$\mathbf{B}_{\mathbf{EQ,(2,4)}}$ =

-10.81	4.6695	6.1407
4.6695	-22.964	3.2942
6.1407	3.2942	-16.935

$\mathbf{B}_{\mathbf{EQ(3,4)}}$ =

-10.11	3.956	6.1538
3.956	-12.418	8.4615
6.1538	8.4615	-14.615

There is noticeable impact of both line outages on the equivalent, and the corresponding equivalent line flows are given in Table 16.

Table 16. Actual equivalent line flows with line outages

EQ line flows, MW, (2,4) =	69.53
	39.21
	-70.54

EQ line flows, MW, (3,4) =	50.71
	35.10
	-60.21

As before, if the external system injection changes are zero, the estimate accurately recovers the true equivalent. Thus, one can discern which line outage has occurred. However, if there are considerable injection changes at PMU and non-PMU buses, the estimate will not match the actual equivalent. Examples are shown in [100]. In this case, estimates may deviate significantly from the estimate that would be expected given zero external injections. Although this phenomenon presents a significant challenge, assuming these estimated equivalents are distinguishable from each other and from the no-outage estimate, it may be possible to do further data mining work to identify and classify these event "signatures."

In summary, even if we obtain the best estimate possible, to determine which line in the original system experienced an outage requires knowledge of the base-case system. Then, one can watch the expected equivalent line flow changes as an indicator, and that area can be examined in more detail following a suspected event.

The changes in equivalent system line flows due to a line outage of each of the ten original lines in the seven-bus system are shown in Table 17. Column values are equal to the new equivalent line flow minus the original (non-outaged) equivalent line flow.

Table 17. MW changes in equivalent line flows due to line outages in seven-bus system

	(1,2)	(1,3)	(2,3)	(2,4)	(2,5)	(2,6)	(3,4)	(4,5)	(5,7)	(6,7)
EQ(1,5)	-16.81	1.09	0. 61	1.67	-21.39	58.75	-17.14	-14.35	14.92	18.08
EQ(1,6)	-8.22	13.20	-4.18	-2.12	28.19	-39.23	-6.24	6.81	-14.92	-18.08
EQ(5,6)	-7.01	-7.60	-4.17	-4.09	-34.14	66.45	6.24	-3.14	19.78	15.65

Comparisons and results of the estimation under various line outages are given in [100]. For this small system, each line has a distinct impact on the equivalent, so it is possible to discern which

line is out. In general, the mapping of line outages in the full system to changes in equivalent system line flows may not be unique.

4.3. Essential Structure Preserving Network Reductions

When examining large interconnected power systems, performing extensive analysis on the full network is often not practical or desirable. Reduced models of the larger system, referred to as network "equivalents," can be constructed. By using data mining techniques and with the aid of available measurement data, we seek to adaptively reduce the system to obtain models which preserve important structural features of a system. These models are denoted *essential structure preserving (ESP)* models. We are interested in examining the impact of different equivalencing or reduction schemes on properties or purposes of interest, where the desirable features to preserve may vary depending on the application. Examples include LMP characteristics and transient stability response. A notable example is generator coherency, which is considered here. Grouping generators together based on coherency is not a new concept, and [105] illustrates the feasibility of using clustering to form coherent groups. This application presents a solution approach in two stages, a grouping stage and an aggregation stage, which are considered separately. There are several key aspects of this project; we focus on the application of data mining insight to suitably group and aggregate power system elements.

In forming ESP models, we leverage our experience with clustering methods as well as with sensitivity analysis, from Chapter 3. The use of clustering to form subsystems from a larger system is presented. The approach has also been developed into a tool to facilitate the systematic creation of reduced systems that accurately represent the full system with respect to some desirable properties.

Network reduction techniques, especially the Ward reduction [106], have been used for a long time. Ward reduction modifies injection locations but not branch values or the topology of the retained network. Other more recent reduction techniques [107], for example, aggregate buses and optimize branch values between areas so that inter-area flows are reasonably well preserved for various dispatches. Another key example is power transfer distribution factor (PTDF) based equivalents [108] which preserve the impact of bilateral transactions. The validity of equivalents

should be maintained as the system is changing, and we need to understand how frequently and under what conditions modifications are needed. On-line external network modeling [109] differentiates between *external equivalent* methods, which operate without the use of real-time data, and *external solution* methods which use data to solve a state estimation or power flow problem. A forecasting-aided state estimator [110] has a prediction step that uses information from past estimates to model how the system state evolves over time. We believe it is important to develop methods which can tune the equivalent to preserve the properties which are advantageous for the problem at hand.

4.3.1. Coherency-Based Equivalents

Much influence for our work on the equivalencing topic comes from the area of coherency recognition, so here we briefly describe some of those key insights. Coherency-preserving equivalents are a type of dynamic equivalent, which are useful for transient stability studies. In the event that the classical machine model is used, there is a relatively simple relationship [111] that can be used to identify coherent generators, which will be described in this section. It has generally been accepted that a classical generator model is sufficient for representing coherency. If this coherency-based aggregation is only required for external systems, this approximation may be appropriate. One of the interests in this work is to facilitate validation for machines with more complicated dynamic models.

Three classifications of dynamic reduction techniques are modal, physical, and topological [111]. Modal approaches are based on controllability, observability, and time-scale properties of a linearized system. Physical approaches choose model types for components based on their role in particular faults. The topological approach is about aggregation into subsystems, and includes coherency methods. This work can be considered a physical and topological approach.

Using a Ward-type approach in [112], machine inertias are allocated according to distribution factors, and a detailed algorithm for computing the reduction is given. Distribution factors relating the eliminated buses to the retained buses can also be obtained. For coherency-based dynamic equivalents, *coherency* refers to the tendency of generators to swing together. In general, coherency recognition algorithms are focused on predicting the behavior of the

trajectory of nonlinear equations before they are solved. Relationships between parameters that indicate coherency, as well as the conditions of both e-coherency (exact coherency) and n-coherency (near coherency), are discussed in [111].

Early efforts often require transient stability runs to be performed before coherency recognition can be done. One work from 1978 [113] requires simulations with a linearized system and contributes some initial results on the subject. The two key results from [113] are that coherency can be determined by considering a linearized model, and that using a classical machine model is sufficient to show the coherent groups. The overall procedure consists of these three steps: (1) identification of coherent groups, (2) network reduction of the buses, and (3) dynamic aggregation of generating unit models. The need is also recognized for on-line application of coherency-based dynamic equivalents and for sensitivity analysis of the coherent groups to fault location and to switching (opening and closing) of lines.

The subject of identifying coherent groups of generators is further addressed in [114]. Coherent groups are identified based on coupling between generators in the state matrix. An analytical procedure to identify weakly coupled groups and strongly coupled groups through the use of a coupling factor is discussed. Strongly coupled groups are coherent. Weakly coupled groups are independent of fault location. Results show that machines which are geographically close may not be in the same coherent group, which may be a function of the transmission system.

Computational algorithms for constructing modal-coherent dynamic equivalents are described in [115]. Inertial averaging is used to aggregate the generators in each of the coherent groups. A coherency measure for different deterministic disturbances can quantify how much error results due to reducing the order of the equivalent. Generators can then be grouped to retain the lowest order equivalent that satisfies a reasonable degree of accuracy.

The interarea model estimation (IME) problem is defined and analyzed in [116]. In this analysis, a two-area system is used to represent power transfer from one coherent area to another, and it is assumed that there is a center point in an equivalent machine for each area where the voltage is fixed. Interarea dynamics of multi-machine power systems are also analyzed in [117]. “Slow”

coherency is defined to group machines with identical slow motions. A singular perturbation is used to separate slow and fast variables. Slow coherency allows machines with different fast dynamics to be grouped together, allowing the same groupings to be applicable for different fault locations. Groups of generators which exhibit slow coherency are identified and grouped in [118], and these groups are used to demonstrate a controlled islanding strategy on a reconstructed pre-blackout case for the August 14, 2003, blackout. The goal is to contain the impact of the disturbance within the formed islands and thus prevent its propagation to the rest of the system. An EPRI program called DYNRED is used in [118] and [119] to perform dynamic reductions. A reduced system does not necessarily represent the modes of interest, so it is important to know ahead of time which modes to preserve.

There are also dynamic equivalencing methods which do not involve coherency. Motivated by the impact of distribution networks with distributed generation (DG) on transmission system dynamics, a more recent paper (2007) [120] considers dynamic equivalencing of distribution networks. According to [120], coherency methods may not be well suited for a distribution network where numerous DGs exist. A dynamic equivalencing method based on the Hankel norm approximation is discussed in [120]; however, the reduced system cannot be represented as an equivalent generator, and the equivalent must be changed for different operating conditions. The Hankel norm is a measure of the effect of a system's past input on its future output, and it often plays an important role in the field of model approximation [121]. Renewable energy integration and the need for equivalent models of wind farms for transient stability studies motivate recent dynamic equivalencing work. The approach in [122] to wind farm modeling is to combine the wind park into a minimal set of equivalent models by combining all turbines which have the same mechanical natural frequency. Offline identification methods from the System Identification toolbox in MATLAB are used in [123] to determine dynamic equivalents based on measured data from disturbances in the distribution network. The system is regarded as a black box, so information about generators is not known, and an equivalent model is derived for the entire system (rather than just the external system) based on observed input/output data.

As evident from the preceding discussion, an abundance of literature exists concerning coherency and the identification of coherent generators in power systems as well as the use of

coherent groups to form dynamic equivalents. Our approach takes advantage of a number of insights gained from the works mentioned above, and before proceeding, we emphasize the goals of our work:

(1) To do coherency recognition without needing to compute the linearization of the full system, its dynamic equations, and its stator algebraic equations.
(2) To do coherency recognition without needing to simulate the system.
(3) To perform the coherency analysis in a computationally efficient way so that the analysis may be computed quickly for a large system. The intention is that the algorithm should facilitate further applications, including those which may require adaptive calculations of an online equivalent for online transient stability analysis.
(4) To be able to incorporate data into the analysis, in addition to whatever model information we have. Data is valuable for everything from identifying groups to updating the equivalent model and creating "equivalent generators."
(5) To develop a strong validation framework. Validation questions include how coupled these coherent groups are for different faults and how different dynamic models impact the accuracy of the results (i.e. how much is coherency dependent purely on the electromechanical modes).
(6) To use the coherency recognition results in an efficient and effective way for computing the dynamic equivalent model to be used in other applications.

In the following sections, a solution approach is presented in two separate stages, a grouping stage and an aggregation stage. The grouping and aggregation are performed according to selectable criteria, and a tool has been created which encompasses the options applicable to each stage.

4.3.2. Bus Grouping Approach

The first stage of the technique is to determine which network elements can be grouped together. For buses or generators, which are the focus of this section, the groups are referred to as "bus groups." The output of this stage is a listing of the bus numbers belonging to each cluster. The grouping is accomplished by performing cluster analysis on vectors which represent some

relationship(s) we want to capture. The representative vectors which serve as input to the clustering algorithms are generically referred to as *feature vectors*. A tool has been implemented for this work which forms feature vectors and then clusters elements based on these criteria. In the tool, the grouping criteria are selectable. Several instances of the criteria to group generators based on similarity are described in this section. There are a variety of clustering algorithms at our disposal for this work. These include the Quality Threshold (QT), the K-means, and agglomerative (hierarchical) methods, all described earlier. Allowing flexibility in the choice of bus grouping criteria is important since the intention is to facilitate the study and preservation of different properties.

First, a criterion is described to form and cluster the feature vectors based on a metric found in the coherency identification literature [111]. This criterion is worth mentioning because it is both fast and parameter based, and results from its use corroborate well other grouping criteria we examine. From this criterion, some insights are gained which have contributed to our generalized grouping strategy.

In order to adequately explain the connection we wish to exploit, it is necessary to first present some key conditions related to coherency. In general, electromechanical coherency is defined as

$$\left|\Delta\partial_i(t) - \Delta\partial_j(t)\right| < \varepsilon \tag{74}$$

for all generators *i, j* within a coherent group of generators {*G*} and for some period of time, t_c-t_0. Using the definition in [111], exact coherency or e-coherency for a group of generators {G} satisfies the following definition:

$$w_i(\delta) = \delta_1 - \delta_i(t) - \delta_{1i}^S = 0 \tag{75}$$

where δ_1 is a reference angle in {G} and the steady-state post-fault angle difference is given by ∂_{1i}^S. As described in [111], this analysis is based on nonlinear systems theory [124], where e-

coherency is considered as a constrained motion. Consider the trajectory of the differential equations,

$$\dot{\mathbf{x}} = f(\mathbf{x}) \tag{76}$$

which, for exact coherency, must satisfy the following constraint:

$$w(\mathbf{x}) = 0 \tag{77}$$

The necessary and sufficient condition for these to be satisfied is given by the following [124]:

$$\sum_{i=1}^{n} \frac{\partial w_j}{\partial x_i} f_i(\mathbf{x}) = 0 \tag{78}$$

In this formulation, the machines are represented by the classical model,

$$M_i \frac{d^2\delta_i}{dt^2} = P_{mi} - P_i(\delta) \quad \text{for } i = 1,...,n \tag{79}$$

where M_i is the inertia, δ is the rotor angle, P_{mi} is the mechanical power, and $P_i(\delta)$ is the electrical power. Then it is assumed that the electrical power is represented by the simplified equation,

$$P_i = P_{si} + \sum_{j \neq i} b_{ij} \sin \delta_{ij} \quad \text{at } i = 1,...,n \tag{80}$$

where $b_{ij} = |B_{ij}||E_i||E_j|$. In combining these criteria, the coherency condition we are examining states that two generators i and j which are coherent must satisfy the following relationship:

$$\frac{b_{ik}}{M_i} = \frac{b_{jk}}{M_j} \tag{81}$$

where k is any bus other than bus i and bus j, and M_i and M_j denote the inertias of the machines. The relationship in (81) is based only on parameters. The terms b_{ik} and b_{jk} represent the equivalent line impedances when the system is replaced by an equivalent at the nodes of all the generators. This criterion is shown graphically in Figure 13.

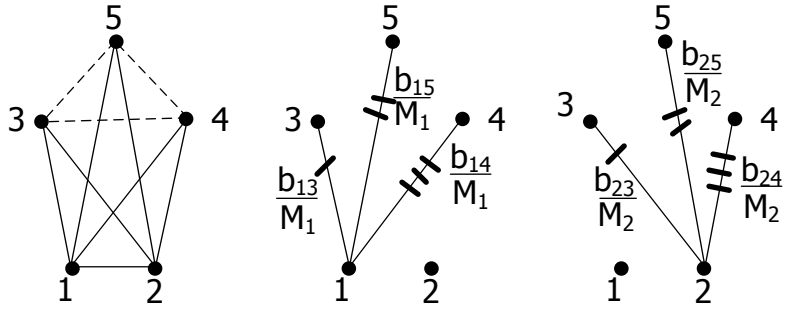

Figure 13. Graphical representation of coherency criteria [111]

The criterion is applied in the following way in [111]. It is assumed that the angles will be approximately equal if the accelerations caused by the fault are equal. The idea is that during the fault, the rotors of the generators are in accelerated motion, so as long as the generators accelerate in the same way, the angle condition

$$\partial_{ij}(t) = \partial_{ij}^{S} \quad \text{for } k \notin \{G\} \tag{82}$$

should be fulfilled, where $t>t_f$, and t_f is the time when the fault is cleared. This criterion is based on several approximations, and one problem it is subject to is a chaining problem. In the simplest case, consider where one generator may be connected radially to another generator which is connected to the rest of the network. These two generators may be coherent, but this criterion would not identify them since they always have different equivalent impedances to the rest of the network.

The above criterion exhibits some desirable characteristics and is useful in identifying coherent generators in many cases. It had also led us to adapt a more general approach of performing the grouping based on sensitivities of the properties of interest. Useful candidate criteria for

determining structural-preserving groups are the elements of the Jacobian inverse sensitivities. The sensitivities in the Jacobian inverse capture the relationships in the way bus angles and voltage magnitudes respond to injections, which captures power flow relationships as well. When considering the Jacobian inverse sensitivities, there are various possibilities in the way one can choose to form feature vectors from its elements. For example, groupings may be based on how angles respond to real power injections, for which it may be appropriate and simpler to use elements of the $\mathbf{B}^{-1}$ matrix of the DC power flow. Alternatively, feature vectors may be formed based on the transpose of the Jacobian inverse, where clustering by this metric groups buses which *provide* or *create* similar changes to bus angles in the system.

Then, there is further flexibility in how many elements each feature vector should contain, and which elements these should be. Consider two extreme cases. In the simplest case, we can examine just one "probe" in the system, i.e. a single injection, and compare the "responses" or (angle) sensitivities at all buses in the system due to that one injection. This has the advantage of being extremely fast to determine, since it amounts to obtaining elements in just one column of the P-θ portion of the Jacobian inverse or $\mathbf{B}^{-1}$, and clustering those, which is simple, since instead of clustering vectors, it is clustering 1-D elements. In the other extreme case, we can cluster the entire Jacobian inverse, considering *all* of its elements equally. The most effective and most representative "probes" for use as the feature vectors lie between these two extremes. That is, to capture enough of the information to effectively represent the relationships or coupling between the buses, only *some* of the sensitivity elements are required.

More investigation is warranted to optimize the selection of the elements. Some ideas for future enhancement in this direction may involve using the SVD to capture the elements representing most of the variability in the data and retaining only the first few corresponding singular vectors. Alternatively, one may choose to look at some probes "near" the region of interest and some far away, and look at how the angles respond to probes distributed throughout the system, without using probes at every single bus. For example, if a bus has no generation and no load (a zero-injection bus), there may be no practical benefit in examining a probe at that location. In summary, identifying groups of buses or generators can be effectively determined according to sensitivities.

4.3.3. Aggregation and Network Reduction

While the clustering described above concerns how to determine element groupings, one must be able to appropriately take the "aggregation" of the devices in these groups and actually produce the simplified model with which computations are to be performed. The behavior of the simplified model must be sufficiently representative of the full system behavior, at least for the characteristics under study. Once groups are formed of elements which have the propensity to respond together, buses in each group must be aggregated to produce an overall equivalent, available for subsequent studies and evaluation. This aggregation is described below.

To generalize the concept fully, transformation theory is useful. Tensor analysis, described by Kron [125], [126], describes the idea that various basic network elements (or basic elements of other types, not just networks) can be connected in different ways, where each way of connection is referred to as a *reference frame*. The problem can thus be conceptually divided into several stages. The idea is to transform the area such that the characteristics of its interconnections remain unchanged. An analogy to this is the commonly used delta-wye transform, which preserves the terminal characteristics as well as the power consumption characteristics. The impedance seen between any two of the terminals, R_{XY}, R_{YZ}, and R_{ZX}, must be preserved. This analogy can be extended to more than three buses, as shown in Figure 14.

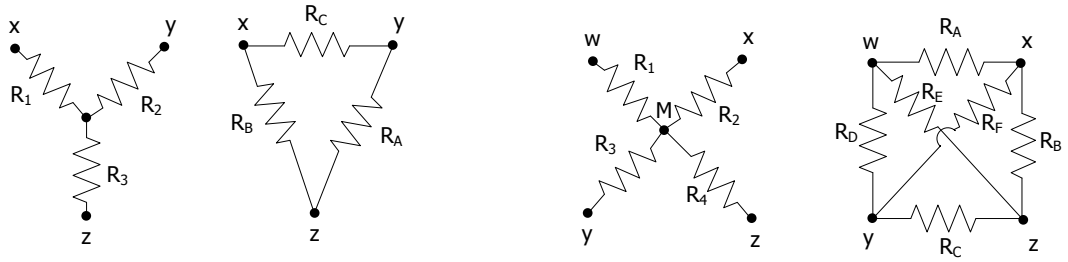

Figure 14. Comparison to delta-wye conversion

The equation to preserve R_{XY} in the three-bus example is given by $R_1+R_2=R_C//(R_A+R_B)$. In the four bus example, R_{WX} is preserved by satisfying $R_1+R_2=Z_{Thev,WX}$, where $Z_{Thev,WX}$ is the Thevenin impedance between terminals *W* and *X*. The bus marked with an *M* is referred to as the mediod bus, which is analogous to the neutral bus in the delta-wye transformation, as shown in Figure 15. The mediod point is arbitrary.

In general, the area is of arbitrary size, and each bus in the area has arbitrary connections to buses inside and outside of the area. Some buses have no connections to the external system. In moving from the first figure to the second figure in Figure 15, we can utilize the Thevenin equivalent impedances between certain buses, determined by calculating elements of the Z_{bus} (Y_{bus}^{-1}) matrix. All generation and load of the area is represented at one bus, the mediod. The external network remains unchanged.

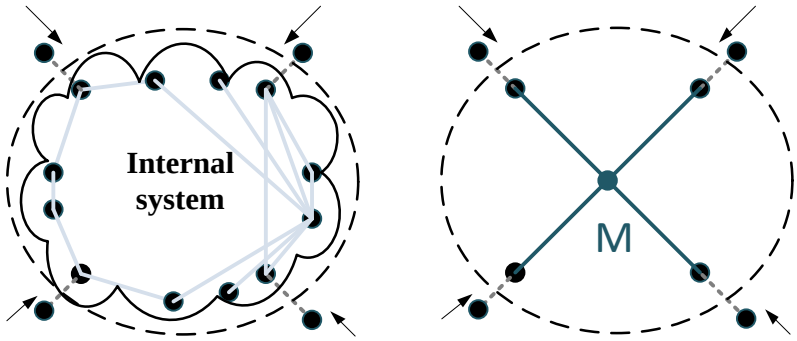

Figure 15. Aggregation overview

The Ward equivalent tends to smear injections among all of the buses, an effect which we deliberately wish to avoid. Rather than lose the identity of the injections, a distinguishing feature of this approach is that characteristics of areas are preserved together. This is facilitated by our grouping approach, described in the previous section.

Now let us describe the types of network transformations that can produce an acceptable equivalent. The problem is complicated (or simplified, depending on perspective) by the fact that there in fact is no unique solution. There are many possible ways to aggregate buses, so that if specific criteria are not applied as constraints, any way may be acceptable. After introducing and implementing several different approaches to arrive at the final reduced system, the point that they are all fundamentally the same, and are all acceptable, has become apparent, as described below. The conclusion from this stage of the work is that all "exact" aggregation methods are essentially variations of a Ward equivalent.

By definition, a Ward equivalent provides an exact matching of the power flow state, **V** and **θ**. In each group of buses denoted *Group_i*, we can easily determine a subset consisting of boundary buses which are connected to both internal and external buses. The internal buses may be eliminated and all of their adjusted injections may be moved to the boundary buses, leaving the external system unchanged. This results in the following, which provides an exact solution:

$$\begin{bmatrix} \mathbf{I_i} \\ \mathbf{I_b} \\ \mathbf{I_e} \end{bmatrix} = \begin{bmatrix} \mathbf{Y_1} & \mathbf{Y_2} & \mathbf{0} \\ \mathbf{Y_4} & \mathbf{Y_5} & \mathbf{Y_6} \\ \mathbf{0} & \mathbf{Y_8} & \mathbf{Y_9} \end{bmatrix} \begin{bmatrix} \mathbf{V_i} \\ \mathbf{V_b} \\ \mathbf{V_e} \end{bmatrix} \tag{83}$$

where $\mathbf{Y_3}$ and $\mathbf{Y_7}$ are zero matrices. From the first equation,

$$\mathbf{V_i} = \mathbf{Y_1}^{-1}\left(\mathbf{I_i} - \mathbf{Y_2}\mathbf{V_b}\right) \tag{84}$$

which is substituted into the second equation of (83) to give

$$\mathbf{I_b} = \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{I_i} - \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{Y_2}\mathbf{V_b} + \mathbf{Y_5}\mathbf{V_b} + \mathbf{Y_6}\mathbf{V_e} \tag{85}$$

$$\overbrace{\mathbf{I_b} - \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{I_i}}^{\mathbf{I_{beq}}} = \overbrace{\left(\mathbf{Y_5} - \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{Y_2}\right)}^{\mathbf{Y_{beq}}}\mathbf{V_b} + \mathbf{Y_6}\mathbf{V_e} \tag{86}$$

The third equation in (83) is unchanged, and the new combined system equations are the following:

$$\begin{bmatrix} \mathbf{I_b} - \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{I_i} \\ \mathbf{I_e} \end{bmatrix} = \begin{bmatrix} \mathbf{Y_5} - \mathbf{Y_4}\mathbf{Y_1}^{-1}\mathbf{Y_2} & \mathbf{Y_6} \\ \mathbf{Y_8} & \mathbf{Y_9} \end{bmatrix} \begin{bmatrix} \mathbf{V_b} \\ \mathbf{V_e} \end{bmatrix} \tag{87}$$

From any system which preserves the system state as in (87), one may obtain any number of other configurations. To see this, consider an arbitrary transformation matrix **T**. It is obvious that multiplying both sides of (87) on the left with the matrix **T** results in a new equivalent system. The more restrictive problem is to define the elements of **T** such that certain constraints are met.

In our formulation, a mediod bus represents the electrical center of the cluster. As described above, the choice of mediod is arbitrary, at least with respect to static properties (more investigation into the impact of mediod choice with respect to dynamic properties may be warranted, as validation proceeds). For the sake of simplicity, we hereafter choose the mediod bus to be one of the boundary buses. All of the generation and load in the area can be aggregated at this bus. The benefit is that the resulting equivalent provides an intuitive representation of the generation and load in the area, and is set up to facilitate further analysis such as dynamic equivalencing and model reduction of generators. If we segment the boundary bus current vector $\mathbf{I_{beq}}$ into the mediod and other boundary buses, $\mathbf{I_{beq}} = \mathbf{[I_M, I_{bb}]^T}$, we can investigate what is needed to move the injections (equivalent generation and load) at the boundary buses to the mediod bus. That is, the goal is to determine $\mathbf{Y_{beq}'}$ and $\mathbf{I_{MEQ}}$ such that $\mathbf{I_{beq}' = [I_{MEQ}, 0]^T}$. Rather than attempt to present all the possible solutions, we simply present the *requirements* for any acceptable solution. Supposing the original system equations are

$$\begin{bmatrix} \mathbf{I_M} \\ \mathbf{I_{bb}} \\ \mathbf{I_e} \end{bmatrix} = \begin{bmatrix} \mathbf{Y_1'} & \mathbf{Y_2'} & \mathbf{Y_3'} \\ \mathbf{Y_4'} & \mathbf{Y_5'} & \mathbf{Y_6'} \\ \mathbf{Y_7'} & \mathbf{Y_8'} & \mathbf{Y_9'} \end{bmatrix} \begin{bmatrix} \mathbf{V_M} \\ \mathbf{V_{bb}} \\ \mathbf{V_e} \end{bmatrix} \tag{88}$$

then the equations which must be satisfied are

$$\mathbf{T} \cdot \begin{bmatrix} \mathbf{I_M} & \mathbf{I_{bb}} & \mathbf{I_e} \end{bmatrix}^{\mathbf{T}} = \begin{bmatrix} \mathbf{I_{MEQ}} & \mathbf{0} & \mathbf{I_e} \end{bmatrix}^{\mathbf{T}} \tag{89}$$

$$\mathbf{T} \cdot \begin{bmatrix} \mathbf{Y_1'} & \mathbf{Y_2'} & \mathbf{Y_3'} \\ \mathbf{Y_4'} & \mathbf{Y_5'} & \mathbf{Y_6'} \\ \mathbf{Y_7'} & \mathbf{Y_8'} & \mathbf{Y_9'} \end{bmatrix} = \begin{bmatrix} \mathbf{Y_1''} & \mathbf{Y_2''} & \mathbf{Y_3'} \\ \mathbf{Y_4''} & \mathbf{Y_5''} & \mathbf{Y_6'} \\ \mathbf{Y_7'} & \mathbf{Y_8'} & \mathbf{Y_9'} \end{bmatrix} \tag{90}$$

where $\mathbf{I_{MEQ}}$ is a free variable to be determined such that the boundary bus injections are made zero, and the external injections are unchanged. Here, **T** is a 3x3 matrix, so 9 variables may be chosen to satisfy the constraints. A typical constraint is to require that the network connecting the external buses remain unchanged. Thus, we desire any matrix **T** which satisfies (89) and (90).

In summary, the only requirement of any system transformation **T** is to satisfy constraint equations of the form above. For the subsequent analysis, we use the simplest instance of such a solution, leaving further investigations into other transformations as an open subject. Hereafter we convert the current injections at the boundary buses into diagonal **Y** elements and then adjust the current at the mediod bus so that all of the area generation and load is aggregated at that bus. This results in the following adjustments. For each boundary bus which is not the mediod, the added shunt element is

$$ysh_{bb} = -\mathbf{I_{bb}} / \mathbf{V_{bb}} \tag{91}$$

where $\mathbf{I_{bb}}$ is the current injected into the bus. Then, at the mediod bus, we adjust the shunt so that the current injection is equal to the net power of the group. Similarly, where $\mathbf{I_M}$ is the current injected into the mediod bus, and $\mathbf{I_{net}}$ is the equivalent current injection into the bus representing the aggregate generation and load in the group, we obtain the equivalent shunt element,

$$ysh_m = -\left(\mathbf{I_M} - \mathbf{I_{net}}\right) / \mathbf{V_M} \tag{92}$$

and new mediod current is $\mathbf{I_{net}}$.

From this frame or representation, we can transform the system further if we wish. A system is readily transformable from one frame to another, as discussed above. The point is that such a system configuration is not unique; there is an entire space of systems which preserve the steady-state characteristics. The existence of this flexibility in the space of equivalent solutions (which preserve the power flow characteristics) may warrant further investigation. One can investigate this transform-space further and investigate the constraints needed on the transformation to obtain equivalents of various desirable forms. All we require for the purposes of this work is *one* "exact" method with which to compare any approximations, yet from the above discussion it is clear that more than one "exact" method may exist and be capable of producing an acceptable network model.

4.3.4. Equivalencing Examples

This section presents examples showing how different reduced systems respond to an event, in comparison with the response of the full system. The events shown were specifically chosen because they are faults occurring just past their critical clearing times in the full system ("critically unstable"). We are interested in examining how the response is impacted in the equivalents. As a case study, let us consider the three groups depicted in Table 16.

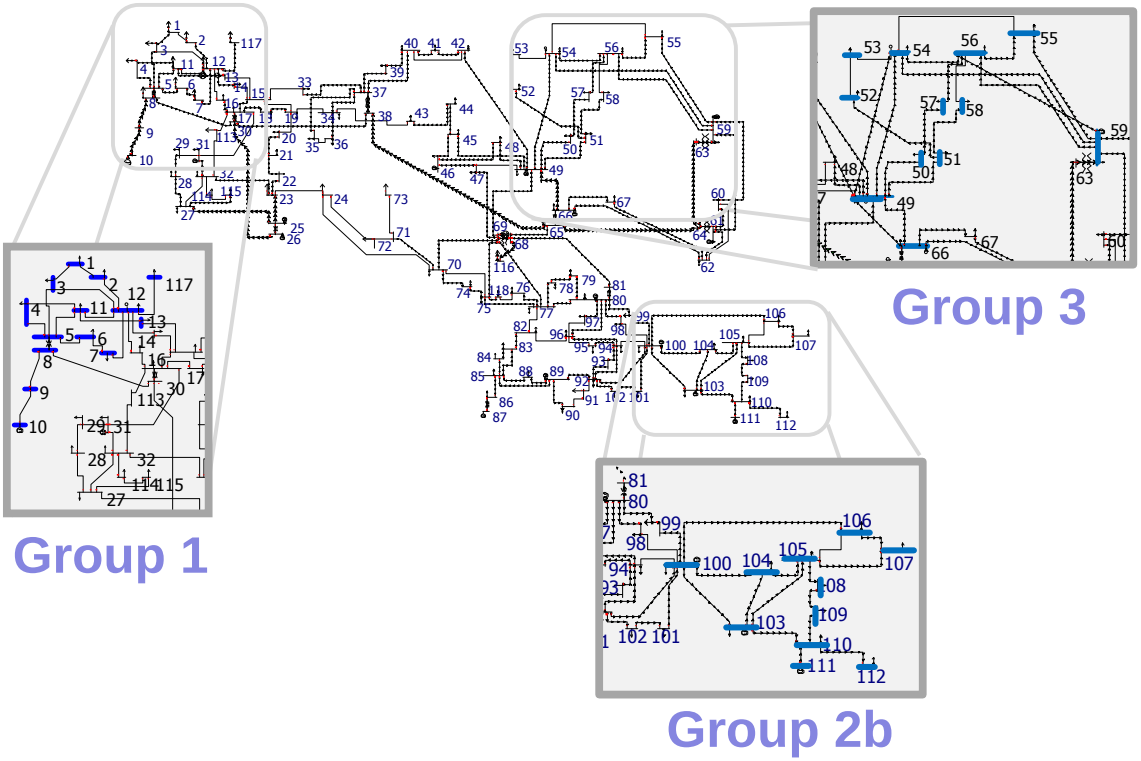

Figure 16. Location of the three groups, shown before aggregation

For these examples, the determination of each group is based on a single probe or real power injection in the same area. This corresponds to a single column of the θ-P portion of the Jacobian inverse; these scalars are contoured as percentages in Figure 17, where the highest values are shown in red. For each of the three groups, a separate equivalent is made. Each of the equivalenced areas after network reduction is shown in the boxes in Figure 18, overlaid onto the original locations of the groups in the full system. Each of the reduced systems is denoted by its group name.

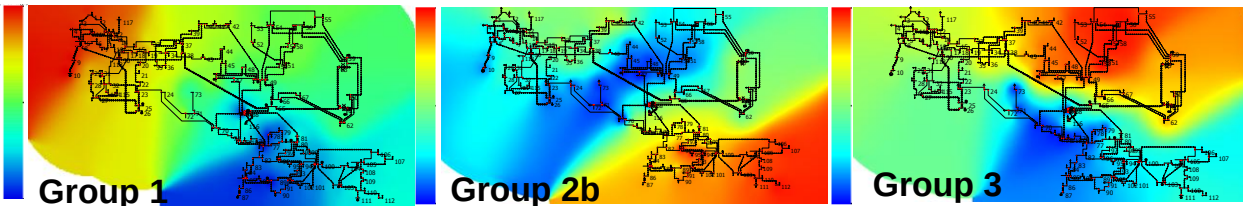

Figure 17. Contours of angle sensitivities showing the three groups

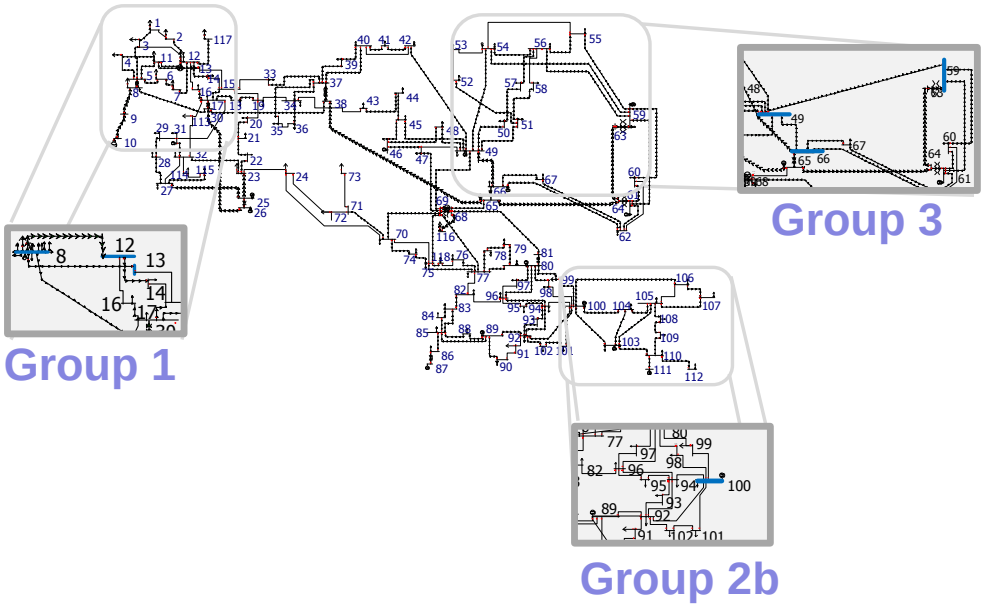

Figure 18. The three groups, shown after aggregation

As described in the aggregation sub-section, the power flow steady-state solution is exact for each of the equivalents. The error vectors of voltage magnitude and angle are plotted in Figure 19 for each of the equivalent systems compared to that of the full system. The x-axis of the plots indicates bus numbers, which are ordered to list the boundary buses first. While all errors are considered negligible, it is apparent that some buses do have slightly more error. Since by design, the constructed equivalents satisfy power flow solution equivalence, the more interesting questions are in looking at what happens as the system changes.

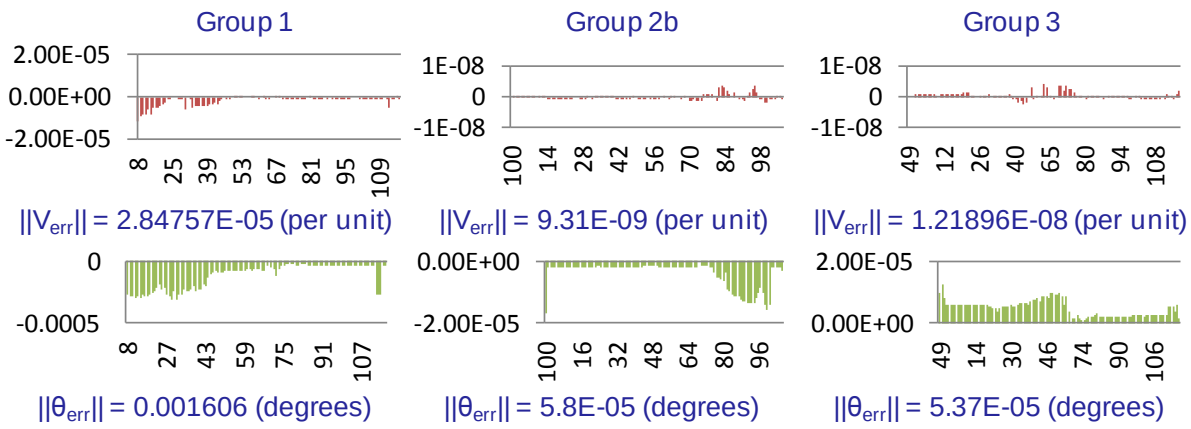

Figure 19. Power flow solution comparisons

Results for three events are presented in this section, depicted in Figures 20-28. For the first event, a fault is simulated near Group 1, at Line 26-30 near bus 30, as shown in Figure 20. The full system is unstable, as seen from the plot of generator rotor angles in Figure 21. As evident from the one-line diagram, bus 30 borders the aggregated area, so this scenario should help stimulate problems which we can subsequently investigate. The response of the rotor angles in each of the three equivalents is illustrated in Figure 22. When comparing the full system response to that of the Group 1 equivalent, Generator 12 responds in a noticeably different manner. One can clearly see the impact of combining Generators 10 and 12, as occurred when forming the Group 1 equivalent. Thus, for certain events, generators 10 and 12 respond independently of each other.

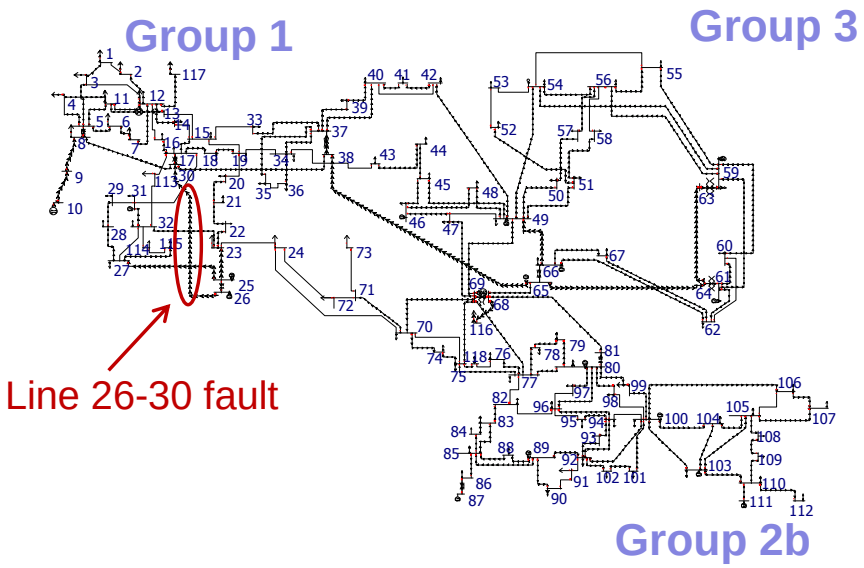

Figure 20. Location in full system of Line 26-30* fault

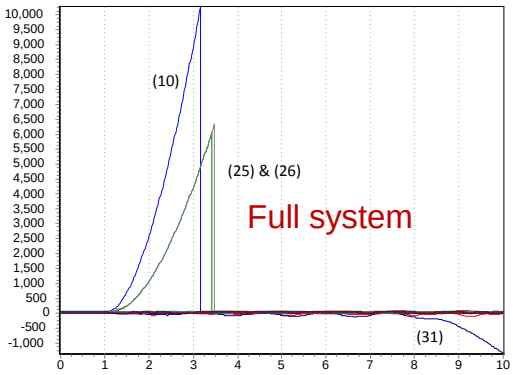

Figure 21. Rotor angles in full system, Line 26-30* fault

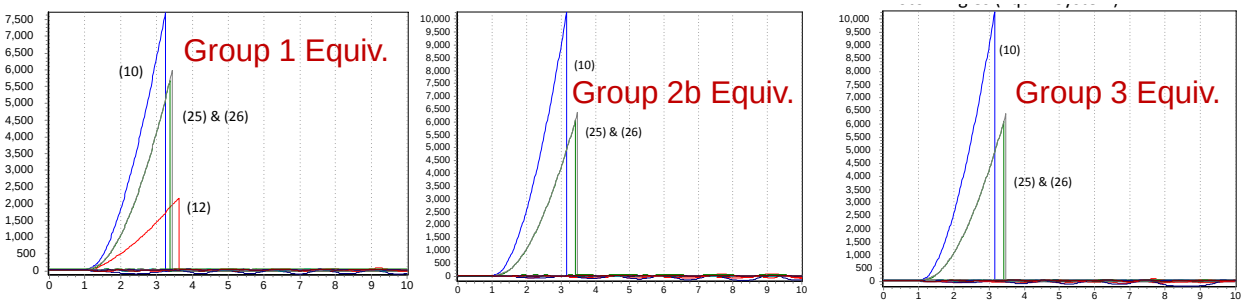

Figure 22. Rotor angles in reduced systems, Line 26-30* fault

The second example considers an event which is *not* near Group 1, a fault on the line between buses 89 and 92, shown in Figure 23. The full system's critically unstable response is shown in

Figure 24. The response of the rotor angles in each of the three equivalents is illustrated in Figure 25. In each equivalent, the response is well-preserved. Considering that the impact of faults tends to be relatively local, this result is not surprising for the Group 1 or the Group 3 equivalent. However, the fault is near Group 2b, and the response of Group 2 is well-preserved as well.

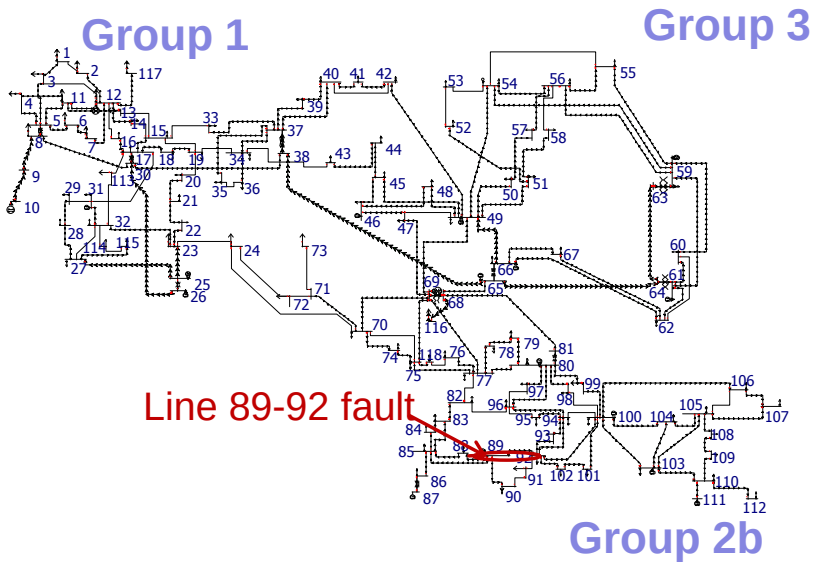

Figure 23. Location in full system of Line 89*-92 fault

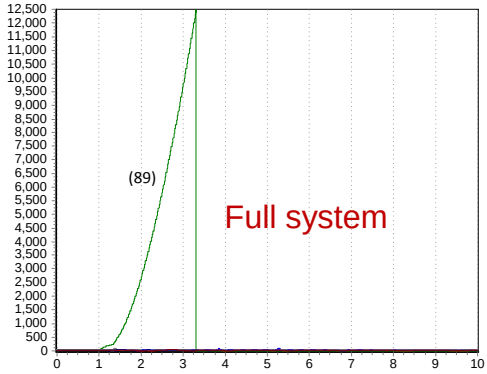

Figure 24. Rotor angles in full system, Line 89*-92 fault

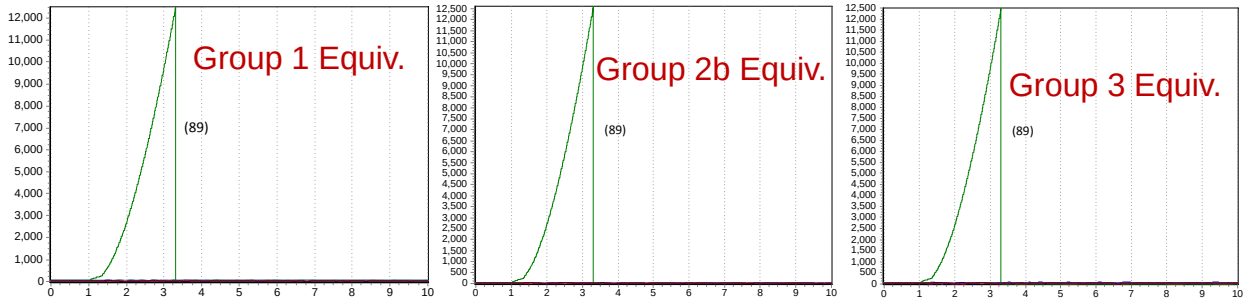

Figure 25. Rotor angles in reduced systems, Line 89*-92 fault

A third event is shown in Figure 26 for a fault at Bus 25, again near Group 1. The response in the full system is shown in Figure 27. When looking at the responses of the three equivalent systems, shown in Figure 28, it is immediately clear that this situation is exactly what we want to prevent. The full model shows that the system is unstable for this fault, yet the equivalent for Group 1 shows that it is stable. The responses in Group 2b and Group 3 are much more similar to the full system, except for a noticeable difference in the behavior of Generator 12.

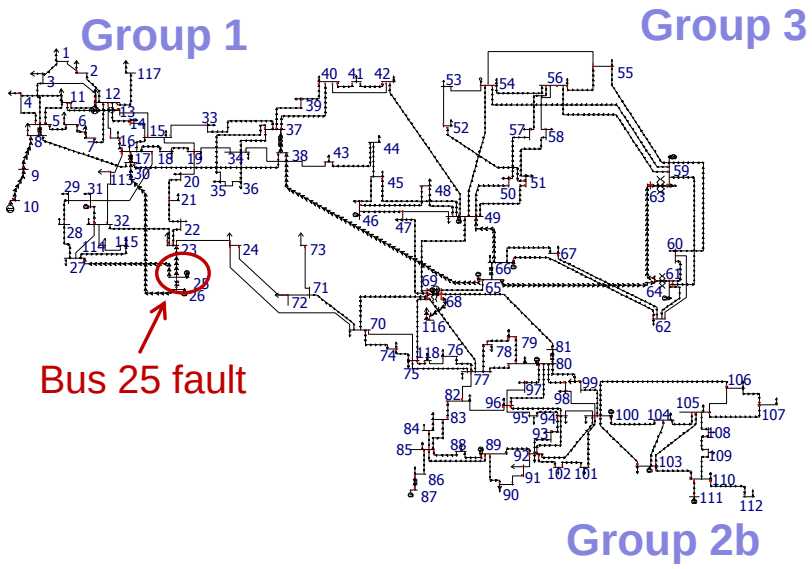

Figure 26. Location in full system of Bus 25 fault

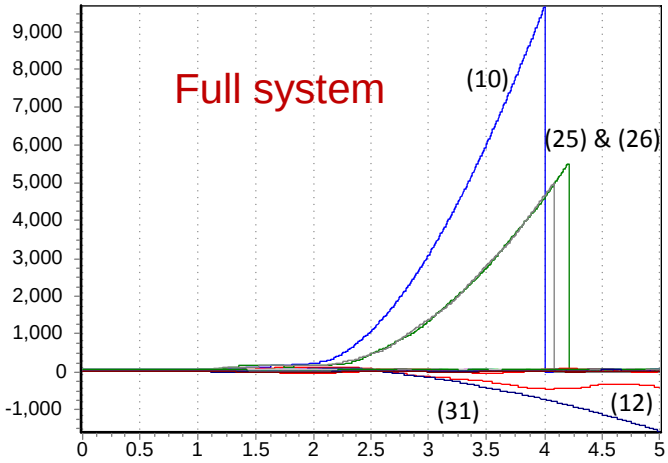

Figure 27. Rotor angles in full system, Bus 25 fault

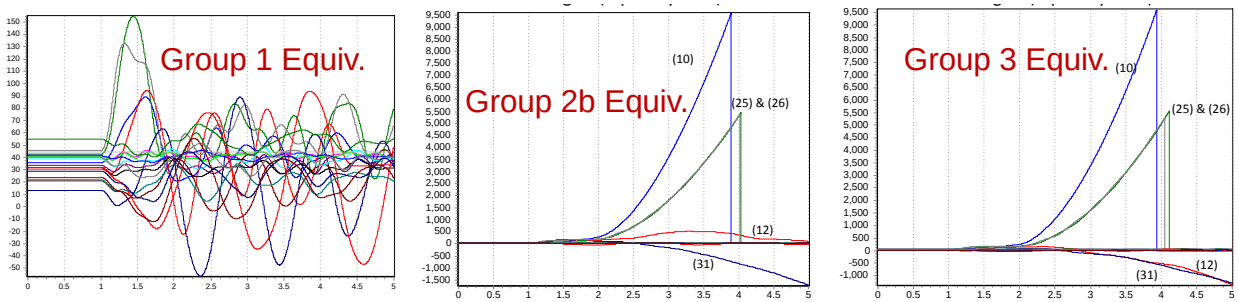

Figure 28. Rotor angles in reduced systems, Bus 25 fault

This example provides compelling motivation to pursue this work, and to develop this type of tool, to gain a better understanding of model limitations. During a transient stability simulation, it is possible that the initial groups may lose their validity as the system exhibits nonlinear behavior and two buses or generators which started out responding together diverge.

4.4. Transmission Line Parameter Estimation

This section discusses the estimation of transmission line parameters from historical SCADA data. Details and results are presented in [127]. A close analogy is presented in Section 4.2 and [100] for using PMU data collected over time to construct a reduced equivalent model at the PMU buses. The focus in this application is on transmission system model errors, which can have an adverse affect on the accuracy of results. Such model errors are often studied in the

context of state estimation. There is existing work in the field aimed at identifying and correcting transmission model parameter errors using telemetry data [128], [129], [130], [131], [132]. Uncertainty in data with respect to state estimation is analyzed in [133]. While a few methods provide a means for updating the estimates online, the approaches generally only consider a single snapshot in time. Also, typically only results for small simulated systems are presented, as opposed to real systems. This work avoids both limitations. The focus is on using the volume of historical SCADA data collected over time, which is readily available to many utilities, to extract transmission line model parameters.

In this section, the theoretical analysis is presented. As described in Section 4.2, the idea is to utilize "data over time," which is advantageous over methods which consider a single snapshot. The application to real systems with real data is presented in Chapter 5. A key aspect of this work involves being able to perform analysis such that any errors in the available data do not corrupt the estimate. The challenges of dealing with uncertainty and errors encountered in real-world data are dealt with in Chapter 5.

First, a brief background of the transmission line parameter estimation problem is provided. Then, the basic line parameter estimation method is presented. The estimation is applied to a simulated test case in Section 4.4.3, and in Chapter 5 its application to real SCADA data from a North American utility is presented and discussed. By estimating the model parameters of individual lines, this approach may be used for validating the parameters of existing network models and identifying potential model errors.

4.4.1. Transmission Line Model and Available Data

In this section, the transmission line model is presented in conjunction with the measured data used in this work. First, it is necessary to consider the possibilities and limitations afforded by the model. This is an idealized or best-case perspective of what can be estimated about the model based on available data. Considering error and real data introduces complications, as examined in Chapter 5.

The transmission line model considered is the *pi model* in Figure 29, often used to model lines of medium length [134] (approximately 50 to 150 miles).

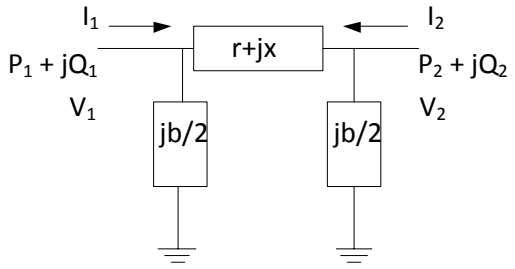

Figure 29. Pi line model

Let V_1, V_2, I_1, and I_2 denote the magnitudes of the phasor quantities. Note that what is physically measured by potential transformers (PTs) and current transformers (CTs) is only voltage and current, respectively; however, these *true* raw measurements are unavailable. The measurements ("raw" SCADA data) for this application are P_1, Q_1, V_1, P_2, Q_2, and V_2, recorded every five minutes.

The goal is to use the available SCADA measurements to estimate values of *r*, *x*, and *b* in Figure 29. The assumption is that the angle difference across the line and the line parameters are unknown. If it is possible to know the angle difference across the line (as with PMUs), the problem is considerably simplified.

The strengths and the limitations of SCADA measurements are now explored. The presence of data errors and noise is temporarily ignored for this chapter. The pi model reflects the relationships that are expected to appear in the data, in light of the assumed *known* or measured quantities.

Several quantities can be derived exactly from this data, at any given time. Since P_1, Q_1, P_2, Q_2, are measured, the complex powers S_1, S_2 and apparent powers $|S_1|$, $|S_2|$ at the line ends are directly known. From apparent powers and voltage magnitudes, the current magnitudes, I_1, I_2, can be directly computed.

$$I_i = |S_i| / V_i \tag{93}$$

From the current magnitudes, power factor angles are also known at both line ends, $\theta_{1,pf}$, $\theta_{2,pf}$.

$$S = VI \angle \theta pf \tag{94}$$

The above quantities can be exactly computed from the SCADA data. Notably lacking are the line parameters and the angle difference across the line.

In the pi model, let *b/2* denote the value of the shunt modeled at each bus. The admittance matrix of the transmission line model is **Y**=**G**+*j***B**, where

$$\mathbf{Y} = \begin{bmatrix} \frac{1}{r + jx} + jb/2 & -\frac{1}{r + jx} \\ -\frac{1}{r + jx} & \frac{1}{r + jx} + jb/2 \end{bmatrix} \tag{95}$$

The real and reactive components of **Y** are matrices **G** and **B**, which consist of the following elements:

$$G_{ii} = \frac{r_{ij}}{r_{ij}^{\ 2} + x_{ij}^{\ 2}} \qquad G_{ij} = -\frac{r_{ij}}{r_{ij}^{\ 2} + x_{ij}^{\ 2}} \tag{96}$$

$$B_{ii} = \frac{-x_{ij}}{r_{ij}^{\ 2} + x_{ij}^{\ 2}} + b/2 \qquad B_{ij} = \frac{x_{ij}}{r_{ij}^{\ 2} + x_{ij}^{\ 2}} \tag{97}$$

The equations for real and reactive power flow at both ends of the line are given by the following:

$$P_1 = V_1^2 G_{11} + V_1 V_2 \left[G_{12} \cos(\theta_{12}) + B_{12} \sin(\theta_{12}) \right] \tag{98}$$

$$P_2 = V_2^2 G_{22} + V_2 V_1 \left[G_{21} \cos(\theta_{21}) + B_{21} \sin(\theta_{21}) \right] \tag{99}$$

$$Q_1 = -V_1^2 B_{11} + V_1 V_2 \left[G_{12} \sin(\theta_{12}) - B_{12} \cos(\theta_{12}) \right] \tag{100}$$

$$Q_2 = -V_2^2 B_{22} + V_2 V_1 \left[G_{21} \sin(\theta_{21}) - B_{21} \cos(\theta_{21}) \right] \tag{101}$$

From (95)-(101), the exact equations for the real and reactive power losses in the line are thus given by P_1+P_2 and Q_1+Q_2 respectively:

$$P_{losses} = \left(V_1^2 + V_2^2\right) G_{11} + 2 \cdot V_1 V_2 \left[G_{12} \cos(\theta_{12}) \right] \tag{102}$$

$$Q_{losses} = -\left(V_1^2 + V_2^2\right) B_{11} - 2 \cdot V_1 V_2 \left[B_{12} \cos(\theta_{12}) \right] \tag{103}$$

Thus, four out of the six equations in (98)-(103) are independent. The loss equations are obviously not independent of the first four since they are each the sum of two other equations. In equations (98)-(101), the unknowns are G_{12}, B_{12}, $b/2$, and $\theta_{12} = (\theta_1-\theta_2)$. Thus, there are four equations and four unknowns.

This looks similar to a traditional state estimation problem except with different unknown variables. The difficulty is that the unknowns appear in the equations in a nonlinear manner. Thus, it follows that use of a linear technique to estimate these quantities will not yield complete success. To see this, observe that what appears in the equations is always a product term of G_{ij} and B_{ij} multiplied with either the term $\sin(\theta_{12})$ or $\cos(\theta_{12})$. Solution requires a nonlinear iterative approach such as Newton's method to find b, x, and r to minimize the difference between the measurements and the calculated quantities. Essentially, this is state estimation. The key difficulty is that from one snapshot, there are more unknowns than equations so traditional state estimation is not sufficient. This application shows that it is possible to improve the ability to directly estimate b, x, and r using data collected over multiple points in time.

The four unknowns, G_{ij}, B_{ij}, and $b/2$ are not dependent on time. This is true generally, at least over some window of time, whereas the angle difference θ_{12} is a function of the operating point and may change over time. Thus, the terms in (98)-(101) which depend on time are the following:

$$\begin{aligned} &G_{12} \cos(\theta_{12}), \quad B_{12} \sin(\theta_{12}) \\ &G_{12} \sin(\theta_{12}), \quad B_{12} \cos(\theta_{12}) \end{aligned} \tag{104}$$

The approach of considering multiple snapshots in time to estimate the transmission line parameters is only successful if there are fewer unknowns which depend on time than the number of linearly independent equations. In this case, there are four linearly independent equations. Each additional time point will produce four more equations but will also produce four more unknowns. Thus, in this problem, some approximation is necessary. The choice of the approximation is therefore important.

4.4.2. Estimation Approach

The equations used to perform the actual estimation are presented in this section. The following two equations for real power losses and reactive power losses contain the three unknowns (*r*, *x*, and *b*) which are time independent.

$$P_{loss} = I^2 r \tag{105}$$

$$Q_{loss} = -V_1^2 \left(b/2\right) - V_2^2 \left(b/2\right) + I^2 x \tag{106}$$

The need for approximation arises since the current magnitude *I* in (105) and (106) is not known exactly from the measurable quantities. That is, due to current injections from the shunt elements, *I* differs from I_1 and I_2 which are known from (93). As an approximation, the average of the two line end current magnitudes I_1 and I_2 is taken for *I*. This avoids the problem of unknowns which depend on time.

Other variations of these estimation equations are possible depending upon which approximations are desirable to make. A bias term *u* is added to the estimation equation for *r*,

$$P_{loss} = u + I^2 r \tag{107}$$

$$Q_{loss} = -\left(\frac{V_1^2 + V_2^2}{2}\right) b + I^2 x \tag{108}$$

which is referred to as the P_{loss} static error and represents a fixed power offset in the measurements.

Based on (107) and (108), a system of equations is obtained by appending the contributions of repeated measurements to the matrix **A**. This may be expressed as (109) or (110):

$$\begin{bmatrix} 1 & I^2 & 0 & 0 \\ 0 & 0 & -I^2 & \left(V_1^2 + V_2^2\right)/2 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} u \\ r \\ x \\ b \end{bmatrix} = \begin{bmatrix} P_{loss} \\ Q_{loss} \\ \vdots \end{bmatrix} \tag{109}$$

$$\mathbf{A}\mathbf{e}_{\mathbf{y}} = \mathbf{f}_{\mathbf{p,q}} \tag{110}$$

As additional rows are added corresponding to additional time points, this becomes an overdetermined problem. Then, least squares estimation minimizes the residual,

$$\mathbf{r} = \mathbf{A}\mathbf{e}_{\mathbf{y}} - \mathbf{f}_{\mathbf{p,q}} \tag{111}$$

and results in the following estimate (109):

$$\mathbf{e}_{\mathbf{y}} = [\mathbf{A}^{\mathbf{T}}\mathbf{A}]^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{f}_{\mathbf{p,q}} \tag{112}$$

where $\mathbf{e}_{\mathbf{y}} = [u, r, x, b]$ denotes a vector of the estimated quantities. The vector $\mathbf{f}_{\mathbf{p,q}}$ contains real power losses P_{losses} and the negative of reactive power losses Q_{gen}. It is possible to estimate P_{losses} and Q_{gen} independently since the equations in (109) are completely decoupled.

4.4.3. Simulated SCADA Data

In this section, the procedure described above is applied to simulated test system data to serve as a reference. To perform these simulations, PowerWorld Simulator is used via SimAuto script commands. A program is written where SimAuto is used to retrieve SCADA data from the simulated system. Then, the program is used to perturb the system to a new operating point. By repeating this procedure, data for a number of operating points is obtained. Simulations provide a base case since both the model and the data are exactly known. The real value of the approach is shown in the Chapter 5, where its application to real data is presented, and error and noise are considered.

The system used for the simulations is the seven bus system shown earlier in Figure 6. The line presented as an example is line (3,4) between buses 3 and 4, with the following actual line parameters: r=0.01 pu, x=0.03 pu, b=0.02 pu. A simulated "error-free" SCADA dataset for line (3,4) is obtained from five different operating points, given in Table 18. The units of P, Q, and V measurements are MW, MVAr and per unit, respectively.

Table 18. Set of five error-free operating points

P_3	Q_3	V_3	P_4	Q_4	V_4
-23.952	-11.235	0.9945	24.021	9.4514	1
-42.009	-4.4517	0.9947	42.189	3.0011	1
-60.035	2.0214	0.9947	60.4	-2.9153	1
-78.055	8.1995	0.9946	78.68	-8.3155	1
-96.086	14.092	0.9944	97.043	-13.211	1

Using only the points in Table 18 with no added error, the estimate matches the model well, as expected. The values of the estimates are given in Table 19. These results exhibit small errors which are attributed to the approximations made, as discussed in Section 4.4.1. An avenue of future work is to investigate the impact of the model approximation on the estimate. Real and reactive power plots of the actual data, the estimate, and the model are shown in Figure 30.

Table 19. Study system noise-free estimates

Line		u (pu)	r (pu)	x (pu)	b (pu)
(3, 4)	Estimate	-1.99e-5	0.010055	0.030161	0.020058
	Model	-	0.01	0.03	0.02

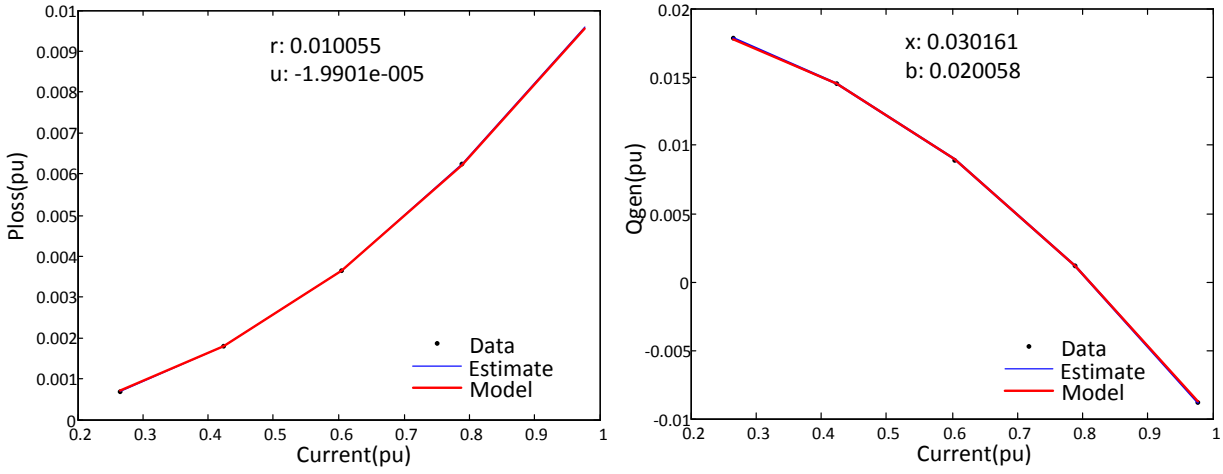

Figure 30. Seven-bus error-free P_{loss} plot (left) and Q_{gen} plot (right)

Cases showing simulated data with error, as well as cases showing the application to real data, are discussed in Chapter 5.

4.5. Conclusions

This chapter provides insight into the nature of network modeling by illustrating the benefit of data-over-time. It is shown that data observed at multiple time points can be used to obtain more information about the system, especially static model parameters. Developing and validating acceptable models is important since such models underlie most power systems analysis. The model needs to represent the underlying physical system, and data can help validate the consistency of these models, even as the system is changing.

The first application of Chapter 4, presented in Section 4.2 and [100], estimates a network model matched to the PMU buses, building upon the analysis in Chapter 3. When underlying models of the system are unavailable or incomplete, such an estimate can be of great practical use. We show mathematically what the estimate gives us, and what the major assumptions are that impact the quality of the estimate. The conclusion from Section 4.2 is that estimation of a static equivalent model from the data, assuming no knowledge of the underlying system, is possible. However, its accuracy depends upon factors over which we seem to have no control, the load variations at each bus in the system. In the circumstance that the injections at the non-PMU buses are zero, and variations occur at all of the PMU buses, then the estimate obtained will be the desired true equivalent. This issue is a key challenge and much discussion is found in [103], whose authors believe that AGC minimizes the error associated with ignoring the external injections. The impact of AGC on the estimate remains to be more fully studied. Future work should explore this estimation procedure applied to real data and investigate the conditions under which the required adjustment term in (54) is zero. It may be possible to find a way to improve the quality of the estimate through the adjustment term. In summary, true insight into the external system load dynamics (57) requires studying this application in conjunction with real-world data; such data was not available for this study. Obtaining and analyzing real data in context with this application would provide a better understanding of how severe the “external factors” are and how much they will truly affect the estimation.

To identify the cause of changes observed in the estimated equivalent, some information about the base case system is needed. If the assumptions are extended to assume that a base-case topology of the system is known, it may be possible to determine a mapping to the equivalent of a desired change in the original system. For this to work, the estimated equivalents must be distinguishable from each other and from the no-outage estimate. Further work can investigate the identification and classification of these event signatures. The challenge is that the mapping of events in the full system to changes in equivalent system is not necessarily unique.

There is great potential to further develop the method in Section 4.2. In addition to addressing the items listed above, the impact of number of time samples and the frequency with which they are taken can be studied. This will become important when real data is used. The estimate can also be acquired recursively using system identification [24], introduced in Chapter 2, or any of the available iterative estimation algorithms [135]. Methods include incremental gradient methods and the extended Kalman filter, which can also be considered an incremental Gauss-Newton method [136], [137]. Incremental methods such as those in [136] and [137] produce an estimate which becomes available as data is accumulated. The role of observability in obtaining the most accurate estimate can also be investigated. There is considerable prior work with respect to PMU placement for observability which can be leveraged for this purpose. The more information we have about the system, the better of an estimate we should be able to obtain. Also, it would be desirable for the estimated equivalent to incorporate additional forms of partial knowledge, such as certain lines with known statuses. It is expected that reactive power and voltage magnitude data can improve the estimate, and such data is also available from PMUs and would be relatively straightforward to include.

The second application in Chapter 4 presents an approach for developing structure-preserving equivalents. Motivated by the prospect of facilitating online transient stability analysis for the quick detection and remediation of events, Section 4.3 creates equivalents with the potential to be updated online as the system is changing. The sensitivity and clustering insight from Chapter 3 allows network elements to be grouped and aggregated to obtain reduced system models which preserve important structural features. The important features to preserve depend upon the application. Buses are grouped by clustering vectors which represent relationships we seek to

capture. To facilitate online transient stability analysis, we intentionally use clustering metrics which are parameter-based, so that the computations may be as fast as possible. The generation and load of each group is aggregated at the mediod bus, and dynamic models are left unchanged.

The structure-preserving network reduction process in Section 4.3 is separated into three stages: grouping, aggregation, and evaluation. A tool has been created for each stage, providing a framework for designing and evaluating reduced models. Flexibilities are necessarily present at each stage and account for the fact that the goals of the equivalent depend on the application, and it must be possible to customize the equivalent to achieve the desired characteristics.

The aggregation in Section 4.3 has the advantage that the generation and load models in an area are retained at the mediod bus to represent the response of the group. A next step is dynamic equivalencing, i.e. changing the generator and load models to represent aggregate behavior. By having successfully "moved" all generation and load to a single bus, we expect that the remaining problem of creating a dynamic equivalent of these models is simplified and can be solved in a standardized way. For example, analysis in Section 3.1.2. can be applied to create dynamic equivalents which preserve certain modal responses.

An interesting network characteristic is explored in Section 4.3. Our analysis shows that a model which preserves the system state can be transformed into other possible "exact" configurations. Investigations into this space of possible transformations is left an open subject. In addition to the genre of exact solutions, we have also found that there a number of solutions which are not quite exact compared to the initial steady-state conditions, but may be more desirable in some situations. It is possible that some variations may be better than others, especially as the network is changing. To minimize the effect of network changes on the quality of the equivalent, one should try to minimize the *use* of external system variables (especially currents and voltages, but possibly line parameters as well) in the direct computation of the equivalent. Naturally, when the equivalent is computed assuming fixed values for these quantities, if these quantities change, the equivalent is no longer exact. One solution is simply to allow the equivalent to be updated as the quantities it depends upon change. Updating the equivalent model as the system changes should not be viewed as a disadvantage. Conversely, the ability to perform effective updates to

the equivalent as needed is desirable to maintain the original degree of accuracy. A key motivation for this work is the ability to perform online transient stability analysis, and thus to facilitate the online updating of groups and of the reduced model. If online transient stability analysis is to be developed and its results are to be trusted, it is imperative that the computations occur quickly and the results have a high degree of accuracy for the scenarios of interest.

In summary, the tools and techniques in Section 4.3 allow us to examine the impact of groupings and of aggregations on the results, with respect to particular properties. It is worth noting that in some cases, the same structural groups tend to arise for different properties. For example, both coherency groups found as described above and LMP groups (as discussed in Chapter 3), indicate the same generators often appearing together in clusters. So, the network structure couples certain generators together in a way which is visible through a number of seemingly different properties. This characteristic is certainly worthy of further investigation. A related investigation would be to quantify how much coherency depends on the transmission network versus on the dynamic models.

To effectively address these intriguing challenges requires improved validation capabilities. The examples in Section 4.3 highlight some of the validation needs. First of all, how should the accuracy of an equivalent be defined and quantified? The need for well-defined tangible metrics is apparent from Section 4.3. In particular, since we are interested in the behavior of the equivalent as the system is changing, metrics are needed concerning how best to compare transient stability response. Transient stability response can be compared by gauging the presence or absence of certain modes, so oscillation monitoring is useful in this context (discussed in Chapter 6). Additionally, there is the question of which events are important to study for particular applications. There are many transient stability events to potentially compare, so we need a transient-contingency-event-selection methodology to help identify which events are important to study in a particular situation. Evidently some events occur which have greater interactions with the aggregated part of the system, and one must consider how to appropriately understand and account for that interaction.

To help address validation, a script-based validation tool was created for this work which interfaces with PowerWorld Simulator via SimAuto and MATLAB. This is a resource to automatically create transient contingencies, run simulations, and retrieve the results for further automated, customized analysis. However, what these suitable comparison metrics should be remains an open question. The hope is that with these tools made available, one can study the impact of different groupings and aggregations on system properties, and that it will be possible to formulate suitable validation metrics, especially with respect to comparing transient stability response.

The third application presented in Chapter 4 is an extension of the first application (Section 4.2). Rather than estimate an equivalent model at the PMU buses, in Section 4.4, the parameters of individual transmission lines are estimated from historical data. Chapter 5 further addresses the estimation of line parameters when real data with noise and error is considered.

In summary, the applications presented in this chapter examine the characteristics of the network which are obtained in the form of a model through the use of data mining techniques. The key result is that power system models can be assessed and improved with the aid of measurement data and data mining techniques. Network modeling and validation is a significant area of intersection between power systems analysis and data mining, and a great deal of improvement can be made in this area.

5. Analysis of Time Series Data

This chapter presents the use of measurement data to find patterns and can be considered the real-data parallel to the sensitivity work in Chapter 3 and the model work in Chapter 4. For the Chapter 4 estimation applications, the statistical properties of the data used are important to the quality of the estimate [23]. Statistical learning methods play a significant role in data mining, and statistical information can be obtained without any knowledge of the system. In this chapter, we examine some of the challenges associated with using raw data to extract relationships. This undertaking can be considerably difficult. In dealing with measurement data, the basic goal is to extract "trends" which should corroborate what is expected and supported by the model. Correlation analysis and visualization of measurement data support this goal. A more practical challenge is dealing appropriately with anomalies in the measurement data that would not be encountered in the model.

5.1. Data as a Diagnostics Tool

In this section, we show some real-world examples where the properties of collected time-series data are important for estimation as well as diagnostics. In Chapter 4, the basics of line parameter estimation are described and demonstrated for a simulated case. Here, we discuss requirements to adapt the estimation method to work on real-world data, susceptible to noise and errors. The proposed methodology to estimate line parameters using real data is presented in [127]. In this section, some key aspects of dealing with real-world data according to the method in [127] are outlined. We also present examples where the data does not exhibit the expected behavior and provide some analysis to determine potential causes.

5.1.1. Dataset Selection and Screening

Before data can be used for analysis, one must select and retrieve the data of interest. For historical data, this data selection can also be designed to accomplish data "screening" functions such as value filtering and initial bad data rejection. Then, for a given set of data, more analysis is applied to detect and reject possibly wrong data based on analysis of a fitted model. The procedure involves several steps which are implemented as described in this section.

A dataset selection procedure, such as Algorithm 1 of [127], is responsible for selecting and returning a set of data on which the analysis is to be performed. For a specified line, the sample dataset for the line parameter estimation application consists of measurements of P, Q, and V at both line ends. This function also provides initial screening and filtering of the raw SCADA data. In the line parameter application, all data is initially screened simply by removing data points which have all values of zero. Additional screening may also be performed. A *data point* or a *row* refers to one set of SCADA measurements from one point in time. Value filters, when applied, allow the procedure to return a dataset for particular conditions, such as $a<x<b$, where x is any of the measured or computed variables. In summary, the dataset selection stage returns a random consecutive data chunk of size *chunkSize* which has undergone the specified initial screening and value filtering.

After an initial fit to the dataset is obtained from least squares, outliers which are outside of a predetermined threshold from the fit are discarded. Outlier filtering can be implemented in a number of ways; one suggested approach is given in [127]. Screening the data is a critical stage of the estimation process, as the dataset which is chosen ultimately impacts the quality of the final estimate.

5.1.2. Simulated Data with Error

While Chapter 4 considers the ideal situation, here we examine the impact of data which is no longer "error-free." In this section, the data is still simulated, so the errors are known. In particular, the five error-free operating conditions in Table 18 are taken, and 200 perturbations are made of each point. This reflects the fact that during each quasi-steady-state operating point, the SCADA values do not change much but may change slightly. Thus, from the five exact operating points, we create 200*5 = 1000 hypothetical measurement data points. Each measurement is per-unitized and is randomized by a normal probability distribution of mean zero and standard deviation 2 MW, 2 MVAr, and 0.02 per unit voltage. These 1000 points are used to compute the estimate. The estimation results from the new dataset are shown in Figure 31 and Figure 32, where Figure 32 shows the real power estimate for the same dataset when the static error is not estimated.

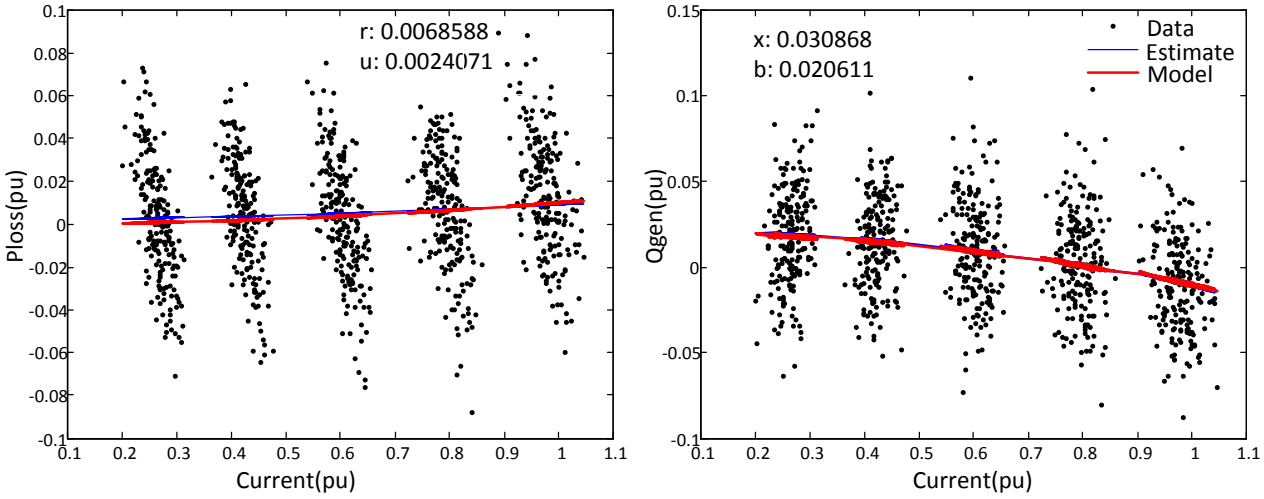

Figure 31. 1000-point simulated dataset with noise, plots of P_{loss} (left) and Q_{gen} (right)

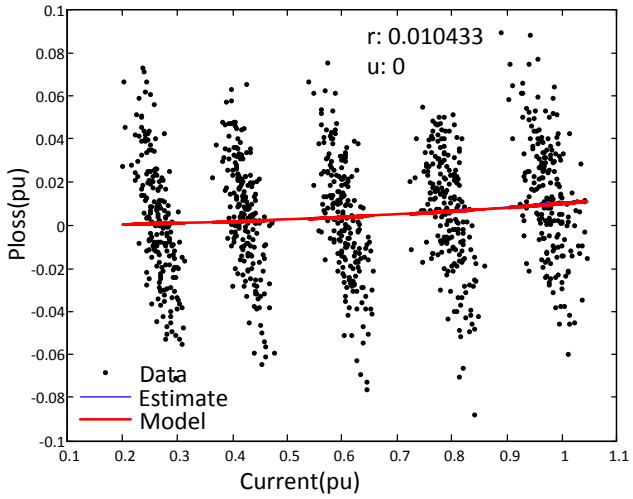

Figure 32. 1000-point simulated dataset with noise, plot of P_{loss}, without estimating static error term

A comparison of the resistance estimates in Figure 31 and Figure 32 indicates that the accuracy of the *r* estimate is considerably improved when the static error *u* is not estimated.

It is interesting to note that the five operating points result in five distinct clusters of points on the graphs. In real data, the presence of these bands is also observed, especially when a relatively small set of points is examined. Thus, multiple noisy measurements of the same operating point appear to cause this characteristic.

Estimates using an entirely different set of operating points are shown in Figure 33 and Figure 34. Here, 60 randomly chosen operating points are simulated, and for each operating point, 500 noisy data points are generated. This results in a dataset with 60*500=30,000 total data points. In Figure 34 as well as Figure 32, Gaussian noise has been added to the measurements, and there appears to be no benefit in including a static error term in the estimation procedure for *r*.

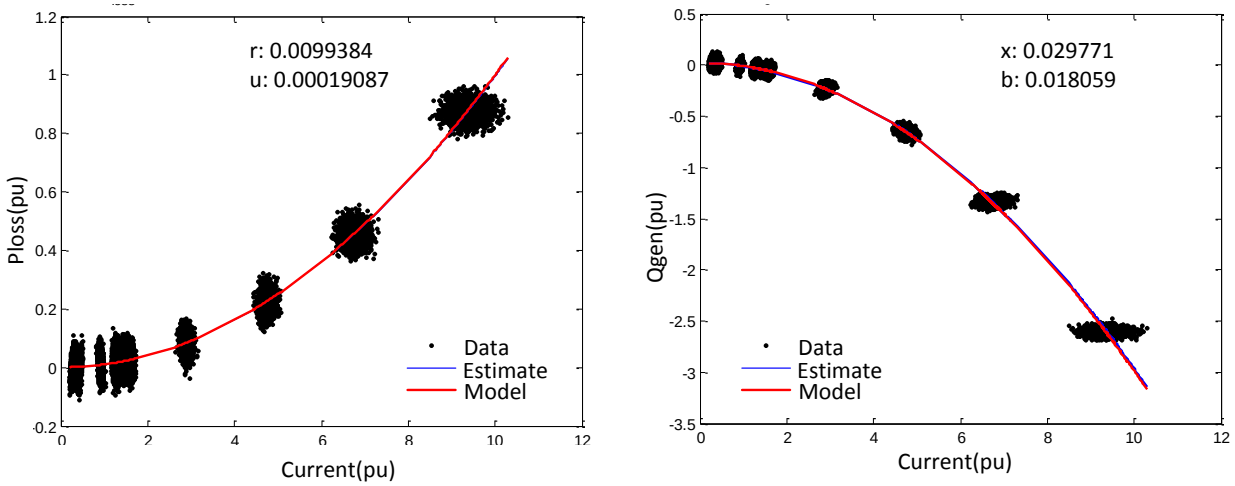

Figure 33. 30,000-point simulated dataset with noise, P_{loss} (left) and Q_{gen} (right)

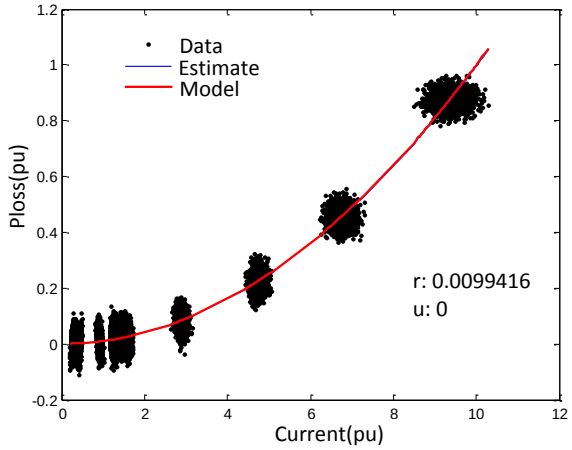

Figure 34. 30,000-point simulated dataset with noise, plot of P_{loss}, without estimating static error term

As may be evident when comparing the results above, some sets of operating points facilitate better estimates than others. The issue is that when simulating data, it is challenging to

reproduce data with a realistic distribution. The question arises of how one should perturb the system to obtain representative operating points since there are infinite possible ways to change the system and obtain a new operating point. Thus, rather than speculate on what representative operating points exist that should be studied, the most value is obtained from testing the procedure on real data, examined in the next section.

5.1.3. Application to Real SCADA Data

The estimation procedure is applied in this section to historical SCADA data from a real North American power system. Identifiers of buses and lines in the real system have been made anonymous. From a practical point of view, it is valuable to be able to present results for actual SCADA data. Estimation on artificially constructed data is merely a special/trivial case. Parameter estimation using real world data can be challenging, as evident by the results and discussions presented in this section. Real world data rarely follows a known probability distribution, much less a Gaussian.

A key feature of this application is that real data provided by a North American utility is used. Several months of historical SCADA data, recorded every five minutes, is available. Large volumes of historical SCADA data collected over time are used to estimate the values of transmission line model parameters. This real data often contains significant errors which do not have known probability distributions. Meters used for measuring data are a common source of error.

The SCADA data, collected over several months, is available for the five transmission lines. The resulting system is presented in Figure 35. The model values of the line parameters for these lines are tabulated in Table 20. Each line identifier $\alpha=(i,j)$ denotes a line α between bus i and bus j, according to Figure 35. This model data has been made available to us to serve as a comparison.

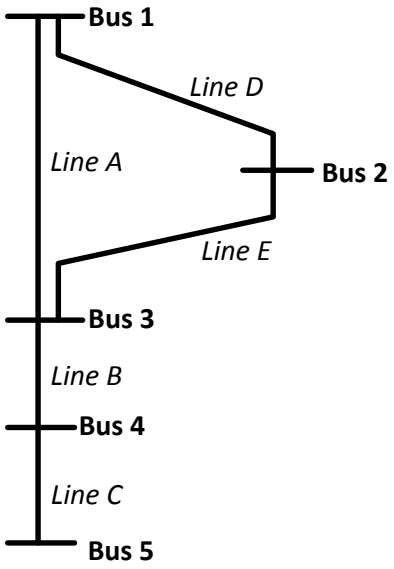

Figure 35. Five lines with available data

Table 20. Existing model values for lines

Line	r (pu)	x (pu)	b (pu)
A=(1, 3)	0.00056	0.01054	0.9079
B=(3, 4)	0.00052	0.00999	0.75194
C=(4, 5)	0.00025	0.00445	0.32042
D=(1, 2)	0.00013	0.00251	0.21230
E=(2, 3)	0.00045	0.00872	0.6724

Comparisons of the raw data to the predicted data values based on both the estimated parameters and the available model parameters are shown for each line in Figures 37-46. These initial model-vs.-data comparisons provide indication of potential problems. Results for all five lines are presented.

Each time an estimation is performed, an initial dataset with a size of 30,000 data points is selected. Initially, the first three lines are considered. Lines *D* and *E* are discussed later. Figures 36, 38, and 40 plot the real and reactive power losses vs. current magnitude for each of the first three lines. The red and blue lines on these plots show respectively the expected data points when using the estimated parameter values and the model parameter values. Figures 37, 39, and 41 show the distribution of residuals about the initial estimate of parameter values. These plots indicate that the distribution of real data is not Gaussian. However, even with such distributions, the use of large datasets available for lines *A*, *B*, and *C* leads to reasonable and consistent

parameter estimates. The estimation for the real and reactive line parameters (*r*, *u*, *x*, and *b*) is performed, as outlined in Chapter 4 and described in detail in [127]. Estimates for the three lines are given in Table 21. Each estimate is based on 30,000 data points. The estimates of the line parameters based on the data above are given in Table 21. As can be seen from the plots, the estimated parameters are close to the available model parameters.

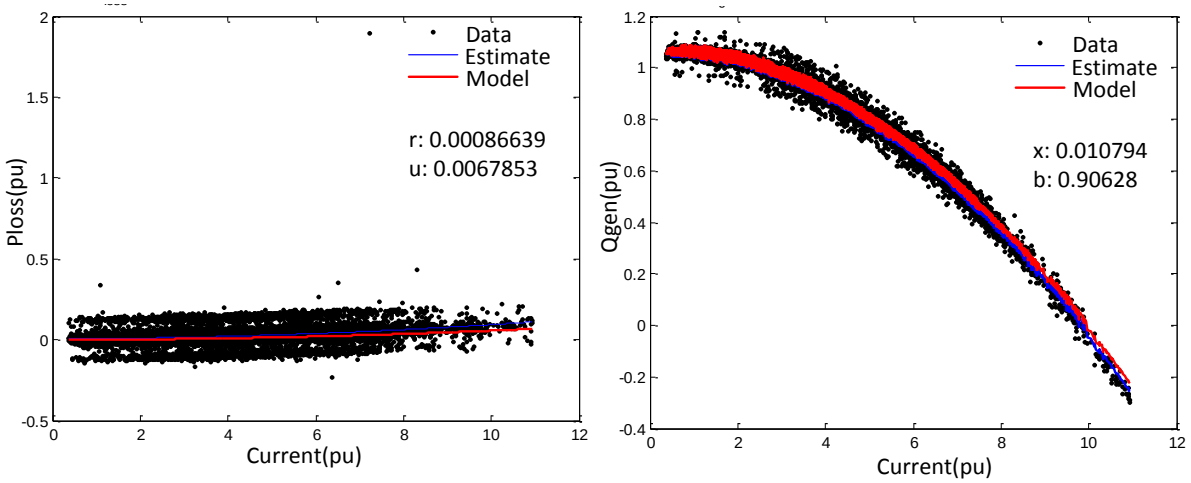

Figure 36. Line *A* -- P_{loss} vs. I (left) and Q_{gen} vs. I (right)

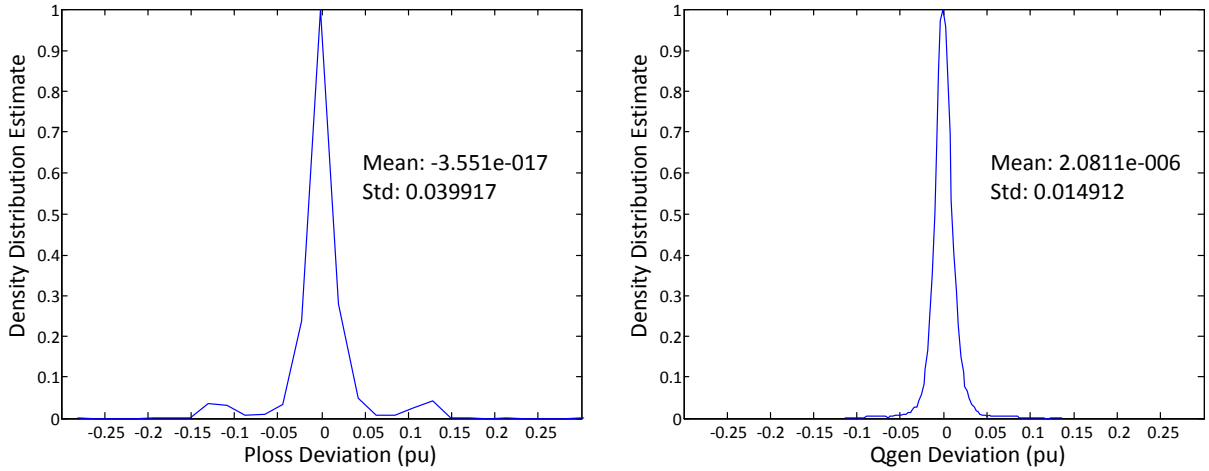

Figure 37. Line *A* -- Distributions of P_{loss} residual (left) and Q_{gen} residual (right)

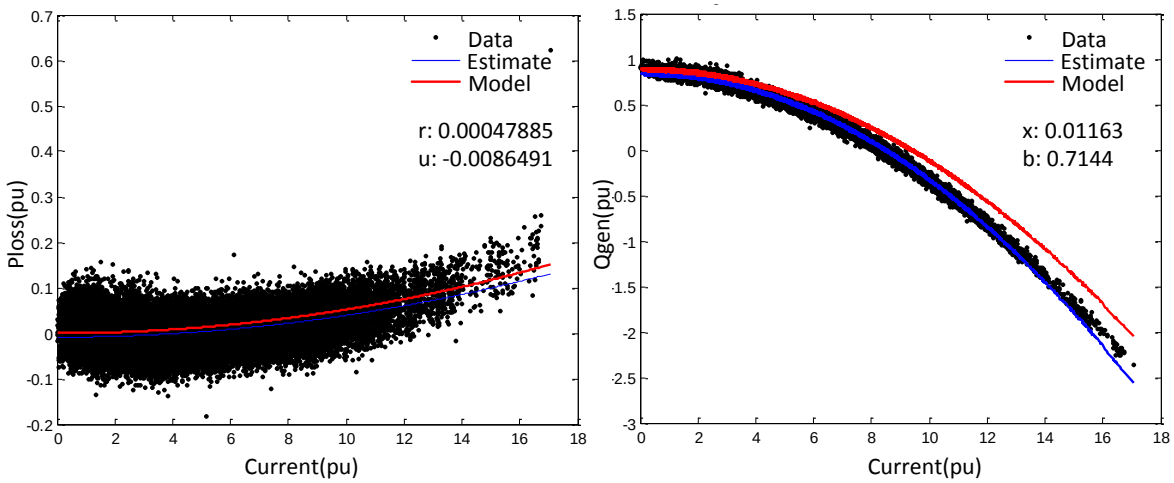

Figure 38. Line *B* -- P_{loss} vs. I (left) and Q_{gen} vs. I (right)

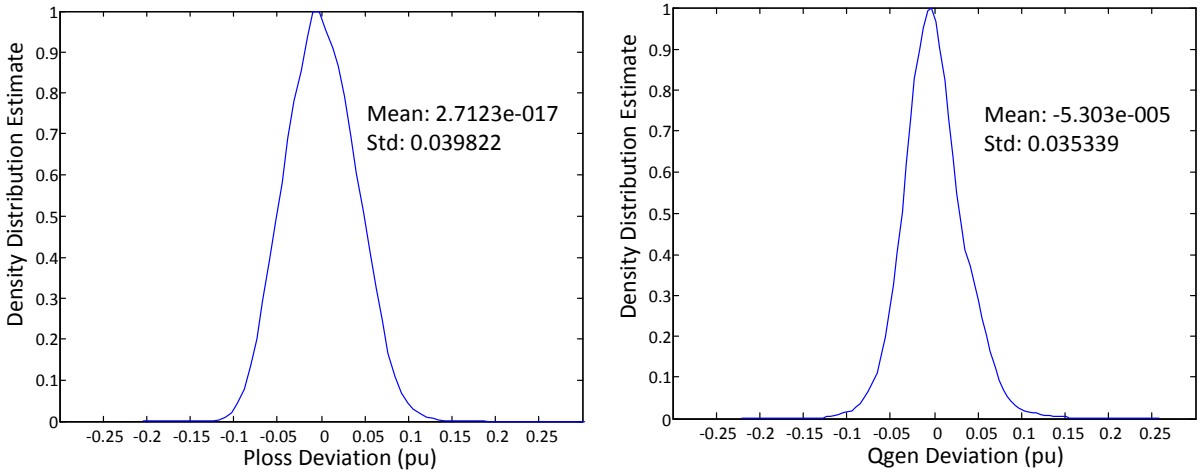

Figure 39. Line *B* -- Distributions of P_{loss} residual (left) and Q_{gen} residual (right)

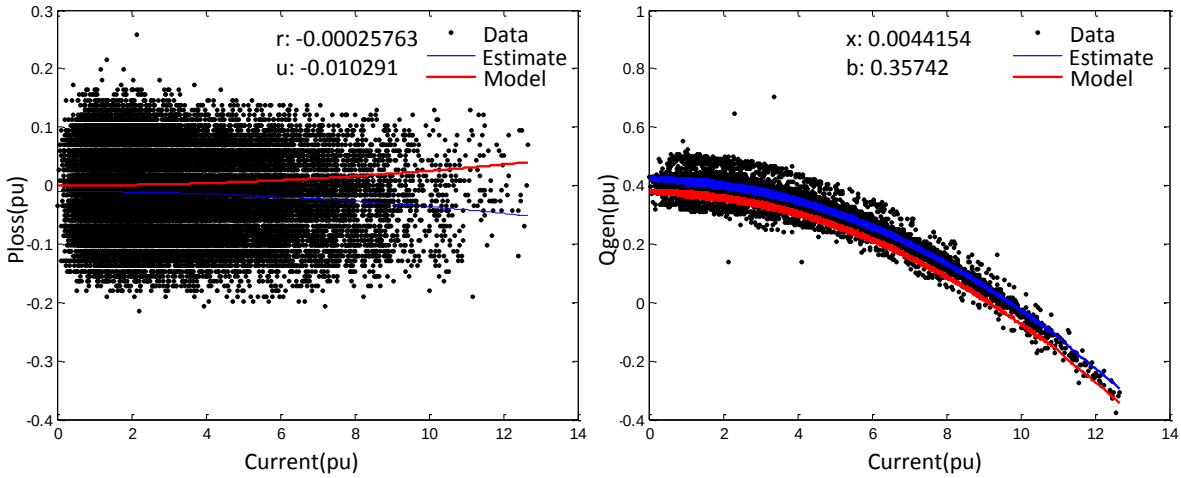

Figure 40. Line *C* -- P_{loss} vs. I (left) and Q_{gen} vs. I (right)

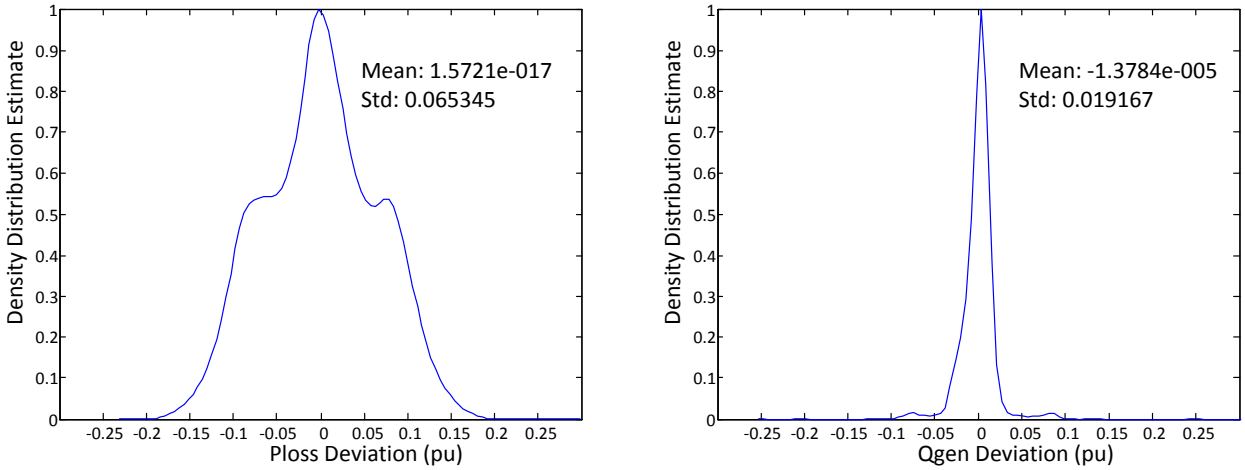

Figure 41. Line *C* -- Distributions of P_{loss} residual (left) and Q_{gen} residual (right)

Table 21. Estimates from 30,000 data points

Line		u (pu)	r (pu)	x (pu)	b (pu)
A=(1, 3)	Estimate	0.0067853	0.00086639	0.010794	0.90628
	Model	-	*0.00056*	*0.01054*	*0.9079*
B=(3, 4)	Estimate	-0.0086491	0.00047885	0.01163	0.7144
	Model	-	*0.00052*	*0.00999*	*0.75194*
C=(4, 5)	Estimate	-0.010291	-0.00025763	0.0044154	0.35742
	Model	-	*0.00025*	*0.00445*	*0.32042*

The estimated resistance for Line *A* is slightly higher than the available model parameter. The estimated value of resistance on Line *C* is negative; this is discussed in the next section.

5.1.4. Identifying Data Problems

Consider the Line *C* estimates from Table 21. The Line *C* estimated resistance is negative when the algorithms are applied on the available raw SCADA data. However, the magnitude of the estimated resistance is close to the model resistance available for this line, which may be an indication that a significant number of data points have sign errors. It is evident that the reactive part of the estimation does not have the same data problems. The suspected sign errors are corrected in the data screening stage, and a corrected dataset is obtained. Figure 42 shows the resulting P_{loss} vs. I plot. The new estimate of the resistance, r = 0.00030502 pu, is considerably improved.

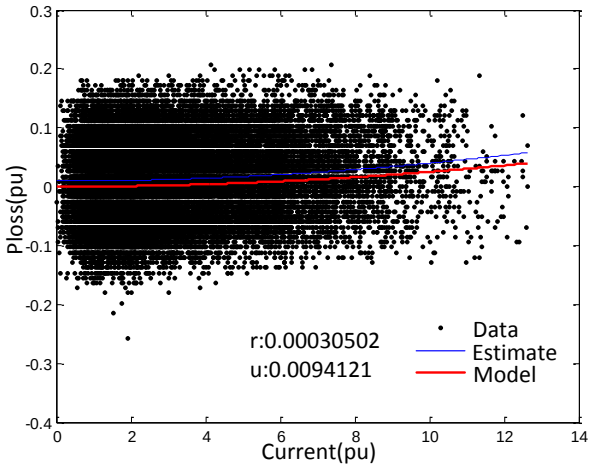

Figure 42. Line *C*, corrected data -- P_{loss} vs. I

The data for reactance and shunt estimation for the three lines has a more normal distribution compared to the data for resistance estimation, as evident from the distribution plots. The real power data contains significant noise with respect to the small P_{loss} values and the small value of the parameter r being estimated.

Two additional lines (*D* and *E*) are considered in Figures 43-46, and this data exhibits even more severe problems. For these two lines, a suitable model that matches the provided SCADA data cannot be identified. This is evident from the plots of the data for lines *D* and *E*, again each consisting of 30,000 data points. The parameter estimates for these lines are given in Table 21.

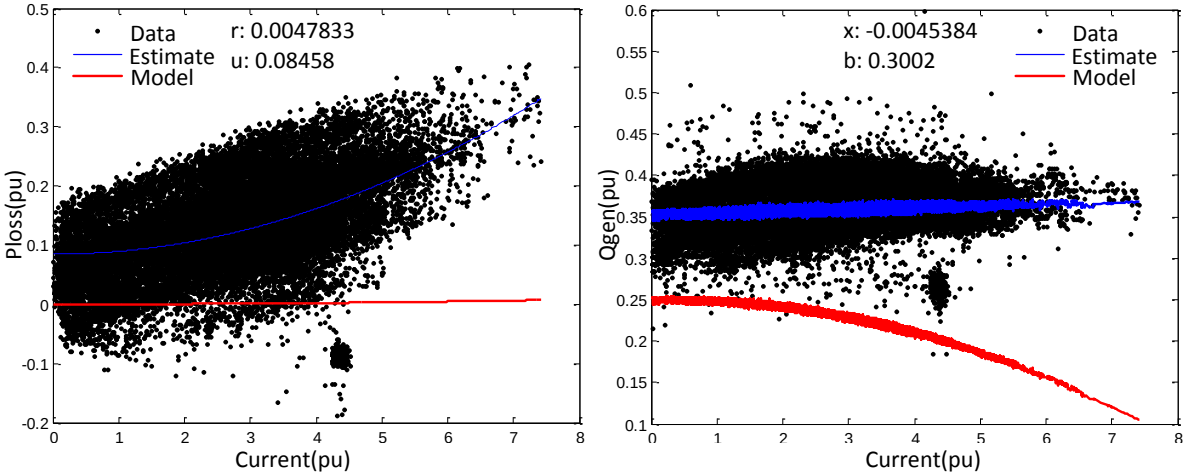

Figure 43. Line *D* - P_{loss} vs. I (left) and Q_{gen} vs. I (right)

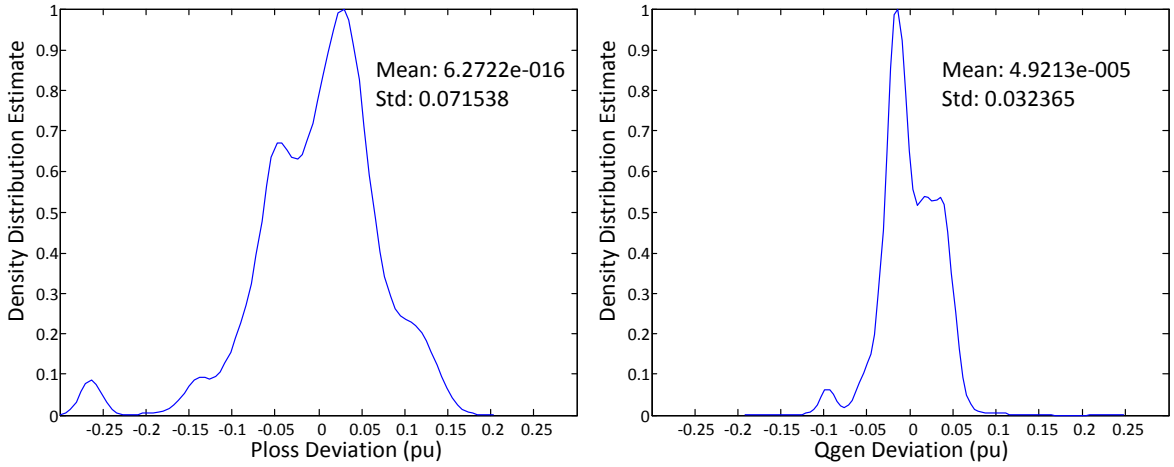

Figure 44. Line *D* - Distributions of P_{loss} residual (left) and Q_{gen} residual (right)

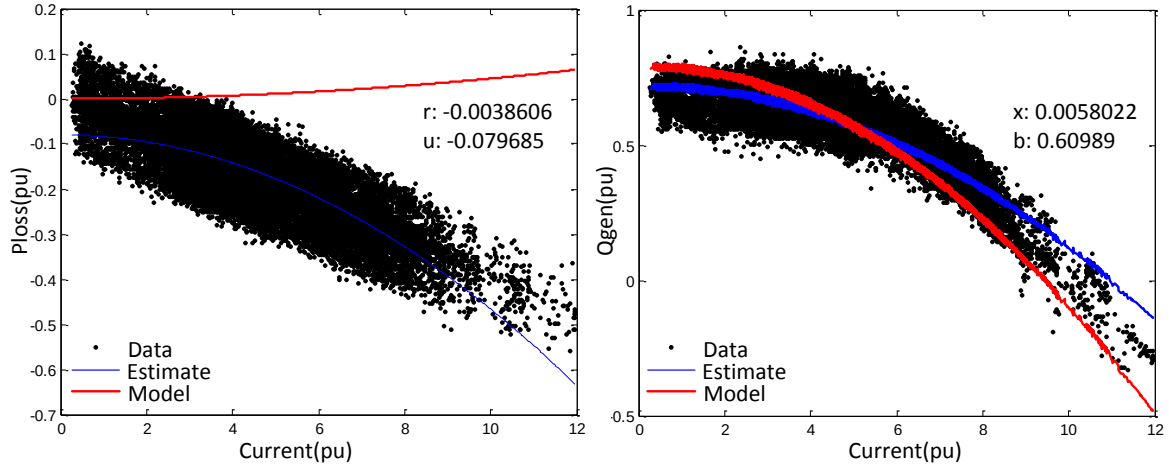

Figure 45. Line *E* - P_{loss} vs. I (left) and Q_{gen} vs. I (right)

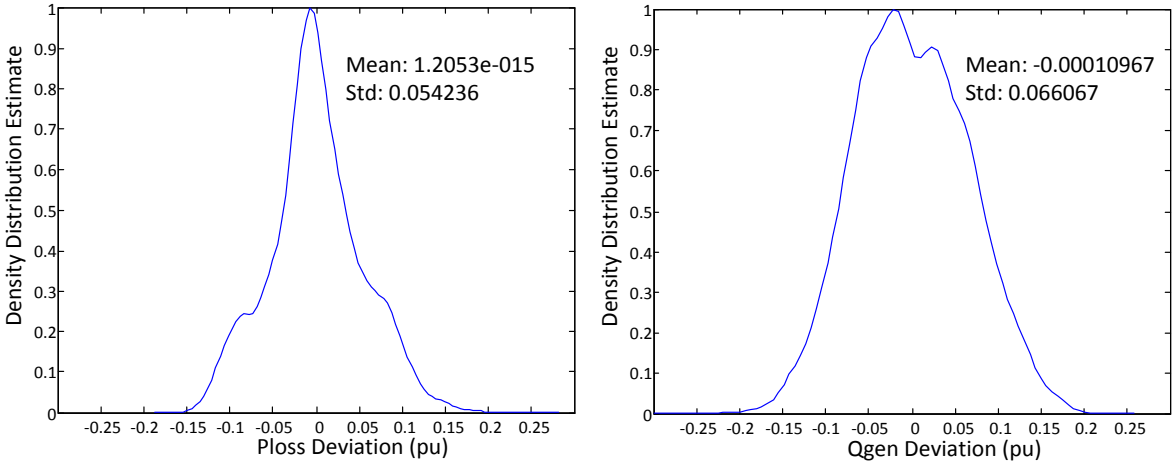

Figure 46. Line *E* - Distributions of P_{loss} residual (left) and Q_{gen} residual (right)

Table 22. Initial estimates from 30,000 data points

Line		u (pu)	r (pu)	x (pu)	b (pu)
D=(1, 2)	Estimated	0.08458	0.0047833	-0.00045384	0.3002
	Model	-	*0.00013*	*0.00251*	*0.21230*
E=(2, 3)	Estimated	-0.07685	-0.00386061	0.00580221	0.60989
	Model	-	*0.00045*	*0.00872*	*0.6724*

The distributions illustrate how well the linear estimate approximates the data but do not directly compare the model with the data. Hence, one interesting observation from the real power residual distribution of Line *E* is that the estimate computed actually fits the data reasonably well, and we see that while the distributions are not Gaussian, they are not worse than the distributions of other lines for which we were able to obtain a consistent and acceptable estimate. For reasons which are unclear, this linear approximation actually differs substantially from what is expected based on the model. Comparatively, for Line *D* we see that this is not the case, that the distributions do not appear Gaussian, so the linear model is not even a good fit of the data.

For Line *D*, the estimated *r* is much higher than the one given by the model, and the relationship between P_{loss} and *I* appears to be linear. For Line *E*, this characteristic is even more apparent; P_{loss} does not seem to depend on I^2 but on I , and the relationship between the two appears completely linear. To consider what may cause this linearity, note that $P = \mathrm{Re}(VI^*)$ would depend linearly on *I* if *V* is roughly constant. The fact that *V* is roughly constant indicates voltage regulation, which may indicate the existence of a generator, since generators typically have voltage regulation. Interestingly, this observation was later independently verified when we were told that there actually is a generator at Bus 2. This situation points to the helpfulness of order statistics, discussed in Section 5.2, which make this even more apparent.

Lines *D* and *E* show that problems seem to exist with the data as well as perhaps the model. Some investigations to track down the source of the data problems are now presented. To help perform these diagnostics, additional information and SCADA data was graciously provided by the utility. This additional information included data for the three generators connected to Bus 2. In the following section, we present some of the correlations we observe between the data of this generator and the model parameter estimates. The existence of this generator was suspected from the data, as mentioned previously, since the data seemed to indicate the presence of voltage regulation. The suspicion was confirmed by the utility, who furthermore encouraged us to verify

the cause of the problematic data by providing us with more information about the generator. Information obtained from the utility consisted of the real and reactive power outputs of each generator connected to that bus.

The additional data source is incredibly valuable since it allows us to start making consistency checks. If we have data which we know should be redundant, and using different combinations of the redundant data does not provide consistent results, then we know one data source is in error. The problem is that it may not be that simple to find. There may be more than one data source in error or they may have errors of a non-obvious form. Potentially the "error" could even be caused by a human altering the data or switching the labels between measurement points, either accidentally or maliciously. Errors which are not random are more difficult to detect and deal with; applications to deal with non-random errors are presented in Chapter 7.

In a way, the problem now is beginning to look more like state estimation and bad data detection. The difference is that here we are using not state estimates which have been smoothed by a state estimator, but rather raw collected data. Another key and useful difference between this and state estimation is that the line parameters are typically fixed. Then, once one knows or gains some appreciable degree of certainty about what the values should be, this can be used to track down the source of the bad data.

Several diagnostic measures were taken to try to determine the cause of the problem. In one approach, an aggregated fictitious line was formed from original lines which were in series. Line *G* represents lines *D* and *E* in series. The model values for Line *G* are shown in Table 23.

Table 23. Existing model values for lines

Line	r (pu)	x (pu)	b (pu)
G=(1, 3)	0.00058	0.01123	0.161355

Combining the two trouble lines, Lines *D* and *E*, into Line *G* yields "better" results (Figures 47 and 48). An interesting point is that there is a generator at Bus 2, yet the losses go high and *positive* which would indicate that it is a *load* and not a *generator*. Thus, something appears to be wrong with the data coming from the generator, since it is negative of what it should be.

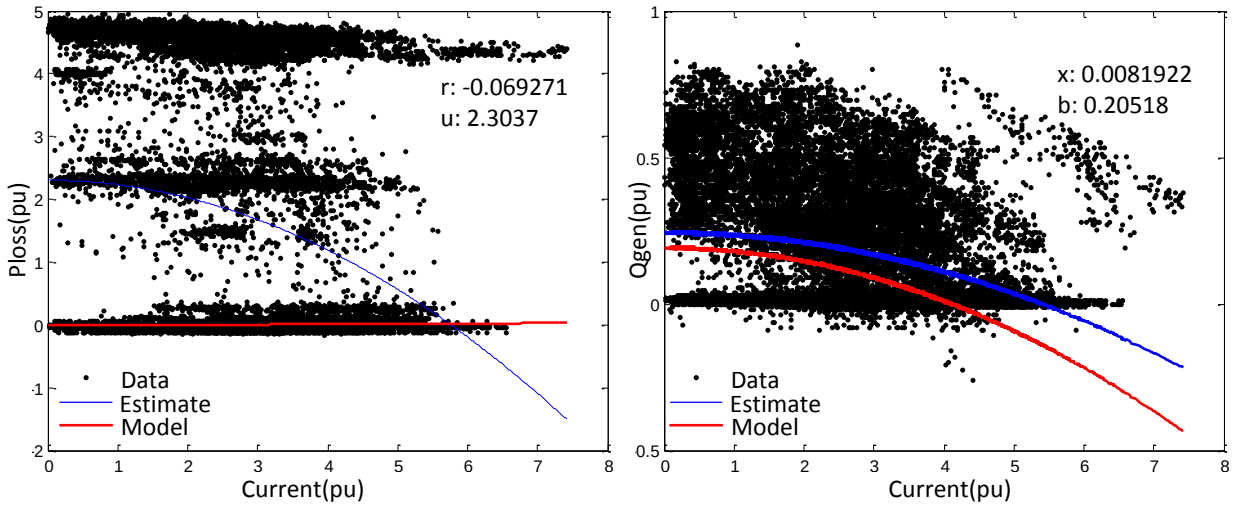

Figure 47. Fictitious Line *G* - P_{loss} vs. I (left) and Q_{gen} vs. I (right)

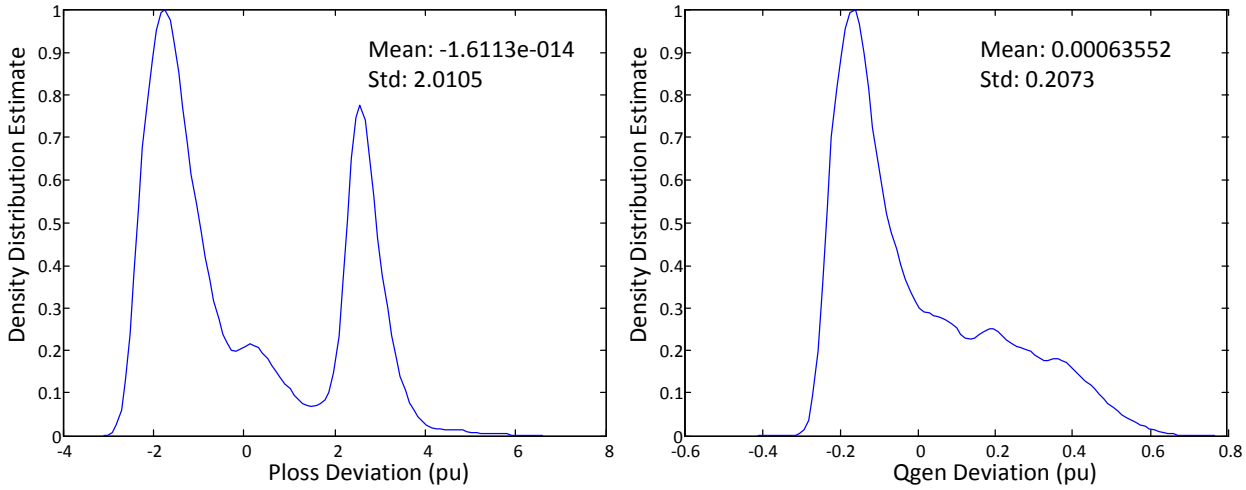

Figure 48. Fictitious Line *G* - distributions of P_{loss} residual (left) and Q_{gen} residual (right)

The estimate for the values of Line *G*, based on the data above, are given in Table 24:

Table 24. Initial estimates from 30,000 data points

Line		u (pu)	r (pu)	x (pu)	b (pu)
G=(1,3)	Estimated	2.3037	-0.069271	0.0081922	0.20518
	Model	-	0.00058	0.01123	0.161355

The poor approximation is caused by the high P_{loss} values, and this makes sense because when the P_{loss} value is high, there is significant real power being injected or withdrawn from somewhere along the fictitious line (this is due to the presence of generators at Bus 2, as discussed). However, points where P_{loss} is close to zero should allow us to get reasonably good approximations of the line parameters. If we filter the above data in an attempt to remove data points with the generator on, by including only P_{Total} = [0.01, 0.2], the estimates in Table 25 are obtained.

Table 25. Initial estimates from 30,000 data points

Line		u (pu)	r (pu)
G=(1,3)	Estimated	0.10586	0.00017121
	Model	-	*0.00058*

The following figures examine the data on Line *D*, one of the lines adjacent to the generator. Value filtering is applied based on the generator data. Let Q_1-Q_3 be the reactive outputs of the 3 generators. Figures 49 and 50 show the correlations between generator outputs and voltage magnitude. As we would expect, the correlation becomes higher and positive as the amount of real power generation is increased. In comparing these correlation plots with the ordered *I* plots, as the data set is filtered to include points where the generator is operating at high outputs, the estimate begins to take a form which is a closer match to the model.

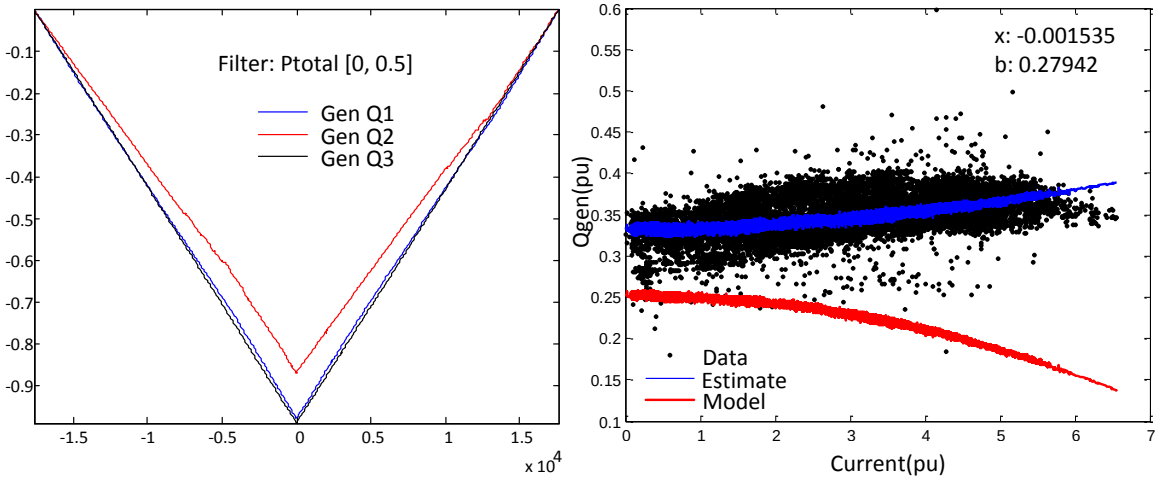

Figure 49. Line *D*, V correlations with Q_{gen}, filtered with $P_{Total} \in [0,0.5]$

The output range where the estimate is most similar to the model is $P_{Total} \in [4.8, 5.2]$, yet Figure 50 indicates that the raw data is not a good match to either the estimate or the model. In Figure 51, all time points with negative Q_{Total} values have been filtered out (points with $Q_{Total} > 0.1$ are retained), since generators cannot normally operate at a leading power factor. The data matches the estimate reasonably well. The model, however, does not match either.

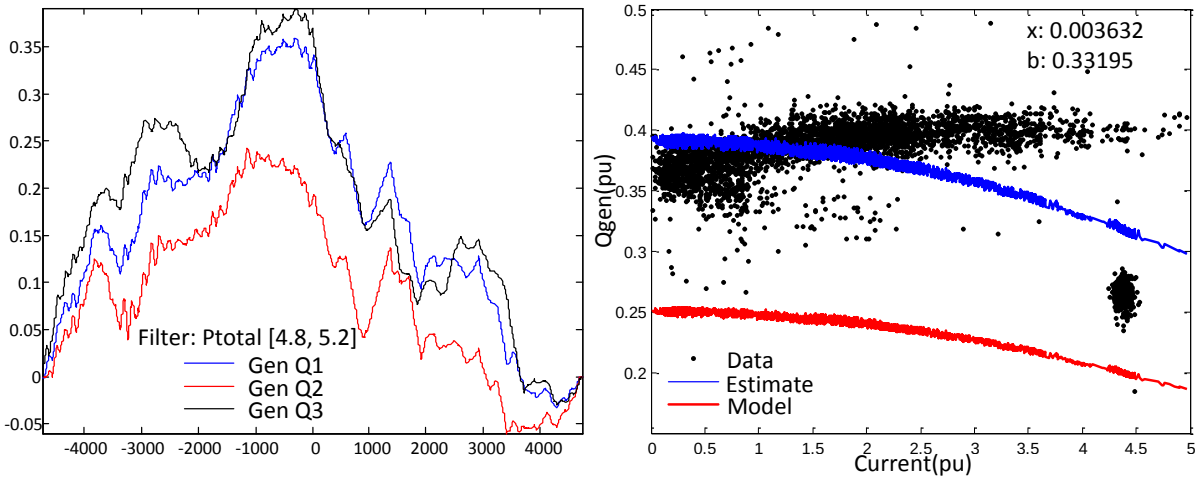

Figure 50. Line *D*, V correlations with Q_{gen}, filtered with $P_{Total} \in [4.8, 5.2]$

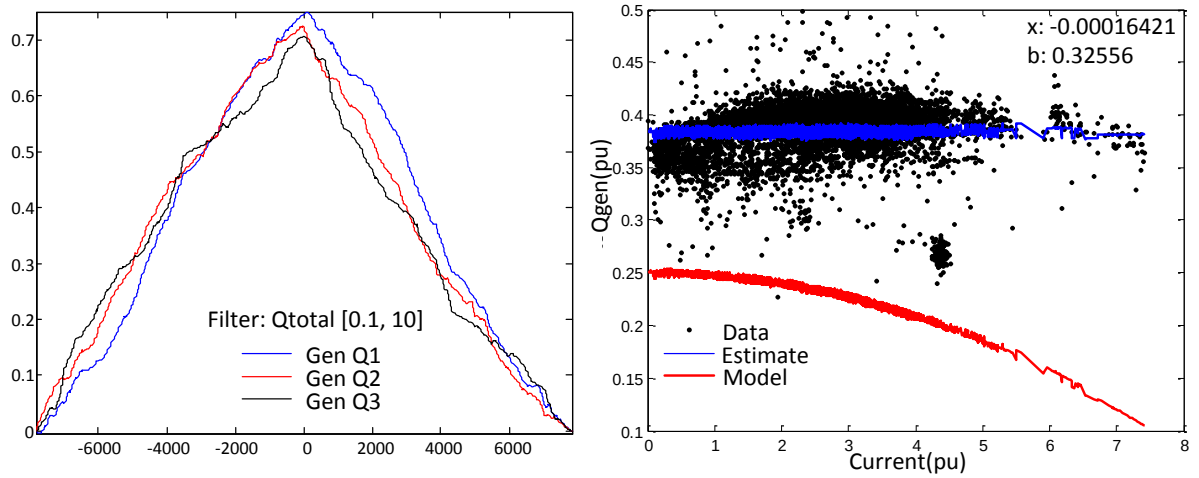

Figure 51. Line *D*, V correlations with Q_{gen}, filtered with $Q_{Total} > 0.1$

5.2. Statistical Properties of Data

Chapters 6 and 7 focus on event and bad data detection. To understand what constitutes an anomaly or an event based on the data, one must first understand what is normal for the data.

This requires the ability to identify normalcy and thus to recognize trends. Several important statistical properties of data are examined in this section, and more background on the subject is given in Chapter 2. In particular, to investigate relationships between variables, the concepts of statistical independence and correlation are useful, and Section 5.2.1 reviews these in the context of power system data. Properties of error are explored in Section 5.2.2 The ability to improve estimates by taking more samples, utilized extensively in Chapter 4, is examined more fundamentally in Section 5.2.5.

5.2.1. Statistical Independence, Covariance, and Correlation

Two measurements are said to be statistically independent if the error of one of the measurements has no effect upon the error of the other measurement. There is a difference between statistical dependence and functional dependence. Functional dependence concerns the relationships expected between values. Statistical independence does not imply the lack of existence of a functional relationship [23]. Two random variables x and y with deviations x-$E(x)$ and y-E(y) are statistically independent if the values of these deviations do not affect each other in any way. Hence, statistical independence concerns the independence of the deviations of random variables. For independent random variables, the joint frequency distribution (the probability that both x is x' and y is y') is equal to the product of the two distributions.

Covariance and correlation are often used to describe the relationship between random variables. Covariance is the expected value of the product of the two deviations,

$$Cov(x,y) = E\left[\left(x - E(x)\right)\left(y - E(y)\right)\right] \tag{113}$$

where the covariance of a random variable with itself is simply equal to the variance.

A *sample estimate of the covariance* may be calculated from data. This is done by calculating the mean of the x and y sample values, taking the product of the deviations of the sample values from the means, i.e. $(x-\bar{x})\cdot(y-\bar{y})$, and computing the average over all sample values. When calculating this average, the denominator is *(N-1)*, as is the case when computing a sample estimate of standard deviation. The magnitude of the covariance depends on the association

between *x* and *y*. If *x* and *y* increase together, the covariance is positive. If one increases while the other decreases, it is negative. However, the difficulty with relying on covariance as a measure of association is that if *x* and *y* have different standard deviations, the covariance calculations will be skewed. That is, if all of the *y* deviations are multiplied by a factor of *m*, the covariance will also increase by the same factor.

Hence, scaling is achieved by dividing the covariance by the product of the two standard deviations to provide standardization to this metric. This scaled quantity is the *correlation coefficient*:

$$\rho(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} \tag{114}$$

A sample estimate of the correlation coefficient can also be calculated from data. Its value always lies between -1 and +1. Note that the coupling index (CI) in Chapter 3 is also essentially the correlation coefficient. If two random variables *x* and *y* are statistically independent, then their correlation coefficient is zero. The reverse is not generally true.

5.2.2. Propagation of Error

One major reason for studying the statistical properties of data is to have a better understanding of its error. It is often convenient to perform theoretical analysis without assuming the presence of errors. However, if error and data properties are ignored, major problems can arise in the implementation of new ideas as they move to use real data. Data with error presents challenges to the applications which rely upon this data and cannot be ignored.

When studying data error, measurements can be considered a combination of random variables,

$$L = ax + by + cz + \ldots \tag{115}$$

where x, y, and z are random variables and a, b, and c are constants. Even if one has multiple measurements of the same quantity and then takes the average of those measurements as the result, the calculated average has the form of (115). The mean of L is the sum of the means:

$$E(L) = aE(x) + bE(y) + cE(z) + ... \tag{116}$$

If the variables are statistically independent, then the variance accumulates. This phenomenon is called the *Law of Propagation of Errors*:

$$V(L) = a^2V(x) + b^2V(y) + c^2V(z) + ... \tag{117}$$

In the more general case, for any two random variables [23, pp. 62],

$$V(x+y) = V(x) + V(y) + 2 \cdot Cov(x, y) \tag{118}$$

where x and y may each be a linear combination of other random variables. If x and y are linear combinations of other random variables, the covariance of x and y is a sum of covariance terms corresponding to the other random variables. The fact that x and y may not be statistically independent is accounted for by the covariance term, which thus represents the degree of interdependence. In the case where x and y are statistically independent, the variances simply add, as above.

In summary, when random variables are statistically independent, the error in a linear combination of the variables is larger than for any one of the variables. If the deviations of the random variables from their means are correlated, as in the general case (118), the correlation may either increase the error or decrease it. For example, if one considers a time measured by a stopwatch, $d = y\text{-}x$, where y is the end time and x is the start time, there may be a constant delay for x and y, corresponding to a fixed time it takes for the operator to push a button. This positive correlation tends to reduce the variance of d because the error is the same for x and y and adds no uncertainty to the measurement difference $y\text{-}x$.

5.2.3. Measurement Replication and Precision

Using the principles of error propagation, the goal is to improve precision as much as possible. Without requiring the population to have a normal distribution, it is still possible to show that replication of a measurement improves precision. The only requirement is that the N observations must be statistically independent and must belong to the same statistical population or at least to populations which have the same variance. Then, the useful result is that the variance goes down as the number of measurements N goes up [23, p 63].

To illustrate the nature of this improvement in precision, consider the average of observations:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{1}{N} x_1 + \frac{1}{N} x_2 + ... + \frac{1}{N} x_N \tag{119}$$

Since these observations are assumed to be statistically independent, the Law of Propagation of Errors gives the variance of (119):

$$V(\bar{x}) = \left(\frac{1}{N}\right)^2 V(x_1) + \left(\frac{1}{N}\right)^2 V(x_2) + \ldots + \left(\frac{1}{N}\right)^2 V(x_N) \tag{120}$$

Based on the assumption that all of the measurements belong to the same statistical population, their variances are all the same, $V(x_1),\ldots,V(x_N) = V(x)$, and (120) is reduced to the following:

$$V(\bar{x}) = \left(\frac{1}{N}\right)^2 V(x) \cdot N = \frac{V(x)}{N} \tag{121}$$

Thus, the variance of the sample mean tends to decrease as N increases. This provides justification for the intuition that repeated observations should yield a more dependable estimate.

5.2.4. Derived Quantities

As described above, measurements are functions of random variables, and other quantities are often derived based on these measurements. The eventual results of interest are denoted *derived*

quantities, which depend on the measured data [23, pp. 68 Mandel]. In examining derived quantities, *systematic error* or *bias* is important. A systematic error or bias $B(x)$ means that $E(x)$ is not equal to X,

$$B(x) = E(x) - X \tag{122}$$

where X is the true value of the measured quantity x. Therefore, for a linear function of random variables, which has a true value given by U,

$$U = aX + bY + cZ + ... \tag{123}$$

its expected value is the sum of the expected values of x, y, and z:

$$E(u) = aE(x) + bE(y) + cE(z) \tag{124}$$

Then subtracting (123) from (124),

$$E(u) - U = a\left[E(x) - X\right] + b\left[E(y) - Y\right] + c\left[E(z) - Z\right] + ... \tag{125}$$

it may be seen that the bias of the measurement u given by $B(u)$ is the linear combination of the biases:

$$B(u) = aB(x) + bB(y) + cB(z) + ... \tag{126}$$

From (126), if the random variables are free of systematic bias, then a derived quantity which is a linear combination of the random variables will also be free of systematic bias. This is a useful property.

Interestingly, the above property does not hold for nonlinear functions. That is, even if random variables are free of systematic bias, it is possible for a non-linear function of these variables to

contain systematic bias. Thus, in general, a derived quantity may have systematic bias even if the random variables it is based upon do not. While this may seem surprising and even counterintuitive, the concept is readily shown in [23, pp. 69] by considering the simple problem of calculation of the area of a circle based on measurements of its radius. It is shown that the error of the area is systematically high and is proportional to the variance of the radius error. A sketch of this example is the following. Consider the area of a circle,

$$A = \pi R^2 \tag{127}$$

where A is the true area and R is the true radius. Suppose the radius measurement has some error, so the measured value is actually r. Then, the measurement value is the true value plus the error:

$$r = R + e_r \tag{128}$$

The *derived* area based on the measured radius is given by a:

$$a = \pi r^2 \tag{129}$$

The error in the derived area a is defined to be e_a:

$$e_a = a - A \tag{130}$$

Then, substituting (127), (128), and (129) into (130), we obtain a relationship between the error of the area and the true radius R:

$$e_a = \pi(2e_r R + e_r^{\,2}) \quad (\text{for } e_r) \tag{131}$$

If the radius error has a symmetric distribution, then values of both $e_r = e$ and $e_r = -e$ occur,

$$e_{a1} = \pi\left(2eR + e^2\right) \qquad \text{(for } e_r = e\text{)} \tag{132}$$

$$e_{a2} = \pi\left(-2eR + e^2\right) \qquad \text{(for } e_r = -e\text{)} \tag{133}$$

which may then be averaged to determine the average error of the area, $\bar{e}_a$:

$$\bar{e}_a = \frac{1}{2}\left(e_{a1} + e_{a2}\right) = \pi e_r^{\,2} \tag{134}$$

which is a non-negative quantity. The expected value of (134) cannot be zero unless the error e_r is always zero. Otherwise, (134) is strictly an average of positive quantities. Thus, the average error of area *a* is positive, and it follows that the expected value of the derived area *a* will always be greater than the true area *A*.

$$E(a) > A \tag{135}$$

That is, the derived area *a* has a bias which is nonzero,

$$B(a) = E(a) - A > 0 \tag{136}$$

even though the radius measurement is unbiased. The unbiased measurement assumption is only that E(r)=R or E(e_r) = 0. The distribution of e_r need not be symmetrical. The exact value of the bias can be shown to be [23, pp. 71]

$$B(a) = \pi V(e_R) \tag{137}$$

where $V(e_R)$ is equal to the variance of r. The fascinating property demonstrated here is that a purely random error in the radius induces a systematic error in the derived area value. This effect occurs because the derived quantity is a nonlinear function of the measurement, and it is important to note that this situation does *not* arise for linear functions. However, these induced errors are often small and can be neglected.

This is an extremely useful property to keep in mind for power systems since we often compute nonlinear functions from measurements. For example, in the line parameter estimation application, the actual quantities measured by telemetry are three-phase voltage and three-phase current from potential transformers (PTs) and current transformers (CTs) respectively. From those, all other quantities are derived.

5.2.5. Control Charts and Order Statistics

An important question when dealing with measurements is whether they all come from the same homogeneous set. For measurements of a given quantity, it is useful to look at the order in which they are obtained. In doing so, the goal is to determine that the data does not show systematic changes such as a gradual increase, decrease, or cycling. That is, the data should appear random. This is important because, for example, if we are collecting data over a period of time, such as a day, and we notice that there is a gradually increasing trend, then it would be incorrect to say something about the average of these values and to apply that to the whole data set. Thus, examining values in chronological order is important. To determine if data comes from a homogeneous set, the simple idea is to divide the set into a number of subsets and examine whether the variability between subsets is greater than within subsets. Note that these goals are very much analogous to the goals of clustering. The question is how to choose the subsets. When choosing subsets, in some cases, the subsets are obvious, like if the same quantities are recorded using several different approaches. In other cases, there is no obvious expected natural grouping.

One tool for studying data homogeneity is called a *control chart* [23, pp. 82] which was introduced by Walter Shewhart in 1924. Data are divided into subgroups, either rational subgroups if such groups are obvious or by grouping together a number of consecutive values. For each group, the average and the range are calculated. One can then take the average of the ranges of each subgroup to obtain a measure of within-group variability. Plotting the averages and ranges over the subsets gives control charts. The standard deviation of the population can also be estimated.

Control charts for each transmission line of the SCADA dataset from Section 4.4 are shown in Figures 52, 53, and 54. Figure 54 also shows control charts for the fictitious line G from Section 5.1. To obtain these control charts, a chunk of 30,000 data points is retrieved. Then, this chunk is broken into 50 chronological groups. Values for each group are calculated and plotted. For each transmission line, plots show the average of P_{loss} and Q_{gen} for the group as well as the estimates of r, x, and b from the data points in the group. These values are all summarized in Tables 26, 27, and 28 following the figures. According to [23], for a normal distribution, about 2/3 of the values should fall within the "one-sigma limits," and about 95% of the values should fall within "two-sigma limits." Thus, by plotting the chronological averages (or estimates) along with the mean of each group and the standard deviations, we can make some conclusions about the data. In the figures which follow, the two-sigma limits are shown with red dotted lines. We should expect 95% of the values to fall within these limits.

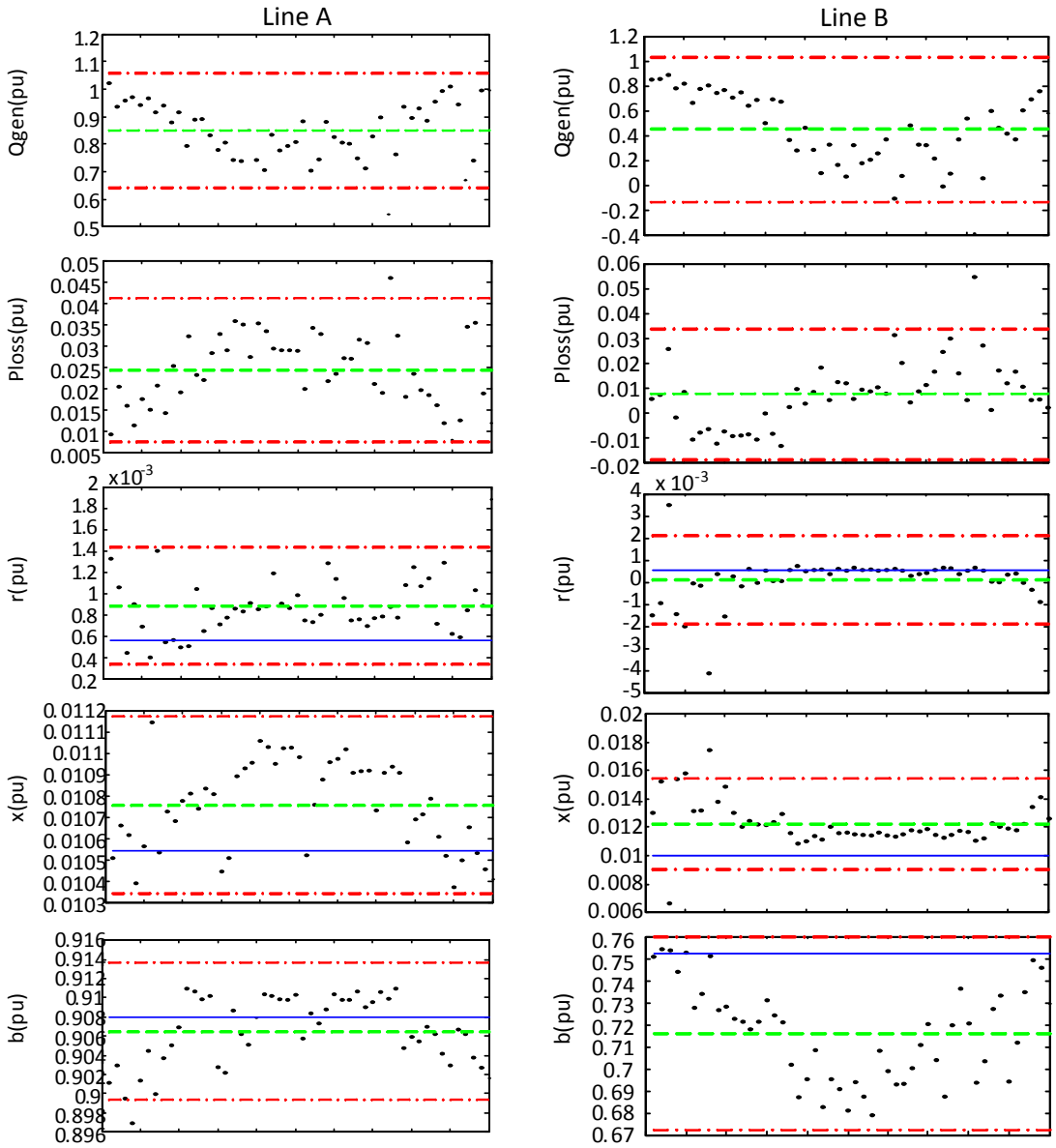

Figure 52. Control charts -- Line *A* (left) and Line *B* (right)

Table 26. Control chart values – Line *A* (left), Line *B* (right)

	Line A			Line B		
	Mean	Std	Model Val	Mean	Std	Model Val
Q_{gen}	0.85069	0.10401	-	0.44866	0.29212	-
P_{loss}	0.02433	0.0084463	-	0.0074734	0.01321	-
r	0.00088435	0.00020751	0.00056	0.00011612	0.0010073	0.00052
x	0.010756	0.00020751	0.01054	0.012259	0.0016	0.00999
b	0.9065	0.0035856	0.9079	0.71587	0.021851	0.75194

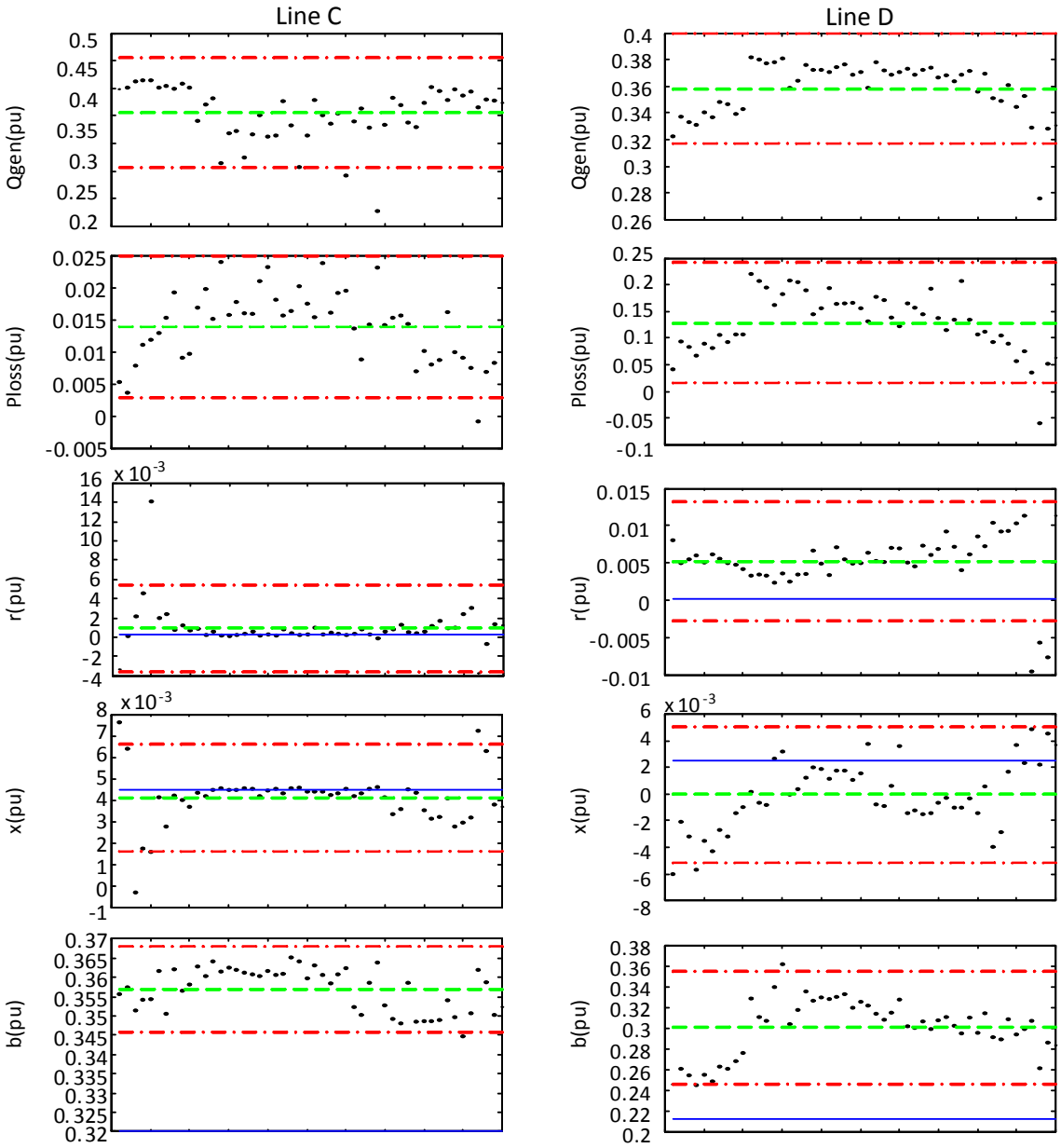

Figure 53. Control charts -- Line *C* (left) and Line *D* (right)

Table 27. Control chart values – Line *C* (left) and Line *D* (right)

	Line C			Line D		
	Mean	Std	Model Val	Mean	Std	Model Val
Q_{gen}	0.355	0.059686	-	0.3581	0.020551	-
P_{loss}	0.013947	0.0054892	-	0.12923	0.056124	-
r	0.00093312	0.0022683	0.00025	0.0051675	0.0039691	0.00013
x	0.0041019	0.0012465	0.00445	-5.6434e-5	0.0025739	0.00251
b	0.35701	0.0055847	0.32042	0.30105	0.027351	0.21230

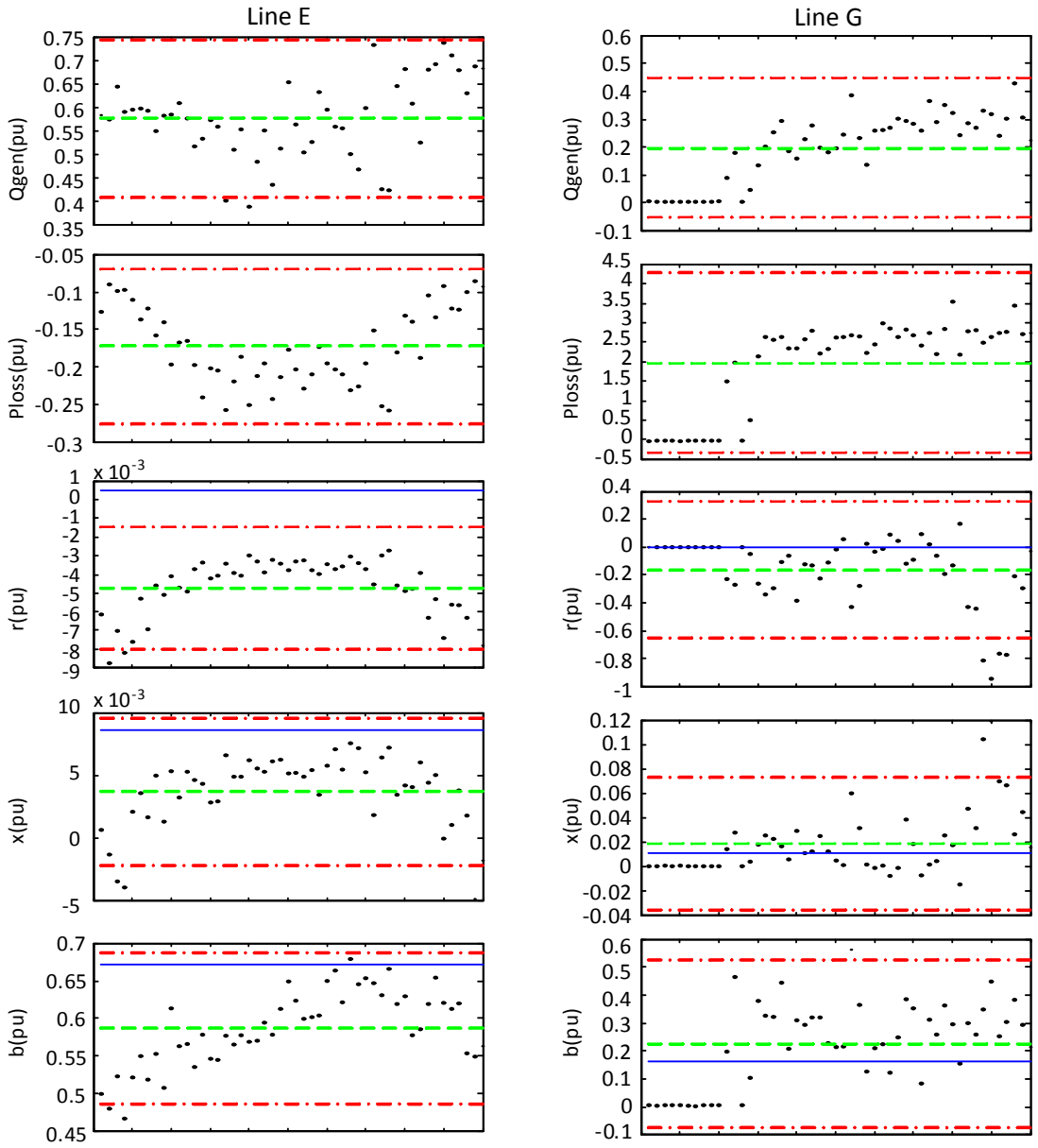

Figure 54. Control charts -- Line *E* (left) and Line *G* (right)

Table 28. Control chart values -- Line *E* (left) and Line *G* (right)

	Line *E*			**Line *G***		
	Mean	Std	Model Val	Mean	Std	Model Val
Q_{gen}	0.57577	0.084501	-	0.19755	0.12484	-
P_{loss}	-0.17276	0.051549	-	1.9703	1.1567	-
r	-0.004759	0.0016349	0.00045	-0.16223	0.24616	0.00058
x	0.0037386	0.0029283	0.00872	0.018352	0.027021	0.01123
b	0.58697	0.050566	0.6724	0.22575	0.14962	0.161355

The objective is to see whether "the values, in the order in which they were obtained, behave like a set of random drawings from a single population [23]." In general, control charts are useful for seeing the randomness or non-randomness of data. Control charts are also useful for revealing *serial correlation*, which is when consecutive values are, on average, less different (or more different) than non-consecutive values. What we see is evidence of serial correlation. Rather than fluctuating in a random matter, the values follow a pattern which seems to indicate some cycling, trending upwards and then downwards. However, most of the fluctuations still occur within the two-sigma lines.

Further investigation into the causes of non-randomness may be warranted. Rather than study the values in chronological order, it is also useful to study the values in order of increasing magnitude. The values arranged in this manner are called *order statistics*. Note that the term "statistic" denotes a function of observations [23, p 85]. That is, a statistic has no unknown parameters. In general, order statistics can be useful for drawing inferences about the nature of a parent population from samples of data. This is done by comparing the ordered data to a known distribution type.

An example is given ([23], pp. 96) which uses order statistics to investigate the consistency and homogeneity of the "measuring process" for 14 different types of paper. This is analogous to our estimation procedure applied to a number of different transmission lines (five, in this example), where the estimated values are different for each line, but the mechanism with which they are measured should have the same characteristics. For the transmission line parameter estimation problem, there are a number of variables for which we can potentially study the order statistics. To form the figures below, the data are first divided into chronological subgroups, as in the control charts above. Then, these values are placed in ascending order. Plots of estimated r, x, and b values are shown in Figure 55, where each plot contains data for all five lines.

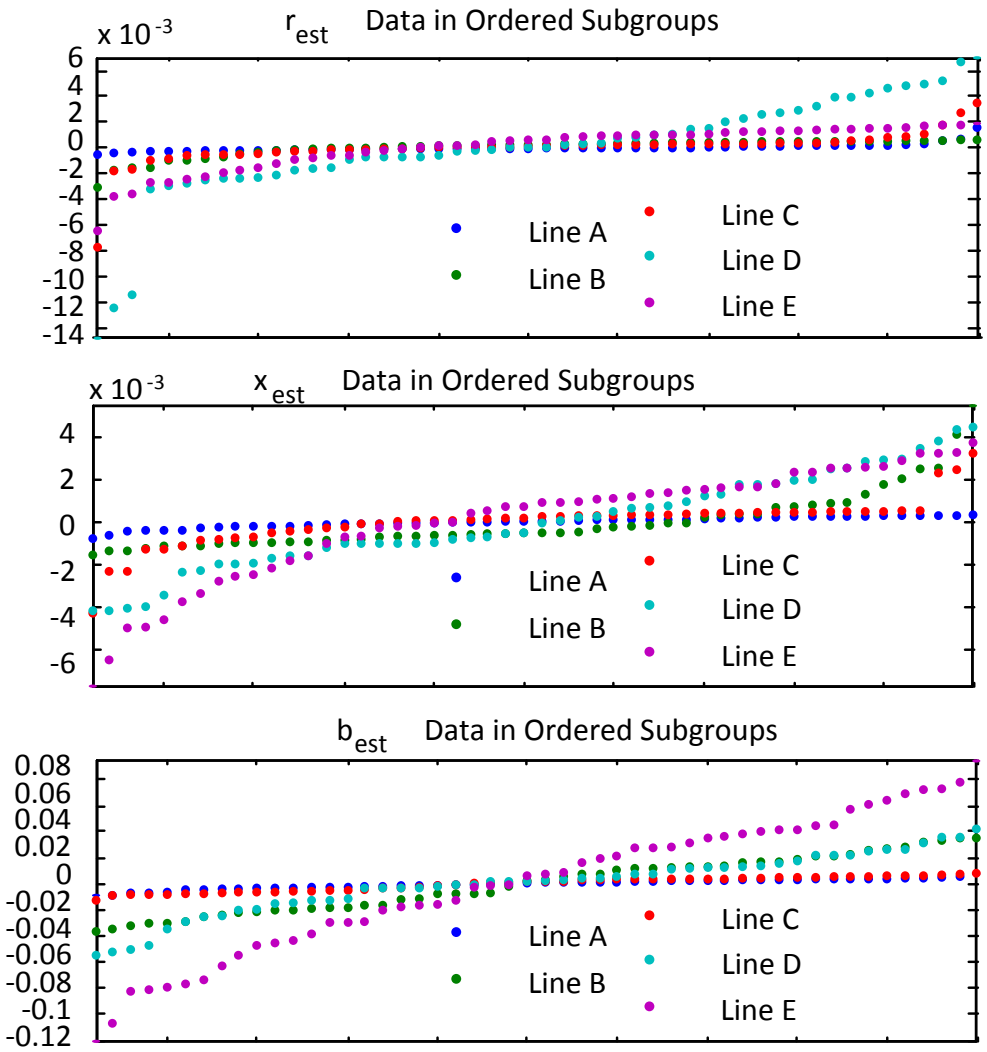

Figure 55. Order statistics of estimated values, with mean removed

Order statistics as in Figure 55 show whether or not the distributions of the populations all have the same shape, differing only by their mean value. In Figure 55, the mean of each series has been removed, which represents the fact that the series are estimating different transmission lines and therefore should have different mean values. Basic anomaly detection, and therefore bad data detection, can also be pursued from order statistics.

5.3. Application of Statistical Properties and Visualization

The normalcy, trends, and relationships for a set of real SCADA data are examined in this section. This dataset is provided by a North American utility and consists of approximately one

month's worth of five-minute SCADA data and several hours of two-second SCADA data for a wind farm. The two-second data is from a portion of the five-minute data time frame when the plant had high output. There are 13 total measured variables at different locations. The measurements are in kV, MW, and MVAr. Also provided are two state-estimator snapshots.

The objective of this application is to construct a well-informed hypothesis for what is going on in the system based only on measurement data. More specifically, the original goal was to determine whether the wind farm is doing the "right thing" with respect to reactive power support. The questions included the following: (1) What are the operators doing? (2) What are they using to make their decisions? (3) Are certain variables being controlled with respect to certain other variables? (4) What switching actions have occurred and why?

A couple key difficulties arise when attempting to use the available data to provide insight into the above questions. The most prohibitive difficulty with this application is that nothing is known about what events may have been captured in the data, so there is nothing with which to verify data mining methods, cross check results, and determine if any patterns found are indeed related to the true events. Other difficulties are due to the need to combine information from various data sources to obtain a working model. The labels on the data are not sufficient to provide a clear mapping of the 13 data points to locations in the state estimator cases. The state estimator cases are full topology cases; we use the "consolidate case" integrated topology processing option in PowerWorld Simulator to make the cases more manageable [138], [139]. In this application, difficulties were also encountered in getting the cases to solve and making equivalents of the wind farm area from the full state estimator models.

The reconstructed view of the system in Figure 56 is based on a drawing from the utility. The measurement points are manually labeled based on point IDs of the SCADA data. Since limited information is available, Figure 56 represents a best effort to correctly match the measurements to their real locations. In Figure 56, all MVAr and MW entries are flows, and the four lines and transformers are labeled A,B,C,D. Measurements xC and xD denote the 230 kV side of transformers C and D. Notations fA and tA denote the "from-bus" and "to-bus" of line A.

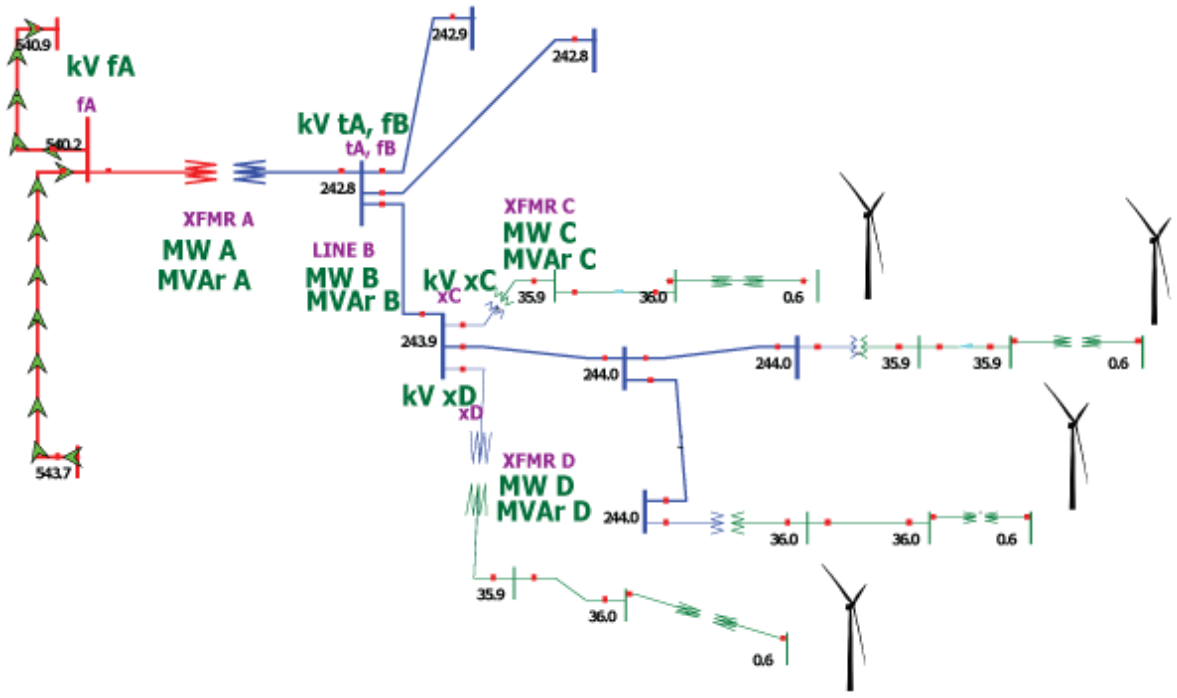

Figure 56. Reconstructed system with SCADA measurement locations labeled

The data contains many discrete changes, as indicated by two of the SCADA time series in Figure 57. These are likely due to deadbands in the recording of data points to the data historian. As evident, high-frequency information about transient changes is lacking.

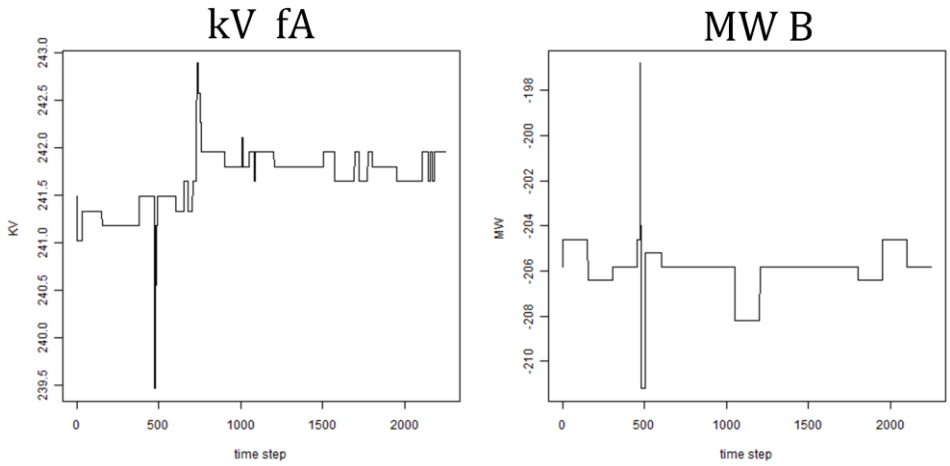

Figure 57. Examples of discrete changes in data

The programming language R is a powerful tool for statistical data processing which has been used to produce most figures in this section. To ascertain system behavior from data, it is important to be able to highlight the usual versus the unusual data characteristics. Statistical summaries such as PDFs and CDFs provide insight into the normal behavior of a system. Histograms are also useful statistical tools for illustrating where data spends most of its time (normalcy). The time series for the MW flow on line D and its histogram are shown in Figure 58.

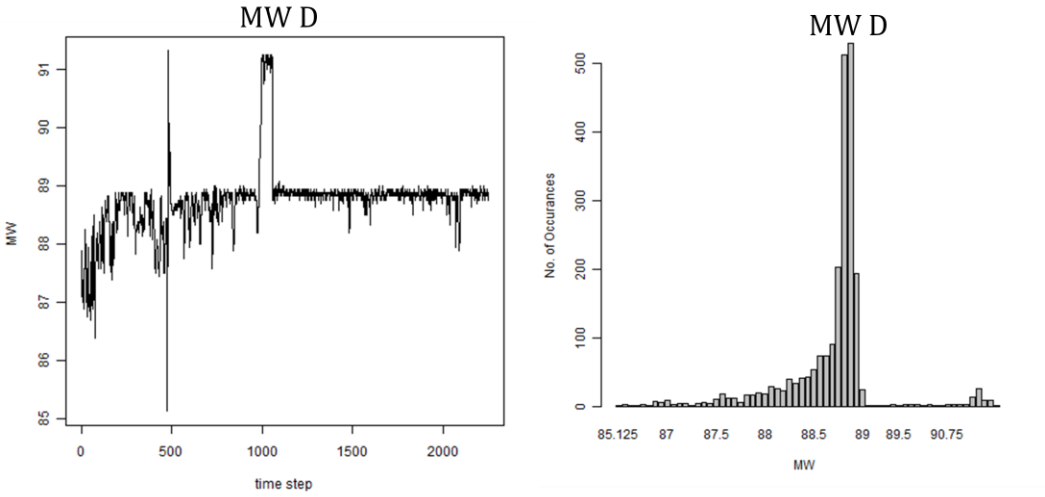

Figure 58. Time series (left) and histogram (right) for MW flow data on line D

While viewing these data distributions is useful, it is difficult to compare many at once and obtain a higher-level picture of the system conditions and relationships.

To obtain such a higher-level view of the system, it is useful to extract relationships in the form of visual indicators. In particular, correlations capture the relationships between variables in an easily interpretable way. Furthermore, it is possible to effectively visualize correlations between multiple pairs of variables at one time and immediately notice patterns of interest such as strong couplings between variables.

Visual indicators of relationships in the data are shown using a technique called *sparklines* [140]. Sparklines use simple visual representations to convey a lot of information simultaneously. For power systems, sparklines provide an initial view where the viewer can see, at a glance, trends in

the data. The sparklines plot in Figure 59 illustrates cross correlations between the variables. Each row and column represents one time-series of measurements from one location in the system. The entries in the matrix are cross correlations of the two time series in the row and column. Entries show at a glance the way measurements at a location are correlated in time with measurements at other locations. The plots were generated using a script written in R. The diagonal entries show autocorrelation, correlation of a signal with itself. The off-diagonal entries show cross-correlations of one time series with the other time series. A peak in the middle of an entry indicates that the highest correlation occurs at a time lag of zero. For autocorrelations, the peak is always at a time lag of zero. If the two time series are independent, the cross-correlation is zero at all times. Sparklines are a general visualization technique and can show other relationships besides correlations.

Relationships between MVAr A and MVAr B, and also between MVAr C and MVAr D, are highlighted by boxes overlaying Figure 59. This is one example; any of the relationships between variables can be picked out. The C-D boxed item shows a high, narrow peak, while the A-B item shows a high peak that is less narrow. Both examples show high correlations, but that of C-D is higher and that of A-B shows more variance. Objects A and B are each negatively correlated to C and D, which is also easy to see. Corroboration for this example is given next.

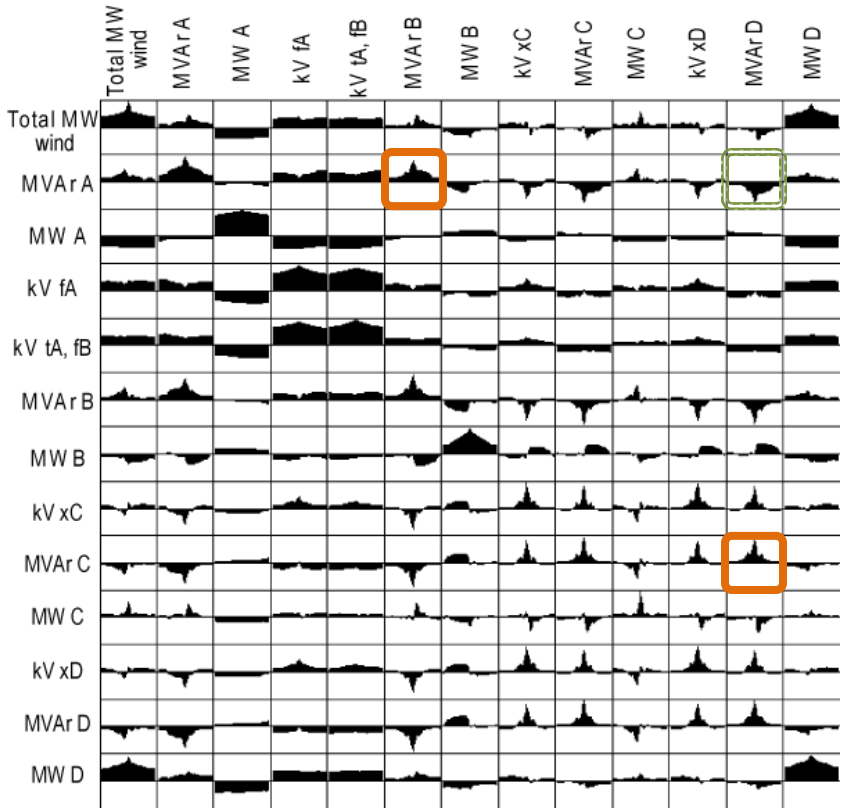

Figure 59. Sparklines plot, illustrating cross correlations in SCADA data

Insight can be gained from SVD, also referred to as principal component analysis (PCA) in the data mining area. A *screeplot* helps visualize the relative importance of the components. A sharp drop in magnitude signals that subsequent factors are ignorable. Figure 60 shows that most of the variability in the data is captured in just the first singular value, or principal component. Thus, we can see that we are justified in using a biplot, introduced in [141].

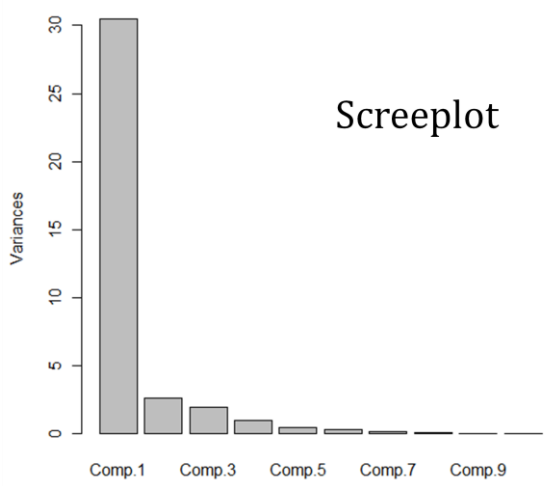

Figure 60. Screeplot of SCADA data indicating that most data can be represented using first singular value

Biplots are used to visualize the first two (for 2-D) principle components of a data matrix to show relationships between rows and between columns. The discussion extends easily to 3-D. As described in Section 2.5, a matrix **Y** can be factored into its SVD:

$$\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}} \tag{138}$$

Then, a rank 2 matrix $\mathbf{Y_2}$ is formed by keeping only the first two singular values,

$$\mathbf{Y}_2 = [\mathbf{u}_1 \quad \mathbf{u}_2]\begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}\mathbf{v}_1^{\mathrm{T}}\\ \mathbf{v}_2^{\mathrm{T}}\end{bmatrix} \tag{139}$$

where $\mathbf{u_1}$ and $\mathbf{u_2}$ are the first two columns of **U**, and $\mathbf{v_1}$ and $\mathbf{v_2}$ are the first two columns of **V**. If a matrix is exactly rank 2, it is exactly represented by its biplot. Thus, a biplot shows a 2-D representation of high-dimensional data. A biplot is a representation of how rows and columns map to the first two singular values. In biplots, generally both rows and columns are plotted, allowing dual representation of both time and variables in this application. For simplicity of illustration, the time information on the biplot of the SCADA data shown in Figure 61 is removed to just focus on the relationships of the variables.

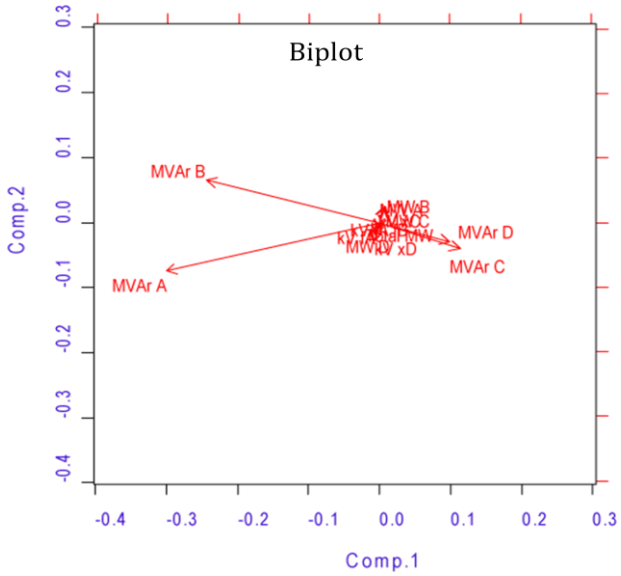

Figure 61. Biplot of SCADA data showing relationships between variables

From Figure 61, it is apparent that only four variables contribute to the variability in the data. These are MVAr A, MVAr B, MVAr C, and MVAr D. MVAr C and MVAr D contribute in an opposite way to that of MVAr A and MVAr B. MVAr C and MVAr D are more correlated than MVAr A and MVAr B. The length of the MVAr A and MVAr B vectors indicates that they have higher standard deviations. The use of SVD to produce biplots is just one instance of its value for many aspects of data mining in power systems. Plotting the largest singular value using a moving time window as in Figure 62 can reveal system stress [142].

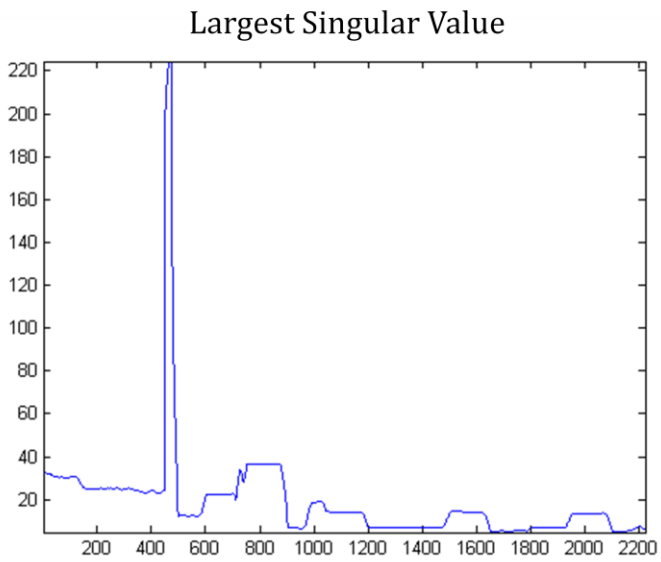

Figure 62. Largest singular value of SCADA data plotted for a moving time window

Following the explanation of SVD in Chapter 2, one can expect that significant changes in operating conditions show up as changes in **S** and **U**. From a geometrical interpretation, this is equivalent to the main direction and scaling of the shape of the data changing. From this simple analysis, it seems clear an event occurred between approximately the 400th and 500th time sample.

However, what happened is unclear, since we have no knowledge of what actually occurred to verify against. The mapping of changes in singular values to events remains to be further investigated; such a study requires the ability to validate true events against hypotheses regarding detected events. Fortunately, even without any additional knowledge, we are still able to observe that the suspected event may have something to do with the MVAr values. This observation is determined by analyzing the SVD and determining the major contributors to the largest singular value. Figure 59 also corroborates this hypothesis. In cases where knowledge about the system is minimal or nonexistent, it is beneficial to be able to cross-check hypotheses using independent methods.

Following the theory of SVD, it may also be insightful to look at the SVD of the sensitivities since they follow the same kind of input-output mapping as the data itself. Then, **V** corresponds

to the (orthogonalized) input data which are the injections, and **US** corresponds to the transformed data which are the voltages and angles of the PMU data, and **A** is the sensitivity matrix. In fact, the estimation applications in Chapter 4 developed as a result of this insight.

Additional enhancements to power system visualization in the area of data mining are possible, especially with respect to visualizing the system over time. In [143], a method for visualizing PMU data by reducing the system to an equivalent at the PMU buses is discussed. The analysis in [143] assumes the network is known, and an equivalencing procedure is performed to reduce the network to a Ward type equivalent at the PMU buses. The work in Section 4.2 does the opposite and reconstructs an equivalent without any knowledge of the system. The insight and visualization techniques developed in [143] are applicable to an equivalent estimated from PMU data. Visualization may include contours of angles, equivalent line flows, and more. Visualization may also involve creating a biplot which is animated in time, used to visualize the PMU-based equivalent (or other data, such as voltage magnitudes and angles) to indicate significant changes. Such a display would show how the contributors to the largest singular values of the system are changing in time. This can provide a quick indicator of problems in the system [142], as only the first two singular values are needed, which can be computed easily. Also, data from every time sample need not be displayed since the human eye is limited in its ability to detect small changes.

5.4. Conclusions

The value of data collected over time is apparent. This chapter parallels the work in Chapters 3 and 4 and presents some of the fundamental challenges associated with using raw measurement data to extract relationships. Section 5.1 applies the transmission line parameter estimation from Chapter 4 to real data. Results in Sections 5.1.2, 5.1.3, and 5.1.4 show that estimation based on real measurement data presents challenges not encountered in simulation. Section 5.2 shows how the data may be used to help identify sources of problems. However, it is difficult to determine whether the errors are present in the model, the data, or both. In Section 5.3, correlation and visualization techniques are shown to be useful for identifying relationships and revealing patterns in a measurement dataset.

The applications in this chapter illustrate the importance of visualization for knowledge discovery (from the KDD field of data mining, mentioned in Section 3.3.3), especially for finding patterns and relationships in power systems. Visualization can help address the need to process data quickly and report updates to the operator in an online fashion, possibly via a *dashboard display*. The idea is that, at a glance, the operator should see a condensed portrayal of important system indicators, like on the dashboard of a car. Animating a biplot over time may provide a useful dashboard display.

Empirical relationships should be compared to what is expected and supported by the model. A conclusion from this chapter is that substantial differences tend to exist between the actual data and the model. The measurements used in the estimation often exhibit noise on the same order of magnitude as the parameter being estimated. Factors other than noise, such as systemic bias, can also disrupt the measurements. Bad data is further addressed in Chapter 7. Anomalies in the measurement data which are not encountered in the model present complications, especially at a practical level. Such anomalies may represent errors or they may represent events. Future work remains in event detection and analysis to characterize how different power system events manifest themselves in the measurement data. For example, a stressed system exhibits higher singular values, and it would be interesting to find out more about what these high singular values mean, and to be able to map changes in key singular values to system events. Chapter 6 presents an application of event detection based on modal estimates of measurement signals.

In considering the use of data for diagnosing problems in the system, one needs to consider the difference between simulated data with errors and real data with errors. For the work in this chapter, a data simulator was created. With the data simulator, it is possible to add any desired error to the measurements. Even with the simulated data, it was found that some sets of operating points facilitate better estimates than others. The issue is that when simulating data, it is challenging to reproduce data with a realistic distribution. The question arises of how one should perturb the system to obtain representative operating points, since there are infinite possible ways to change the system and obtain a new operating point. Thus, rather than speculate on what representative operating points exist that should be studied, the most value is obtained from testing procedures on a wide variety of real data. The differences between simulated data with

error and real data with error provide further motivation for model validation, discussed in Chapter 4.

The final important point is that *real data* is important for developing data mining applications. It is difficult to represent real data errors in simulation. There is no substitute for validating applications with real data. It is especially important to obtain data corresponding to *known events* for learning and verification purposes.

6. Oscillation Monitoring Data Event Analysis

An oscillation which occurs as the system changes state following a disturbance is called a *ringdown*. This chapter examines the relationship between disturbances in the system and the measured modal content of the consequent ringdowns. The identification of patterns or distinguishing characteristics from oscillation data is studied in the context of linear system theory and control theory. The goal is to extract information from the ringdown about the disturbance or the event that causes it. This chapter examines how modal content is estimated and how that information enables the automatic analysis of events, creating tools for automated "grid detective" analysis.

Extraction of modal content can be accomplished in conjunction with an oscillation monitoring system (OMS). An OMS implements a set of methods to estimate the modal components of oscillation signals measured throughout the network. These estimates can reveal important characteristics of the system. The core of this work is to facilitate a better understanding of the way modes manifest themselves in measurement signals. One objective is to employ the information contained in a signal to help determine the cause of the oscillation.

For example, suppose there is just one PMU or measurement signal in the system. Then the question is to determine how different events manifest themselves in that signal and if it is feasible to distinguish them. Suppose, for example, a fault occurs at the terminals of a generator. The PMU will capture oscillation data that occurs as a result of this event. The goal is to take the estimated modal content of the event, as viewed from the PMU measurements, and ascertain which possible events may have caused the ringdown. The question is whether it is possible from a single signal to identify the characteristics of the actual event, and if it is, what prior information is needed. For example, if we are looking at generator outages, do we need to know information about the type and parameters of the generator? Or, can we figure out enough based on network location, in which case, what information do we need to know about the network? If we have knowledge of the network, we can reasonably understand what would happen to the steady-state values and dynamic states of the system under various events. The drawback is that if we have only steady-state information, many possible events could cause a particular change

in a value. The challenging question is uniqueness, or how to figure out what makes each event unique such that it can be identified.

The manner in which a particular set of characteristic modes is excited is related to the event itself, both its type and its location in the system. It is desirable to better understand what information about the event is captured in a ringdown. In this chapter, the modal content of oscillations is used to recover associations and patterns related to events in the system. In particular, we examine the modal content associated with particular generators or events. Using information about events and their signatures, additional tools can be developed to perform forensics and facilitate improved system operation and control.

The chapter is organized as follows. In Section 6.1, oscillation monitoring systems are described in the context of this work. Then, Section 6.2 traces the eigenvalues (modes) associated with generators to their impact on measurements. This conveys the content of what may be observed in a ringdown signal. Then, in the sections to follow, applications are presented where the estimated modal content of ringdowns is used to identify particular events. In particular, a methodology for using OMS data to classify ringdown events resulting from generator faults is presented in Section 6.3 and Section 6.4. Results of the approach are presented using both a 37-bus case and a large WECC case. Extension to another event type, exciter setpoint changes, is considered in Section 6.5.

6.1. Oscillation Monitoring Systems

Fundamental to this work is the existence of a means to routinely extract the model content of signals; such functionality constitutes an oscillation monitoring system (OMS). When oscillations are present, it is often desirable to characterize the signal with respect to its frequency content. An OMS [144] evaluates the modal content of measured oscillations in power system signals, cross-checks the results using several methods, and generates alerts when poorly damped modes are discovered [144]. Estimates are computed of the frequency, damping, amplitude, and phase of the components in the sign al [145].

Useful information about operational reliability is obtained from the estimated modes. Such an OMS can serve as an early warning of serious events. Alerts can prompt actions designed to enhance operational reliability. The report [144] discusses methods to detect in real time the danger of growing or poorly damped oscillations, while still in the early stages of an event. In this system, to perform cross-check or validation, moving time-window analysis is used, as well as multiple signal groups, and also different types of signal processing engines. The three methods analyzed in [144] are Prony analysis, the matrix pencil method, and the Hankel total least squares (HTLS) method. The matrix pencil method is from the system identification and spectrum estimation areas. All three methods use singular value decomposition (SVD). In summary, the main purpose of an OMS is to recover accurate oscillation information about the power system quickly so that it may be useful in detecting the onset of and thereby preventing catastrophic events.

An overview of power system applications of linear analysis techniques appears in [146] (2001). Some notable algorithms which are described are the Prony method and the eigensystem realization algorithm (ERA). Key on-line applications include estimating the critical mode (poorest damping) in the system; this is implemented by a critical mode estimator (CME) based on Prony analysis and used in an on-line dynamic security assessment (DSA) tool. Spectral analysis methods can be classified as *non-parametric* since they are direct techniques for estimating transfer functions without first selecting a set of possible models to which the system must belong [24].

In this section, oscillation monitoring is described in the context of this work. First, Section 6.1.1 examines ringdowns from a theoretical perspective. Then, Section 6.1.2 presents Prony analysis, used throughout this chapter. Section 6.1.3 concludes with an outline of our approach for using this modal content to characterize events, presented throughout the chapter.

6.1.1. Power System Ringdowns and Linear System Theory

In power systems, it is imperative to obtain information about damping and stability. This information is present in the oscillations. Ringdowns commonly appear as exponentially damped sinusoids. In this section, we discuss the content of oscillation signals.

For a time-invariant linear system, the impulse response is a complete characterization of the system [147], [24]. If the input into a linear system is the Dirac delta function, *δ(t)*, then the response of the system is a linear combination of all the modes of the system. The formal definition of the Dirac delta function is given in [17].

$$h(t) = b_0 \cdot \delta(t) + \text{ characteristic modes} \tag{140}$$

$$h(t) = b_0 \cdot \delta(t) + \left[P(D)\, x_0(t) \right] us(t) \tag{141}$$

The *us(t)* function represents the unit step function which is a 1 for time greater than 0. The unit step function enforces causality by considering the state to be zero at, as well as before, time equal to zero. Ideally, if the system could be perturbed by this input, the characteristic modes of the system would be clearly seen in the response.

Theoretically, a delta function perturbs the system by an infinitely short amount of time at all the frequencies, and then it is suddenly removed as the input. Thus, the way the system responds is completely governed only by its own characteristic modes, so these characteristic modes may be retrieved. When considering power system oscillations, the response will not be ideal for several reasons. Power systems are certainly not linear systems, and the input is obviously not a delta function. Also, it is impossible to perturb a system by a delta function because this would require an infinite magnitude and an infinitely small duration.

The total system response is the sum of a zero-input component and a zero-state component [147]. This should not be confused with the natural response and the forced response, which is a similar decomposition but not the same. The zero-input portion is the response of a linear system when there is no input. This response reveals the inherent behavior of the system. The zero-input component is the solution of the following equation:

$$\left(D^N + a_1 D^{N-1} + \ldots + a_{N-1} D + a_N \right) x_\phi(t) = 0 \tag{142}$$

where D represents the derivative operator d/dt. Equation (142) is satisfied by a solution of the following form:

$$x_{\phi}(t) = ce^{\lambda t} \tag{143}$$

Then, (142) can be written as

$$ce^{\lambda t}\left(\lambda^N + a_1\lambda^{N-1} + \ldots + a_{N-1}\lambda + a_N\right) = 0 \tag{144}$$

A non-trivial solution requires that the polynomial in (144) be equal to zero,

$$\lambda^N + a_1\lambda^{N-1} + \ldots + a_{N-1}\lambda + a_N = 0 \tag{145}$$

which may be rewritten as

$$\left(\lambda - \lambda_1\right)\left(\lambda - \lambda_2\right)\ldots\left(\lambda - \lambda_N\right) = 0 \tag{146}$$

so there are N possible solutions. Thus, (143) can be any linear combination of the exponential solutions,

$$c_1e^{\lambda_1 t}, c_2e^{\lambda_2 t}, \ldots, c_Ne^{\lambda_N t} \tag{147}$$

For proof, see [147]. These exponentials represent the poles, eigenvalues, roots, or characteristic modes of the system [147], [148]. Which one of the infinite number of possible solutions we actually obtain as the response is dependent upon the system input and upon the initial conditions.

The zero-state component is determined by assuming an initial state equal to zero. The zero-state component is the result of convolving the system input with the impulse response,

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) \tag{148}$$

where y is the output, u is the input, g is the impulse response, and t enumerates the sampling instants. Then, the output can always be calculated if the input is known. The impulse response in (148) is also a weighting function. Because of the fact that input is generally present, the zero-state response will generally contain both characteristic modes and non-characteristic modes of the system.

In reality, even if we could know the input and the impulse response, it is not generally possible to completely know the output. Other inputs affect the model in (148). Often, these external effects are lumped together into a noise or disturbance term, $v(t)$.

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t) \tag{149}$$

From (149), if the output is measured, the input is known, and the impulse response is known, $v(t)$ can be found. In a practical sense, noise is only as important as its effect on the output. Measurement noise can come from sensors as well as other signals which are not controllable inputs to the user. However, in some cases, the disturbances can be measured or are known. Then, the analysis can be appropriately compensated.

From these relationships, we can arrive at a transfer function for the input-output relationships in the system, and many system identification methods are aimed at estimating such transfer functions. Spectral analysis to determine transfer functions of linear systems is often used in time-series analysis. For oscillation monitoring, these methods are particularly important. Determination of transfer function equations for a second order system is discussed thoroughly in [149].

Often, the response observed in power systems is largely dominated by only one or two modes, so it may resemble a second order response. That is, the same damped sinusoidal waveform we

observe in time-domain power system ringdowns has a Laplace transform to a second order system in the frequency domain [149]:

$$Ae^{-\sigma t}\cos\left(\omega t+\phi\right)\longleftrightarrow\frac{a_1 s+a_0}{\left(s+\sigma\right)^2+\omega^2} \tag{150}$$

$$A=\left[a_1{}^2+\left(a_1\sigma-a_0\right)^2/\omega^2\right]^{1/2} \tag{151}$$

$$\phi=\tan^{-1}\left(\frac{\left(a_1\sigma-a_0\right)/\omega}{a_1}\right) \tag{152}$$

Interestingly, such an observed second order response is consistent with the type of response one would expect from a generator. A linearized second order model of a synchronous generator represents its electromechanical modes (machine angle and rotor speed),

$$\begin{aligned}\Delta\dot{\delta}&=\Delta\omega\\ \Delta\dot{\omega}&=-k_S\Delta\delta-k_D\Delta\omega\end{aligned} \tag{153}$$

where k_S is the synchronizing coefficient and k_D is the damping coefficient. The eigenvalues of (153) are

$$\lambda=\frac{-k_D}{2}\pm\frac{1}{2}\sqrt{k_D{}^2-4k_S} \tag{154}$$

In general, such a second order system is represented as

$$s^2+2\zeta\omega_n s+\omega_n{}^2=0 \tag{155}$$

where the poles may be expressed as

$$s=\sigma+j\omega \tag{156}$$

$$\omega=\omega_n\sqrt{1-\zeta^2} \tag{157}$$

$$\sigma=\zeta\omega_n \tag{158}$$

where ζ is the damping ratio, ω_n is the natural frequency, and ω is the damped natural frequency. Characteristics of the measured response can then be analyzed in order to determine the parameters ω_n and ζ of a transfer function, as in (159).

$$C(s) = \frac{\omega_n^{\,2}}{s^2 + 2\varsigma\omega_n s + \omega_n^{\,2}} R(s) \tag{159}$$

The system may be said to experience insufficient damping when ζ<0.03 and "acceptable" damping when ζ≥0.1 [150]. In a system with *N* machines, there are *N*-1 electromechanical modes, at the minimum. Thus, the true problem is more complicated.

Measured ringdowns in power systems can be better understood based on an investigation of their origin and by recognizing them as indicative of a disturbance. Estimated modal content can be used to our advantage to help diagnose events. The next section describes this modal content extraction, using Prony analysis.

6.1.2. Modal Estimation with Prony Analysis

Extraction of modal content from a signal is a fundamental part of what an OMS provides. This estimated modal content can provide a useful summary of transient stability runs as well as of measured data. Oscillations in power systems are often well represented as the summation of damped exponentials, even though power systems are nonlinear. Parameters of measured sinusoids can be estimated using Prony analysis [145]. Since Prony analysis is widely used for modal analysis in power systems, it is now briefly described.

Prony analysis represents the signal in the following form [145]:

$$y = \sum_{i=1}^{p} B_i e^{\lambda_i t} \tag{160}$$

and estimates parameters B_i and λ_i of (160). The parameters of the fit are estimated such that (160) fits the observed signal. The signal in (160) is also commonly written in terms of sinusoids,

$$y(t) = \sum_{i=1}^{q} A_i e^{\sigma_i t} \cos\left(2\pi f_i t + \phi_i\right) \tag{161}$$

where σ_i is the damping, A_i is the amplitude, f_i is the frequency in Hz, and ϕ_i is the phase in radians.

Many variations exist in Prony analysis implementations. For our work, Prony analysis is performed following the implementation of the Prony toolbox available for MATLAB [151], which is described in detail in [152] in its original context for analyzing modes in reactor measurements. The properties of least squares (LS) and generalized least squares (GLS) to estimate a transfer function of the system and then identify oscillatory modes for power systems are investigated in [153]. Prony analysis can be considered a variation on the LS method, and GLS is an iterative approach to allow online updating of modal estimates. Prony analysis finds linear prediction coefficients and then solves for roots of a polynomial. Prony analysis algorithms are examined in [154] for the purpose of analyzing power system harmonics. The authors of [154] also discuss the tuning of parameters and compare the Prony method with the Fourier transform. In their analysis, SVD is found to be the algorithm with the best overall performance for estimation of the coefficients in the linear prediction model of the Prony method. The details on the algorithms are omitted here.

The work in [155] extends typical Prony analysis [145] to allow multiple signals to be analyzed to determine one set of modal estimates. That is, this allows information from multiple measurements which are redundant to be supplemental to each other by recognizing that they are all measuring the same modes. This type of technique is valuable since one of the major practical issues with Prony analysis is its behavior in the presence of noise [155], [144]. Other issues such as sample size, data de-trending, and selection of model order are also problematic, and are examined in [156]. A case study of a TCSC in [157] examines the response of the system to step inputs and noise inputs used to probe the system. The modal content is examined with the purpose of studying the damping effectiveness of the TCSC.

Issues of model accuracy and validation, emphasized previously in this document, naturally arise here as well. Prony analysis can help to validate models in the following way. As demonstrated in [157], it is possible to "probe" the real power system, and gauge the response of the generator of interest. The actual response can be compared to the simulation response to the same probe, and the model can be updated as necessary. This process can be automated using system identification techniques, and a lot of work is underway in this area. While model validation is not our primary focus, we expect to be able to extend techniques developed in this work to model validation. Unless otherwise stated, we assume that the models in the simulations are valid.

In summary, Prony analysis can be used to quantify the modal content of a ringdown by determining a best-fit reduced-order model both in time and frequency domains. Now we describe how one may use modal estimates to facilitate pattern recognition.

6.1.3. OMS as a Precursor to Pattern Recognition

In this section, we describe our framework to find patterns in modal content related to events, based on OMS modal estimate results. This section positions our work in the OMS context. The functionality of an OMS, and the information it provides, have been described, so here we examine how an OMS serves as a precursor to more in-depth analysis. The applications in this chapter provide specific examples of these possibilities. In a way, the OMS is a filter for dynamic data similarly to how state estimation is a filter for static data. It is true that the more accurately the modal parameters can be estimated, the better and more useful the post-processing event detection framework can become. Future work in the signal characterization area may thus have implications for power system event identification.

The framework presented here facilitates the use of OMS modal estimates for the discovery of event signatures in the ringdown data. Intuitively, different events have different impacts on the system and will thus produce different ringdowns. A plot of an angle measurement taken at the same location for different events is shown in Figure 63. Each time series on the plot shows the ringdown caused by a fault at a different generator.

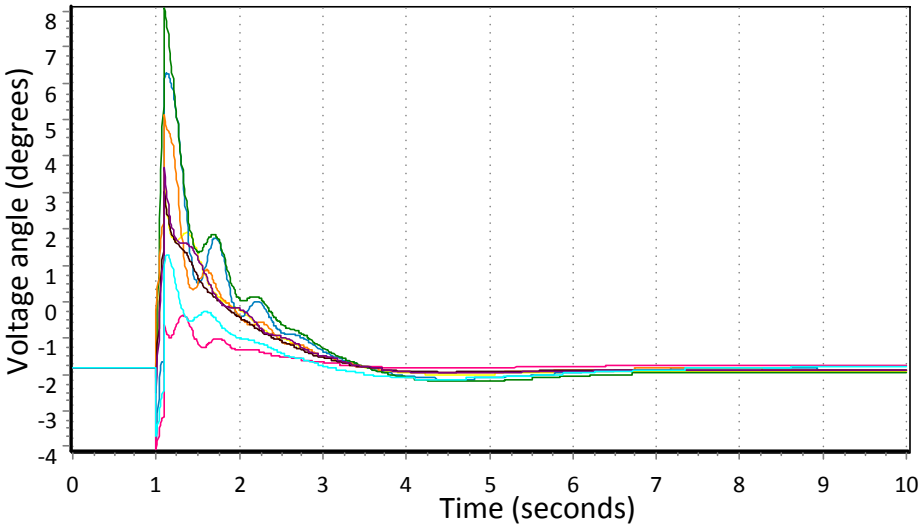

Figure 63. Ringdowns for different events, measured at one location

These ringdowns are all measured at the same place, and they are all different in some way. The question is whether this type of information is enough to distinguish between events. Validating associations or patterns between modal estimates and events make it possible to better recognize what is occurring based on the data.

Power systems are constantly changing. Small disturbances include load changes, while large disturbances include network topology changes and faults. In this work, the focus is predominately on what occurs during large disturbances. Changes in modal content associated with significant events motivate our work to recognize and classify patterns in oscillation data. The goal is to better understand what information is captured in these ringdowns so that they may serve as tools for performing forensics on the system. By building up a knowledge base with these relationships, we can help discover and validate new associations between measured data which is observed and what may have occurred.

In this work, we characterize events in terms of signatures which are then sought in the incoming data. In order for a ringdown to provide useful information about an event, it is necessary to identify attributes which are unique characteristics of the event. There is boundary which will likely be encountered where ringdowns or events become too similar to distinguish. Analysis is needed on the extent to which certain events are distinguishable from other events. For example,

do generator outages appear distinct in OMS results? What circumstances are necessary to assure that certain events will present distinct signatures? This is a challenge, and the inherent limitations of signature-searching remain to be fully explored.

The modal content of measurements is governed by the characteristics and parameters of every component in the system. Machine exciters, governors, power system stabilizers (PSSs), etc., all have some impact, as well as the types and locations of loads. In addition, there are effects due to the structure and parameters of the transmission network. Which modes are excited is related to the event itself, both the type of event and its location in the system. For example, an interesting association between the modal content of angle measurements and increased wind generation in Texas is recognized in [158]. Other event types of interest may include faults, generator trips, and system topology changes. In summary, we seek to discover signatures of system events and control actions and determine the impact on OMS results.

In the applications presented in this chapter, the focus is on recognizing events occurring at specific generators. A single event type must be considered at a time for consistency and comparison purposes. It is the intention to gradually extend the framework by adding the capability to detect a greater variety of events. The approach is outlined in Figure 64.

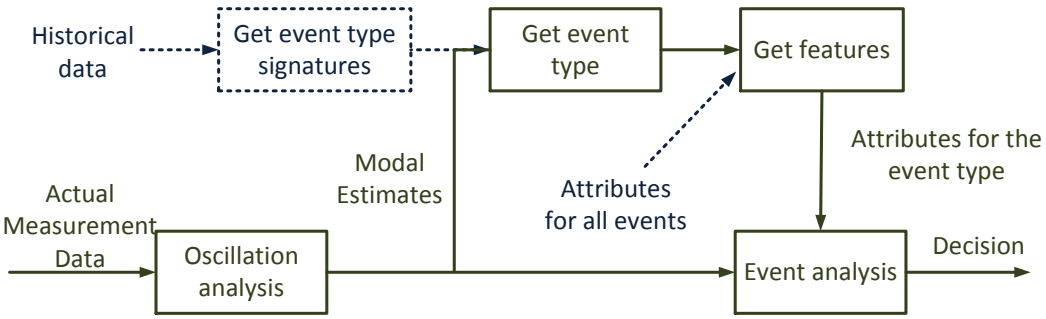

Figure 64. OMS-based event analysis framework

The links shown by blue dotted lines indicate computation and analysis procedures that can be performed completely offline. The green solid lines represent an online process to quickly make decisions about the event. These online decisions would be based on information obtained offline as well as on the incoming measurement data. For this pattern recognition system, the features

investigated in this chapter are related to modal content. The same framework can be followed for other event types with other sets of features. Some particular applications of this framework are presented in the following sections.

6.2. Mapping Modes to Measurements and Events

As described in the previous section, the modes in a ringdown can be used to identify patterns in measurement data for classifying events. Results on estimating oscillation parameters of power systems in [159] corroborate our theory by showing that for a small disturbance, a measurement can be modeled as the output of a linear system excited by a delta function. Then, the signal is completely characterized by the coefficients of the numerator and denominator polynomial. One difficulty is that the events which cause ringdowns are not necessarily "small" disturbances for which the above theory is valid. The system may move from one state to another state or it may return again to the same state. We are interested in the damped oscillatory responses that occur as the system moves. Discovery of such patterns and relationships can lead to improved situational awareness by providing event information to operators. In this section, we investigate the relationships of the modal content in oscillations to known modes of the power system.

In this section, we develop an analytic representation for power system measurements in terms of the system eigenvalues for a single-machine-infinite-bus (SMIB) system. The generalization of this analysis to an *m*-machine, *n*-bus system is then described. These equations provide insight into what phenomena are being captured in the measurements and what is expected to be obtained from a mode meter. The goal is to see analytically how certain modes of this system appear in any of the "measureable" state variables.

6.2.1. Two-Axis Single-Machine-Infinite-Bus Approximation

Single-machine-infinite-bus (SMIB) eigenvalues are calculated by modeling one machine in detail while modeling the rest of the system as an infinite bus [160]. The SMIB eigenvalues for each generator can be obtained if the dynamic models are known. A SMIB model allows the local modes of a generator to be examined and can greatly simplify analysis. Detailed investigation of SMIB models is found in [161]. In this section, we develop a representation of power system measurements in terms of the eigenvalues of a SMIB system.

For a SMIB, the equations are representative of local phenomena in the full system, and the associated equations are easy to analyze. The SMIB system in Figure 65 [160, pp. 100] uses the two-axis machine model representation. The infinite bus voltage magnitude is V_∞. The equivalent transmission line which connects the single machine to the infinite bus is modeled by parameters R_e and X_{ep}.

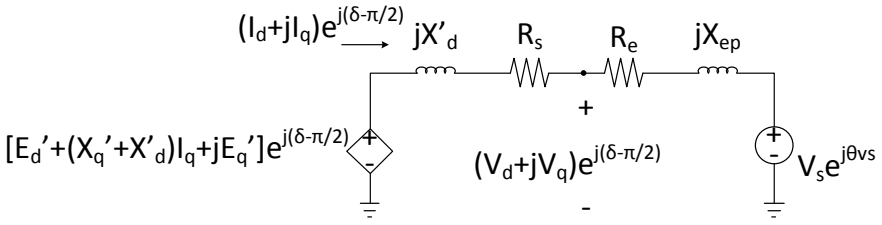

Figure 65. SMIB two-axis machine model [160]

The two-axis machine can be represented by four differential equations or states. Additionally, other models can be represented along with the two-axis model. Typically, generator models also include exciter and governor dynamics, at minimum. Addition of the IEEE Type 1 exciter to the model [160] results in two additional states. Numerous generator models and corresponding control system models can be found in the literature [162, 163, 164, 165]. The block diagram model for the IEEE Type 1 exciter is shown in Figure 66, from [166].

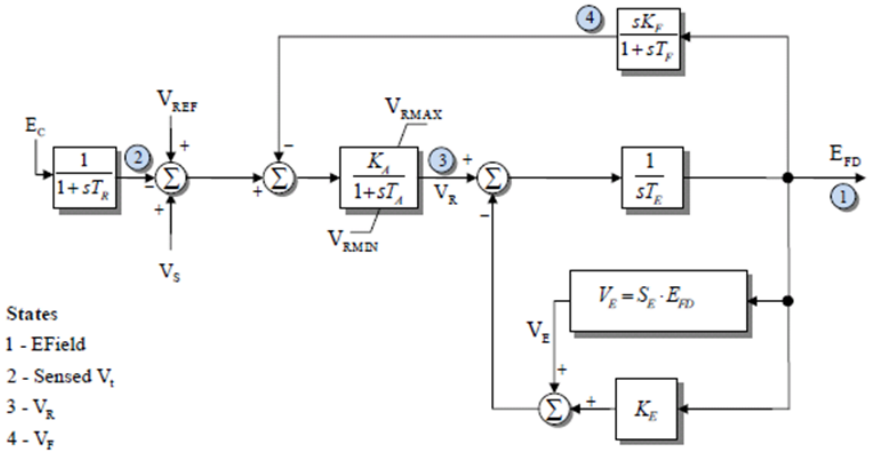

Figure 66. IEEET1 exciter model [166]

The four differential equations representing the machine dynamics correspond to states of rotor angle $\delta(t)$, speed $\omega(t)$, $E_q'(t)$, and $E_d'(t)$, where the (t) is dropped from further notation,

$$\dot{E}_d' = \left(-1/T_{q0}'\right)\left(E_d' + \left(X_q - X_q'\right)I_q\right) \tag{162}$$

$$\dot{E}_q' = \left(-1/T_{d0}'\right)\left(E_q' + \left(X_d - X_d'\right)I_d - E_{fd}\right) \tag{163}$$

$$\dot{\delta} = \omega - \omega_S \tag{164}$$

$$\dot{\omega} = \frac{\omega_S}{2H}\left(T_M - E_d' I_d - E_q' I_q - \left(X_q' - X_d'\right)I_d I_q - T_{FW}\right) \tag{165}$$

and a damping term of $T_{FW} = D(\omega - \omega_s)$ is assumed. In the above system, E_{fd} can be modeled as an input or as a state of the exciter. The IEEE Type 1 exciter is represented by three state equations,

$$\dot{E}_{fd} = \left(-1/T_E\right)\left(\left(K_E + S_E\left(E_{fd}\right)\right)E_{fd} - V_R\right) \tag{166}$$

$$\dot{R}_f = \left(-1/T_F\right)\left(R_f - \frac{K_F}{T_F}E_{fd}\right) \tag{167}$$

$$\dot{V}_R = \left(-1/T_A\right)\left(V_R - K_A R_F + \frac{K_A K_F}{T_F}E_{fd} - K_A\left(V_{REF} - V_t\right)\right) \tag{168}$$

where V_{REF} is an input, and limits on V_R are given by

$$V_R^{\min} \le V_R \le V_R^{\max} \tag{169}$$

Turbine governor dynamics are neglected, which is synonymous with modeling T_M as a constant.

Note that there is also a simpler one-axis model called the flux decay model [160]. The simpler model represents most of the impact we seek to capture, the relationship between machine states and terminal algebraic variables. In summary, there is little simplification on the relationships of interest for this application when the one-axis model is used. The relationship we seek to capture can be seen just as easily in the two axis model, which is a more accurate model.

6.2.2. From State Variables to Measurements

For a two-axis machine model as in Figure 65 and a network model, as represented by the infinite bus and transmission line, one can compute the effect of state variables on measureable variables such as terminal voltage and current. The stator algebraic equations for the two-axis model can be written as

$$E_d' - V_t \sin(\delta - \theta) - R_S I_d + X_q' I_q = 0 \tag{170}$$

$$E_q' - V_t \cos(\delta - \theta) - R_S I_q - X_d' I_d = 0 \tag{171}$$

where is V_t is the terminal voltage magnitude,

$$V_t e^{j\theta_t} = \left(V_d + jV_q\right) e^{j(\delta - \pi/2)} \tag{172}$$

and after eliminating V_t, the stator algebraic equations simplify to

$$V_d = E_d' - R_S I_d + X_q' I_q \tag{173}$$

$$V_q = E_q' - R_S I_q - X_d' I_d \tag{174}$$

The terminal current into the network is

$$\left(I_d + jI_q\right) e^{j(\delta - \pi/2)} = \frac{\left(V_d + jV_q\right) e^{j(\delta - \pi/2)} - V_\infty}{R_e + jX_e} \tag{175}$$

which may also be simplified to obtain

$$V_d = R_e I_d - X_e I_q + V_\infty \sin\delta \tag{176}$$

$$V_q = X_e I_d + R_e I_q + V_\infty \cos\delta \tag{177}$$

Combining equations (173) and (174) with equations (176) and (177), leads to the following set of equations:

$$E_d' - R_S I_d + X_q' I_q = R_e I_d - X_e I_q + V_\infty \sin\delta \tag{178}$$

$$E_q' - R_S I_q - X_d' I_d = X_e I_d + R_e I_q + V_\infty \cos\delta \tag{179}$$

where V_d and V_q have been eliminated. After some manipulation of (178) and (179), I_d and I_q can be solved for in terms of the machine constants, network constants, and state variables,

$$I_d = c_1 E_q' + c_2 \cos\delta + c_3 E_d' \tag{180}$$

$$I_q = c_3 E_q' + c_4 \cos\delta + c_5 E_d' \tag{181}$$

where values for the constants in the system of Figure 65 are given by

$$c_1 = \frac{\left(X_q' + X_e\right)}{\left[\left(R_e + R_S\right)^2 + X_d'\left(X_q' + X_e\right)\right]} \tag{182}$$

$$c_2 = c_1 V_\infty \left[\frac{j\left(R_e + R_S\right)}{\left(X_q' + X_e\right)} - 1\right] \tag{183}$$

$$c_3 = c_1 \frac{\left(R_e + R_S\right)}{\left(X_q' + X_e\right)} \tag{184}$$

$$c_4 = c_2 \frac{\left(R_e + R_S\right)}{\left(X_q' + X_e\right)} - \frac{jV_\infty}{\left(X_q' + X_e\right)} \tag{185}$$

$$c_5 = c_3 \frac{\left(R_e + R_S\right)}{\left(X_q' + X_e\right)} - \frac{1}{\left(X_q' + X_e\right)} \tag{186}$$

Thus, I_d and I_q are linear functions of cos(δ), E_q', and E_d'. Similarly, expressions can be found for V_d and V_q in terms of the system constants and state variables:

$$V_d = \left[c_1 R_e - c_3 X_e\right] E_q' + \left[c_2 R_e - c_4 X_e - jV_\infty\right]\cos\delta + \left[c_3 R_e - c_5 X_e\right] E_d' \quad (187)$$

$$V_q = \left[c_1 X_e + c_3 R_e\right] E_q' + \left[c_2 X_e + c_4 R_e + V_\infty\right]\cos\delta + \left[c_3 X_e + c_5 R_e\right] E_d' \quad (188)$$

The current (180), (181) and voltage (187), (188) equations show specifically how the state variables impact the measurements. The rotor angle δ, E_q', and E_d' are again explicitly present. This analysis thus reveals the relative weights of state variables in a measurement.

A simple view of the problem is as a sensitivity problem, the sensitivity of measurements to changes in machine states. Then, machine states may be approximated as a sum of damped exponentials.

$$\frac{\partial z}{\partial x},\ z = \text{measurements},\ x = \text{machine states} \quad (189)$$

$$x(t) = g(\lambda, u) \quad (190)$$

$$z(t) = g_2\left(x, I_d, I_q, E_q', E_d'\right) \quad (191)$$

From the above analysis, the states which appear directly in the measurements are only three states, δ, E_q', and E_d'. It is therefore appropriate to focus on identifying the modes which appear in these three state variables; this is the focus of the next section.

6.2.3. From Eigenvalues to State Variables

Examining the transformation from eigenvalues to state variables reveals how certain modes may be expected in a measurement. Prior to this point in the analysis, no linearization has been used or required. To obtain the eigenvalues and eigenvectors to continue this discussion, linearization is necessary.

As described in [145], when a disturbance is applied and then removed to a linear time-invariant (LTI) system, the system will "ring down" according to the solution of the system differential equations,

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{192}$$

where **A** is the linearized system matrix from the dynamic equations in the power system. The eigenvalues, right eigenvectors, and left eigenvectors of **A** can be represented respectively as λ_i, $\mathbf{p_i}$, and $\mathbf{q_i}$.

This zero-input [147] response is governed by the characteristic modes of the system. There are a number of representations for the solution to (192). A desirable form is the following:

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x_0} \tag{193}$$

$$e^{\mathbf{A}t} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P^{-1}} \tag{194}$$

$$\boldsymbol{\Lambda} = \mathrm{diag}\left[e^{\lambda_1 t}, \ldots, e^{\lambda_n t}\right] \tag{195}$$

As discussed, the solution actually obtained depends upon the system input and the initial conditions. The eigenvalues, right eigenvectors, and left eigenvectors of **A** can be represented respectively for each mode i as λ_i ,$\mathbf{v_i}$, and $\mathbf{w_i}$, and C_i is a scalar. Then, as discussed in [145], the state of the system as a function of time may be written as

$$\mathbf{x}(t) = \sum_{i=1}^{n} \mathbf{v_i} C_i e^{\lambda_i t}, \quad C_i = \mathbf{w_i}^{\mathbf{T}}\mathbf{x_0} \tag{196}$$

where $\mathbf{x_0}$ is the initial system state and n is the dimension of the system.

For the model in Figure 65, the states form **x**, and the system **A** matrix is obtained by linearizing about the operating point and eliminating I_d, I_q, V_d, and V_q. Let its eigenvalues simply be represented by λ_1, λ_2, and λ_3. Then, keeping the initial condition vector distinctly represented, the solution may be written as

$$\mathbf{x}(t) = \left[\mathbf{v_1}\mathbf{w_1}^{\mathbf{T}}\right]\mathbf{x_0}e^{\lambda_1 t} + \left[\mathbf{v_2}\mathbf{w_2}^{\mathbf{T}}\right]\mathbf{x_0}e^{\lambda_2 t} + \left[\mathbf{v_3}\mathbf{w_3}^{\mathbf{T}}\right]\mathbf{x_0}e^{\lambda_3 t} \tag{197}$$

In summary, the state values depend on the modes as expressed by (197), while the terminal voltage and current depend on these states as described by equations (180), (181), (187), and (188).

6.2.4. General Impact on Measurements

While the previous section analyzed the relationship between modes and measureable algebraic variables, it is important to extend the discussion to a multi-bus, multi-machine system. The general impact on measurements is described in this section.

Since **A** is determined by the differential equations of generators and represents the eigenvalues or modes associated with the generators, it follows that a measured ringdown in the system will contain information about those modes. A measurement signal, such as line flow deviation or voltage deviation, is a linear combination of the states [159],

$$s(n) = \sum_k w_k x_k(n) \tag{198}$$

where w_k are weights, and samples are taken at $t=nT$, where T is the sample period. Thus, a measured signal contains information about the system modes and is also a sum of damped exponentials.

For a general system, the stator algebraic equations can be rewritten so that they are equations for I_d and I_q ($\mathbf{I_{d\text{-}q}}$) as functions of power flow variables **V** and machine states **x**. The stator algebraic equations define the relationship between measureable variables and the machine states, as sought in (189). The whole system can be linearized as in [160] to obtain the following:

$$\dot{\mathbf{x}} = f_0\left(\mathbf{x}, \mathbf{I_{d\text{-}q}}, \mathbf{V}, \mathbf{u}\right) \tag{199}$$

$$\mathbf{I_{d\text{-}q}} = h(\mathbf{x}, \mathbf{V}) \tag{200}$$

$$0 = g_0\left(\mathbf{x}, \mathbf{I_{d\text{-}q}}, \mathbf{V}\right) \tag{201}$$

Hence, the terminal equations are essential for the defining relationship between **V** and **x**. The linearized system simplifies to the following [160]:

$$\begin{bmatrix} \Delta\dot{\mathbf{x}} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{A_1}' & \mathbf{B_{AE}} \\ \mathbf{C_{AE}} & \mathbf{J_{AE}} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{bmatrix} \tag{202}$$

$$\Delta\dot{\mathbf{x}} = \left(\mathbf{A_1}' - \mathbf{B_{AE}}\mathbf{J_{AE}}^{-1}\mathbf{C_{AE}}\right)\Delta\mathbf{x} \tag{203}$$

$$\Delta\mathbf{y} = -\mathbf{J_{AE}}^{-1}\mathbf{C_{AE}}\Delta\mathbf{x} \tag{204}$$

The algebraic variables **y** include $\mathbf{I_{d\text{-}q}}$ and the power flow states **V**. Any other power system measurement variables can then be calculated from **V**. The eigenvalues and eigenvectors of the matrix in (202) and the algebraic equations (204) thus provide a way to quantify the impact of modes on states and on measurable variables.

Additionally, a ringdown can be considered to be in one of three categories. (1) The system is stable and the value returns to the equilibrium value $\mathbf{x_e}$. (2) The system is stable but returns to some new equilibrium point, denoted $\mathbf{x_{e,new}}$. (3) The system is unstable. The focus of this work is on the oscillations for the first two types. Future work is possible to obtain more information by simultaneously considering the steady-state changes that occur in situation (2).

6.3. Generator Fault Event Classification

The focus of this section is on identifying a single event type, three-phase to ground faults at the terminals of generator buses. Sections 6.3.1 and 6.3.2 present algorithms well-suited for this identification task. Results for a 37-bus case are presented in Section 6.3.3. Examining a single event type is necessary for consistency, and the present choice is sufficient to demonstrate the concepts of this work. When a fault occurs, an immediate change in the terminal characteristics of the generator results. These changes are propagated through the machine states. If the fault is cleared quickly, the event appears as a brief spike in terminal values, followed by an oscillation.

Many issues involved with determining system damping are described in [150]. According to [150], oscillatory modes with negative damping (positively growing exponentials) are generally

not encountered in a system where speed governors and voltage regulators have been eliminated. Thus, oscillatory instability is almost entirely due to the effects of speed governors and voltage regulators. When attempting to map SMIB eigenvalues to Prony analysis ringdowns, it is important to keep in mind the fundamental influence that regulators have on damping. This importance appears especially with regard to low frequency oscillations [150]. Another difficulty which we encounter is also described in [150]: "Qualitative determination of damping from swing curves may be difficult when the contribution of various oscillation modes is important." That is, depending on what the event is, any modes in the system may be excited, not just the ones near the event. In such cases, locating the precipitating generator is more difficult; it may be advantageous to examine the coupling between the different generators' modes [161] and incorporate this into the solution.

For the generator faults considered, the goal is to use oscillation data to identify which generators are likely to have been involved. Key eigenvalues and associated generators may be known *a priori*, especially if a transient stability model of the system is available. Analysis and experiments show that the electromechanical modes associated with angle and frequency tend to be the most prominent in the ringdown for three-phase faults at the generator terminals. For a SMIB system, these modes can be found by examining participation factors.

Participation factors are a measure of the effect of a particular state on a particular eigenvalue,

$$p_{ki} = \frac{\partial \lambda_i}{\partial a_{kk}} = \frac{\mathbf{w_{ki}} \mathbf{v_{ki}}}{\mathbf{w_i}^{\mathbf{T}} \mathbf{v_i}} \tag{205}$$

where they relate the *k*th state variable to the *i*th eigenvalue. These are determined from right and left eigenvectors **v** and **w** respectively. Eigenvectors may be chosen with any convenient scaling; they are typically scaled to make $\mathbf{w}^{\mathbf{T}}\mathbf{v} = 1$. Thus, participation factors show which state variables reflect or capture certain modes. Then, by inspection one can determine which modes represent the angle and speed states.

Two distinctions concerning the assumed measured signal should be made before proceeding. First, a two-bus equivalent system can be constructed based on a SMIB model: one bus represents the machine, while the other bus and transmission line represent the rest of the system. Ringdowns in the two-bus equivalent represent modes of the particular generator more clearly since the dynamics of all other generators are not modeled, whereas ringdowns in the full system contain the impact of all generators. It is useful to compare the modal content of a ringdown in the two-bus system to that of the full system. Additionally, it is important to distinguish between the rotor angle and the bus angle. In simulation, it is possible to obtain the internal rotor angle which is a state of the generator, and this angle tends to more clearly preserve the modes. Conversely, the bus angle, which is possible to measure, is a function of solving the network's algebraic equations, and may not preserve the modes as well.

There are properties which make some modal estimation approaches better suited than other approaches for the goal of performing mode-matching for events. Ideally, the two components of the problem, modal parameter estimation and event signature identification, can be considered together. This may be directly integrated into the OMS, leading to a more online versus a post-processing event analysis approach. In reality, some combination of the two will be used. We aim to take advantage of the most integrated approach possible, but it is likely that some post-processing will inevitably be required.

The next two sub-sections present modal estimation methods as part of this integrated approach. In particular, these two methods have special application towards obtaining the goals of this chapter. These are the mode-matching least-squares (LS) algorithm and the Steiglitz-McBride algorithm. A method is sought with the ability to constrain the modal estimation problem to prefer to return certain modes. These methods are both tailored towards incorporating initial estimates of dominant modes. As examined in [167], these two methods can be modified to account for some knowledge of the modes which are to be estimated. The ability to account for known or expected modes is a feature which is particularly appealing in this application. The results of these methods are also compared to the conventional Prony method, described in Section 6.1.2. In this chapter, Prony analysis essentially serves as a base-case modal analysis method for examining and validating results. While Prony analysis provides a good starting

point, the purpose of examining other methods is to be able to compare the response at different frequencies. It is important to point out that these methods for obtaining the modal estimates used in this chapter are by no means the only approaches. Performance comparisons of several methods are given in [168].

We obtain a set of eigenvalues associated with each generator, and these eigenvalues are compared to the results of using Prony analysis on the oscillations that occur as a result of different events in the system. As an example, suppose the eigenvalues associated with a certain generator or set of generators (or all the generators in the case) are known. Many of the key eigenvalues in a system and their associated generator can be known approximately, assuming a transient stability model of the system is available. Thus, this is a relatively reasonable assumption to make. Then, when an event occurs near one of the generators, a measurement device such as a PMU located somewhere in the system will capture the oscillation data that occurs as a result of this event. The question is whether we can take the ringdown data of the event and ascertain what possible events may have been the cause. We may be able to tell which generator or generators are likely to have been involved in the event. Specifically, we look for patterns which relate the modal content of oscillations in a signal measured at one location in the system to the "SMIB eigenvalues" of the generators in the system.

The methods in this section are used to investigate the extent to which events at different generators have distinct signatures in OMS results. The coefficients of an estimated transfer function are termed a linear predictive code (LPC) or model (LPM) [159]. This LPM gives a model of the signal, and from this signal model we can compute various signal attributes. Important attributes are frequency, presence or absence of critical modes, degree of damping, and amplitude of signal components. The future evolution of the signal can also be predicted from the LPM.

Approaches for LPM characterization of the signal are discussed and demonstrated in this section. Differences are often present in how the coefficients of the LPM are estimated. Understanding the benefits and limitations of particular identification methods with respect to signature-seeking is important. The methods explored in this section are the Prony method

[145], the Steiglitz-McBride algorithm [169] with initial mode estimates, and mode-matching least-squares (LS) estimation. Signatures of different events are evident. These approaches illustrate the feasibility of distinguishing different events from each other in the resulting ringdown data. Research on techniques to further characterize the signatures and to appropriately distinguish between events is warranted.

6.3.1. Steiglitz McBride Algorithm with Initial Parameters

The Steiglitz-McBride (SM) algorithm, introduced in [169], is also used to estimate the modal content of power system ringdowns [168], [171]. The SM algorithm iteratively computes Kalman estimates to identify a transfer function $N(z)/D(z)$:

$$N(z) = \alpha_0 + \alpha_1 z^{-1} + \ldots + \alpha_{n-1} z^{-(n-1)} \tag{206}$$

$$D(z) = 1 + \beta_1 z^{-1} + \ldots + \beta_n z^{-n} \tag{207}$$

Each iteration of the algorithm determines updates of the coefficients $\alpha_0,\ldots,\alpha_{n-1}$, and $\beta_1,\ldots\ \beta_n$ until the estimated response converges to the true response [170]. Convergence properties of the algorithm are examined in [170].

The applicability of the SM algorithm in power systems for the purpose of studying low frequency electromechanical oscillations and for identifying low-order models from simulation data is examined in [171] and [168]. Performance comparisons of the Steiglitz-McBride algorithm, the Eigensystem Realization Algorithm, and the Prony method are given in [168]. As explained in [171], the estimated transfer function can be written as a sum of the residues over the poles:

$$H(s) = \sum_{i=1}^{n} \frac{r_i}{s - \lambda_i} \tag{208}$$

The SM method can be executed in two forms, either with or without the use of initial parameters [169]. Executing the SM algorithm with specific initial parameters is advantageous for cases where certain modes seem to be of interest, as in this application. The SM algorithm is especially useful because we can specify modes to serve as an initial estimate for the modes of

the LPM. In particular, SMIB eigenvalues and modal estimates from other methods serve as useful initial conditions. The SM algorithm implementation used for this work is the Seiglitz-McBride function in MATLAB's Signal Processing Toolbox [172].

6.3.2. Mode Matching Least Squares

The approach described in this section seeks a direct mapping between the estimated modes of the response and the SMIB eigenvalues. Instead of estimating all of the modal parameters, it is possible to hold certain modes fixed and estimate only their amplitudes. Using least-squares estimation, it is thus possible to force the estimate to include certain known modes. For this application, the specified modes can be SMIB eigenvalues. Specified modes (damping and frequency) are held fixed while their amplitudes and initial phases are estimated.

The objective is to solve the following:

$$\min_{A_i,\phi_i} \left\| \mathbf{f}(t) - \mathbf{y}(t) \right\|^2 \tag{209}$$

where the vector **f**(t) is the estimated signal and the vector **y**(t) is the true signal. For this method, consider the estimate **f**(t) to be a summation of unknown coefficients multiplied by known sinusoids:

$$\mathbf{f}(t) = \sum_{i=1}^{q} A_i e^{\sigma_i t} \cos\left(\omega_i t + \phi_i\right) = \sum_{i=1}^{q} \underbrace{A_i e^{j\phi_i}}_{unknown} \underbrace{e^{\sigma_i t} \cos\left(\omega_i t\right)}_{known} \tag{210}$$

The magnitudes and initial angles of modes in the signal can be estimated by solving a linear matrix, as long as the dampings and frequencies are known. These estimates are obtained by solving the following linear matrix equation for **a**,

$$\mathbf{E}\mathbf{a} = \mathbf{y}(t) \tag{211}$$

where matrix **E** and vector **a** have the following structures:

$$\mathbf{E} = \begin{bmatrix} e^{\gamma_1(0\Delta t)} & \cdots & e^{\gamma_n(0\Delta t)} \\ \vdots & \ddots & \vdots \\ e^{\gamma_1(N\Delta t)} & \cdots & e^{\gamma_n(N\Delta t)} \end{bmatrix} \tag{212}$$

$$\mathbf{a} = \left[A_1 e^{j\theta_1}, \ldots, A_n e^{j\theta_n} \right]^T \tag{213}$$

n is the number of modes, and *N* is the number of time points. Then, the least-squares solution is

$$\mathbf{a} = \left[\mathbf{E}^\mathbf{T}\mathbf{E} \right]^{-1} \mathbf{E}^\mathbf{T}\mathbf{y} \tag{214}$$

where **a** is a vector of complex numbers which, in polar notation, give the magnitude and initial angle of the signal.

This method is particularly useful for this application because one may choose a priori that the modes include the SMIB eigenvalues of interest. Then, the SMIB eigenvalues are forced to appear in the Prony analysis results, and only their amplitudes and initial phases need to be determined. Once these are determined, one may examine the amplitudes of the modes to determine which generators are significant contributions to the event.

The quality of the resulting approximation depends extensively on which modes are specified. If a careful choice of representative modes is made initially, then **f(t)** is approximately equal to the original signal. Conversely, if the modes are poorly selected, a meaningful approximation may not be found. We have experimented with several ways of choosing the initial modes for this algorithm. Suggestions include the SMIB eigenvalues, possibly in conjunction with estimates obtained from other methods. In conclusion, this method is not as robust as the SM algorithm, since if the modes chosen to match for this algorithm are not a good match to the true signal, the results of the estimation will not be valid.

6.3.3. Results for 37-Bus Ringdowns

This section presents experimental results of events at several locations and examines the identifying information that is present in the ringdowns from these events. The example system is the GSO 37-bus case [173]. The one-line diagram for the system is shown in Figure 67. The

same study system is used throughout the chapter. Results of Prony, Mode-matching LS, and SM methods applied to the GSO case are presented in detail in [167], along with discussion of tradeoffs to be made in choosing the various methods. In this section, some results are highlighted.

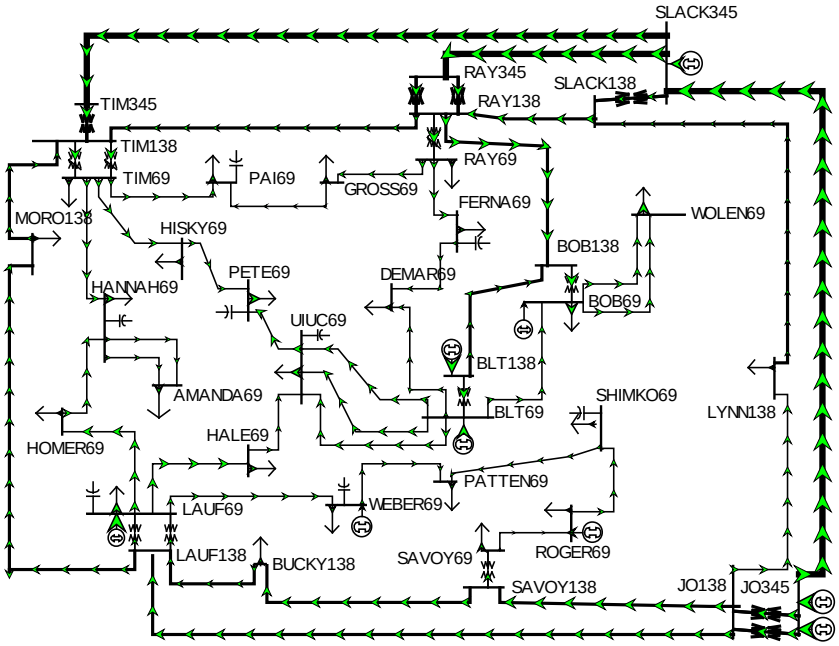

Figure 67. GSO 37-bus system

The bus angle measurement "BUCKY138" in this system is used as the simulated PMU signal. The angle reference in the simulations is the average of all generator angles.

In Figure 68a, several SMIB eigenvalues which have low damping are shown. These correspond to generators BLT138, LAUF69, ROGER69, BLT69, and BOB69. The Prony analysis results of ringdowns due to simulated three-phase-to-ground faults at the generators followed by the generator tripping are shown in Figure 68b.

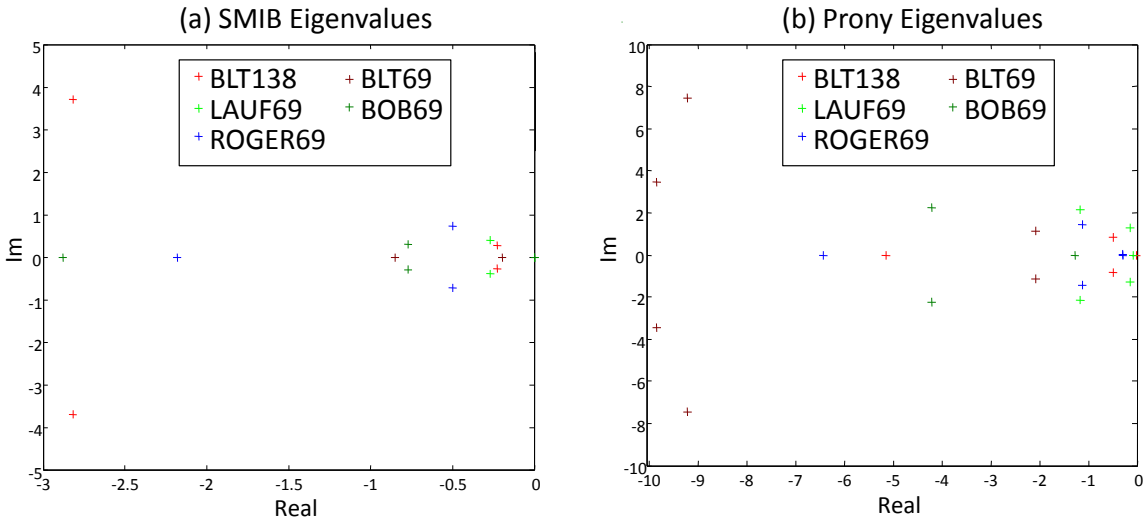

Figure 68. (a) SMIB eigenvalues for generators, (b) Prony analysis of events at generators

By inspection of Figure 68, there is some similarity in structure between SMIB eigenvalues and Prony analysis modal estimates, but more analysis is needed. Initially, we note that the SMIB eigenvalues do exhibit a similar pattern to the estimated Prony modes, but the scaling is substantially different.

Below, we compare the ringdowns for events in the GSO case. The event at BLT138 from Figure 68b is used to illustrate the application of the mode-matching LS method to the GSO case. The SMIB eigenvalues are chosen as the specified modes, then the generators with electromechanical modes corresponding to the highest estimated amplitudes should be noted. The identified generator or generators may serve as an indication of the likely source of the event. Figure 69 shows the results from fitting 29 generator SMIB eigenvalues to the ringdown using mode-matching LS, and the original signal is also plotted.

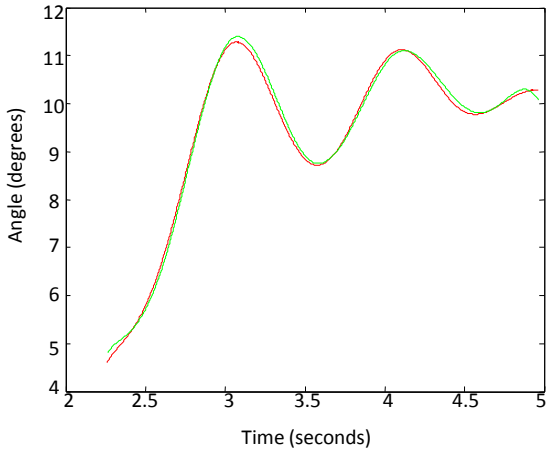

Figure 69. Reconstruction of event ringdown with mode-matching LS

For the ringdown in Figure 69, the SMIB modes with the highest amplitudes and their associated generators are identified. The generators with the largest contributions are identified in Table 29. As indicated in the table, the SMIB modes of generator BLT 138, which is where the event occurred, have higher amplitudes than the other SMIB modes. Thus, the precipitating generator is correctly identified as BLT138.

Table 29. Mode-matching LS results for estimating SMIB amplitudes

Generator	A_i	σ_i	ω_i
BLT138	17525.0000	-2.8161	3.6975
BLT138	17525.0000	-2.8161	-3.6975
ROGER69	9.9970	-1.9498	-12.603
ROGER69	9.9968	-1.9498	12.603
LAUF69	8.0001	-1.6795	14.944
LAUF69	7.9998	-1.6795	-14.944
BLT138	122.1300	-2.661	-15.716
BLT138	122.1300	-2.661	15.716

It is important to note that identification of a single contributing generator will not always be this apparent. Which modes are present in the ringdown depends on which modes are excited, and this is based on both the system input and system state. Further investigation on more general cases is warranted.

A useful result regarding the SM method is apparent from this application, as discussed in [167]. It is found that the SM approximation, when the initial transfer function poles are chosen, more

closely matches the true signal than the Prony approximation. Thus, a key insight is to use the SM approach with specified initial parameters, since a noteworthy improvement in the results is observed. When fewer modes are estimated, the discrepancy between the Prony and the SM methods is more obvious. For a 0.1 second balanced three-phase to ground fault at BLT138 in the two-bus equivalent, the bus angle ringdowns with both methods approximating only four modes are shown in Figure 70. The approximations for 30 modes are shown in Figure 71.

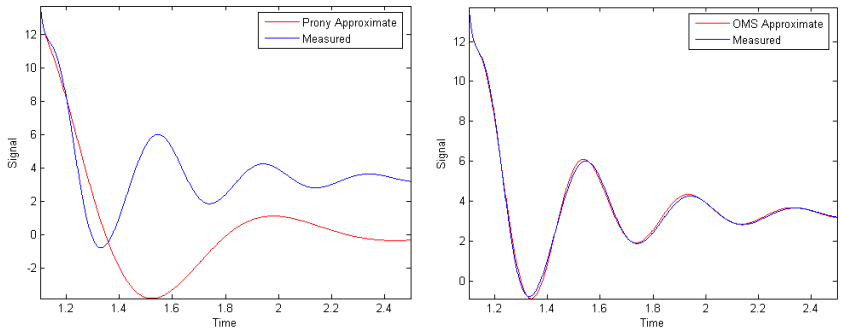

Figure 70. Four-mode estimate for BLT138 fault in two-bus equivalent – Prony method (left), SM method (right)

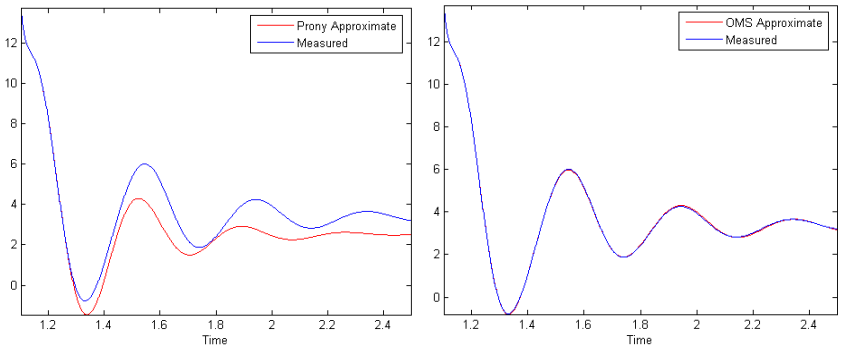

Figure 71. 30-mode estimate for BLT 138 fault in two-bus equivalent – Prony method (left), SM method (right)

Interestingly, when no initial parameters for the SM method are specified, its results essentially match the Prony method. Suggestions for choices of initial parameters include the largest-amplitude eigenvalues from an initial Prony analysis and the SMIB eigenvalues. This effect is observed generally, regardless of which ringdown signals are studied.

To conclude the section, a comparison is given for ringdowns at all generators in the 37-bus case due to the same type of event. In particular, three-phase to ground faults lasting 0.1 seconds are applied at each of the eight generator buses. Table 30 provides a summary of results from the three modal estimation algorithms for each event. The goal is to be able to pick out characteristics that make each event appear distinct.

Table 30. Modal estimate comparisons for events at each generator

(1) Gen 53 – "BLT138," s = -2.66±j15.7162		
Prony	**SB**	**LS**
s =-3.78±j10.92 *(d: 4.92, 4.80)*	s = -3.51±j16.63, *(d:1.25, -0.91)*	s = -3.78±j10.92, *(d:4.92, 4.80)*
s =-3.51±j17.13, *(d:1.65, -1.41)*	s = -2.97±j10.82, *(d: 4.91, 4.90)*	s = -3.51±j17.13, *(d:1.65, -1.41)*
(2) Gen 44 – "LAUF69," s = -1.68±j14.94		
Prony	**SB**	**LS**
s =-6.02±j9.51 *(d:6.95, 5.44)*	s =-7.06±j14.97, *(d:5.38, -0.03)*	s =-6.02±j9.51, *(d:6.95, 5.44)*
s =-2.10±j16.15, *(d:1.29, -1.21)*	s =-2.13±j16.70, *(d:1.82, -1.76)*	s =-2.10±j16.15, *(d:1.29, -1.21)*
(3) Gen 50 – "ROGER69," s = -1.95±j12.60		
Prony	**SB**	**LS**
s =-2.674±j12.221 *(d:0.82, 0.38)*	s =-2.862±j11.95 *(d:1.12, 0.65)*	s =-2.674±j12.221 *(d:0.82, 0.38)*
(4) Gen 54- "BLT69," s = -1.90±j11.28		
Prony	**SB**	**LS**
s =-2.09±j11.09 *(d:0.27, 0.19)*	s =-1.32±j11.20 *(d:0.59, 0.08)*	s =-2.09±j11.09 *(d:0.27, 0.19)*
		s =-1.90±j11.279 *(d:0.00, 0.00)*
(5) Gen 31- "SLACK345," s = -3.68±j8.17		
Prony	**SB**	**LS**
s =-2.86±j12.07 *(d:3.98, -3.90)*	s =-2.94±j10.79 *(d:2.72, -2.62)*	s =-3.683±j8.171 *(d:0.00, 0.00)*
s =-3.52±j14.45 *(d:6.28, -6.28)*	s =-1.91±j13.9 *(d:6.00, -5.73)*	
(6) Gen 28 – "JO345," s = -2.25±j14.69		
Prony	**SB**	**LS**
s =-2.362±j12.74 *(d:1.95, 1.95)*	s =-2.872±j11.49 *(d:3.26, 3.2)*	s =-3.517±j14.46 *(d:1.29, 0.23)*
s =-6.224±j16.63 *(d:4.42, -1.94)*	s =-2.09±j14.40 *(d:0.33, 0.29)*	s =-2.86±j12.07 *(d:2.69, 2.62)*
(7) Gen 14- "WEBER69," s = -3.75±j15.25		
Prony	**SB**	**LS**
s = -2.22±j13.30 *(d:2.48, 1.95)*	s = -2.13±j13.32 *(d:2.52, 1.93)*	s = -3.75±j15.25 *(d:0.00, 0.00)*
(8) Gen 48 – "BOB69," s = -3.37±j12.96		
Prony	**SB**	**LS**
s = -3.08±j11.34 *(d:1.65, 1.62)*	s = -3.32±j11.33 *(d:1.63, 1.63)*	s = -3.37±j12.96 *(d:0.00, 0.00)*

The results in Table 30 provide a comparison of estimated modal content of the measured signal with SMIB eigenvalues of known generators. The first row in each portion of Table 30 denotes the generator and its electromechanical SMIB modes.

An important challenge becomes apparent, which is how best to quantify similarity of modes. In these results, we give both the Euclidean distance and the Manhattan distance in the $j\omega$ direction. Beneath each estimated mode in Table 30 are two numbers, indicated in the *(d:a,b)* term, where *a* is the Euclidean distance and *b* is the Manhattan distance of each mode to the SMIB mode. Low distance values indicate an estimate which is well-matched to the SMIB electromechanical mode. The Manhattan distance shown is the SMIB eigenvalue frequency minus the estimated frequency, in radians. The reason for providing this metric is that the value of the frequency term may be of greater significance than the value of the damping term. Frequency is based on physical characteristics of the generator, whereas damping may more readily change based on controls in the system. More research is needed to better quantify the "closeness" of modes, or how best to compare similarity of modal content.

6.4. Automated Event Recognition

The OMS pattern recognition scheme is further developed in this section, and results of its application to a large system are provided. An important aspect of the problem is addressed, namely, the automatic classification of the ringdowns. The event recognition approach is formulated to facilitate its automatic implementation and the expansion of its functionality. In Section 6.4.1, the construction of mode maps for automatic ringdown analysis is presented. The approach requires the "similarity" of modes to be defined. Specification of this distance function is discussed in Section 6.4.2. Then, Section 6.4.3 demonstrates the success of the approach for oscillations originating at generators in the western interconnect (WECC).

It is important to point out that the approach is developed to work irrespective of having knowledge of the network or measurements. Of course, having any of these obviously makes the job easier. While in a substation or a control room, under normal circumstances, everything is mostly known about the system or can at least be queried, and it is understood that this

immediate application of identifying fault location is probably not necessary. Generally, there are much simpler ways of detecting such an event; for example, a relay will simply report that a three-phase fault occurred. If one has access to this status information, there is no need for complicated oscillation monitoring event detection. Where this application may be particularly valuable are in situations where little or nothing is known about the underlying system. This is similar to the model-free assumptions of the PMU-based equivalencing application Chapter 4. Model-free applications are designed to work when all that is available may be a spreadsheet of data. This may happen, for example, if one is recording oscillations from a wall outlet. These techniques may also be useful for post-event reconstruction; that is, after a significant disturbance. Model-free tools are useful for determining what may have occurred, without needing to piece together information from many different devices and sources. One may be able to discern significant identifying information of the event just from one device.

6.4.1. Generator Mode Maps for Ringdown Comparison

The goal of the decision-making process is to assign a given ringdown to one of *m* categories based on a vector *d* of features. Here, the *m* categories or classes are the generators, and the feature vector *d* is from the OMS data (frequency, damping, etc). Results in [167], highlighted in the previous section, show the effectiveness of three methods at finding the dominant modes. However, in Section 6.3, the responsibility to make a final conclusion that the ringdown is at a particular generator is ultimately left to the discerning eye of a viewer examining the results. Automatic analysis is desired instead. This motivates the need for an appropriate metric to quantify "closeness" of ringdown results to SMIB eigenvalues. Especially when considering large systems, a quantifiable way of interpreting the results is needed. An approach to interpreting the modal estimates is presented here which is designed to handle the fact that there is some uncertainty in the eigenvalues. In Section 6.4.3, this is demonstrated for events in the WECC.

The ringdown comparison approach presented in this section involves constructing *mode maps* for each generator. In general, this mode map should be representative of a particular generator or event. In this case, the mode maps are obtained by considering generator SMIB eigenvalues. Each point in the space of an image can be considered to receive some contribution from each of

the eigenvalues. That is, a set E of eigenvalues for a particular generator is positioned at fixed points in the complex plane. Then, there is some uncertainty surrounding these eigenvalues. As a point becomes farther away from the points in E, the probability that it corresponds to that generator decreases to zero. A function may be used to describe this uncertainty, but finding the most appropriate function may be a challenge. In summary, the purpose of a mode map is to define the expectation that any point in the complex plane belongs to a particular generator or event.

Additionally, the SMIB eigenvalue mode maps provide a 2D field of scalar values which can be mapped to colors and visualized, thus taking advantage of the extraordinary capability of the human visual system for pattern recognition. The value at each 2D coordinate is a weight reflecting the likelihood that a particular point belongs to that generator. These mode maps or *modal images* form the *knowledge base*. Once modal images are constructed, they are available for easy comparison with measured ringdowns.

The mode maps described above are calculated offline. Then, when a new ringdown is encountered, the likelihood of the instance belonging to each of the predefined classes needs to be evaluated. Then, the instance may be assigned to the class, or generator, with the highest likelihood. Consider events which occur at different generators, represented by the classes G_i. The Prony ringdown data R is the "evidence." The problem can be considered probabilistically [174],

$$\Pr\left(G_i \mid R\right) = \frac{\Pr\left(R \cap G_i\right)}{\Pr\left(R\right)} \tag{215}$$

where the intersection of R and G_i is the joint probability, and is determined from the modal image. Rather than deal with probabilities, we deal with weights,

$$s \cdot \Pr\left(R \cap G_i\right) = \sum_k image_{G_i}\left(\sigma_k, f_k\right) \tag{216}$$

where s is a scale factor, and $image_{Gi}(\sigma_k,f_k)$ denotes a point on the modal image G_i for each of the k modes. If it is assumed that all generators are equally probable, the scale factor, s, can be neglected.

As indicated in (216), the estimated modal components give coordinates in the generator modal image, and the corresponding retrieved values are then compared to classify the event. The event is thus assigned some likelihood of belonging to each class.

It is important to continue this work to develop classification methods for power system event data so we can better facilitate the automatic recognition of events. The ability of the analysis to occur *automatically* is emphasized. In this work, the classifier is described by the mode maps. Then, similarity of modal characteristics is evaluated by comparing the mode measurements in a ringdown response to mode maps of possible contributing generators.

6.4.2. Choice of Distance Function

For the mode maps used in this chapter, the value at any point is based on the distance of that point to the SMIB eigenvalues. However, there are potentially an infinite number of distance function choices that could be defined. Comparing similarity of modal content requires specification of this distance function.

In examining the choice of a distance function, we note that the problem of representing the extent to which the true modes are located near the SMIB eigenvalues lends itself naturally to graphics. Thus, in this work, the weighting function used is one which is adapted from Shepherd's method [175]. Sheppard's method interpolates points on a grid and is also used in power system visualization to create color contours (e.g. voltage) on one-line diagrams [176]. The value at any point on the 2D plane is a weighed combination of the values at points with specified values:

$$image_{G_i}\left(\sigma_k, f_k\right) = \frac{\sum_j w_j f_j}{\sum_j w_j} \tag{217}$$

In this work, the points with specified values are the generator eigenvalues. The modal images show the interpolation of the likelihood of these eigenvalues. The weighting function used in (217) is the following inverse distance function:

$$w_i = \frac{A}{\sqrt{\alpha\left(\sigma_{\text{mode}} - \sigma_{\text{current}}\right)^2 + \beta\left(f_{\text{mode}} - f_{\text{current}}\right)^2}} \tag{218}$$

where α and β reflect the fact that the damping and the frequency components may be assigned a different level of importance.

This distance function choice is made to illustrate the feasibility of the approach, and it appears to represent the problem well. Many other functions are possible, as discussed in [177], so this is not a unique choice. This distance function does reflect the uncertainty of eigenvalues, which is an important feature. A clear opportunity is present for further research to develop a more general or optimal way of constructing modal images. The distance function can be considered as part of the *knowledge base*, and new information can be incorporated to enhance its accuracy.

A difficulty in designing applications is that some assumptions must be made concerning what information or starting knowledge one may already have, and the available information will obviously influence the course of action one takes. There is a wide spectrum of starting knowledge one might possess when performing this analysis. Thus, it is worth emphasizing that one must be able to adapt this approach based on the amount and extent of available starting information. The mode maps are one place where extra information can be incorporated, and the distance function choice is another. For example, it will be useful to collect and incorporate information about how modes move in the complex plane and show this in the modal images. Once modal images are constructed, the possible origin of a measured ringdown can then be estimated.

6.4.3. Results for WECC Ringdowns

A system representing the WECC is now examined. The WECC system has approximately 3300 generators and 8200 loads. There are numerous dynamic models, and each dynamic model potentially contains a large number of states and parameters. In this section, we examine simulated events in the WECC system and demonstrate the use of our approach on the resulting ringdowns. The angle reference in the simulation is the average of all generator angles. Then, in the next sub-section, we discuss these results and propose useful extensions.

Four WECC generators are studied for this example. To construct the modal images, the SMIB eigenvalues of the generators are required. These are plotted in Figure 72. The horizontal axis is damping and the vertical axis is frequency in Hz. Each symbol denotes a different generator. From Figure 72, one can observe that certain eigenvalues are not distinguishable from those of other generators. Fortunately, however, the pattern of the eigenvalues for each generator is still distinct. The question is whether these distinct patterns can be captured and recognized in the ringdowns.

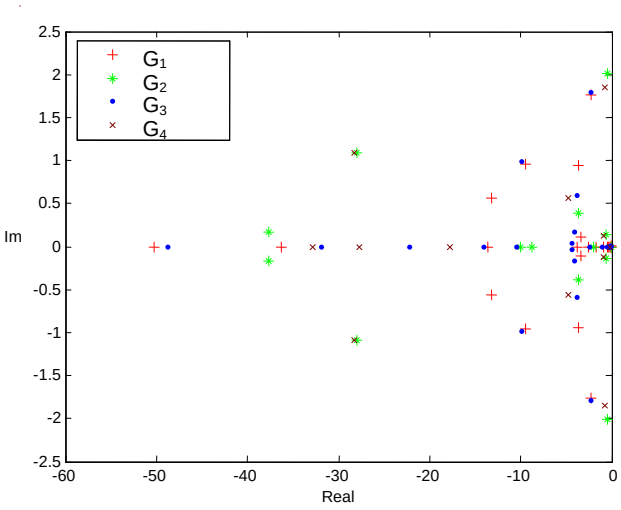

Figure 72. West SMIB eigenvalues

The model types of these four generators are given in Table 31. The ringdowns for three-phase-to-ground faults at generators G_1, G_2, G_3, and G_4 are shown in Figure 73, and they are labeled respectively R_1, R_2, R_3, and R_4. The modal images constructed for each generator are shown in Figure 74.

Table 31. Machine model types

	Machine	Exciter	Governor	Stabilizer	Other
G_1	GENROU	TEXS	GGOV1	PSS2A	OEL1
G_2	GENTPF	EXAC3A	IEEEG1	WSCCST	OEL1
G_3	GENROU	EXAC3A	-	IEEEST	OEL1
G_4	GENROU	TEXS	GGOV1	PSS2A	OEL1

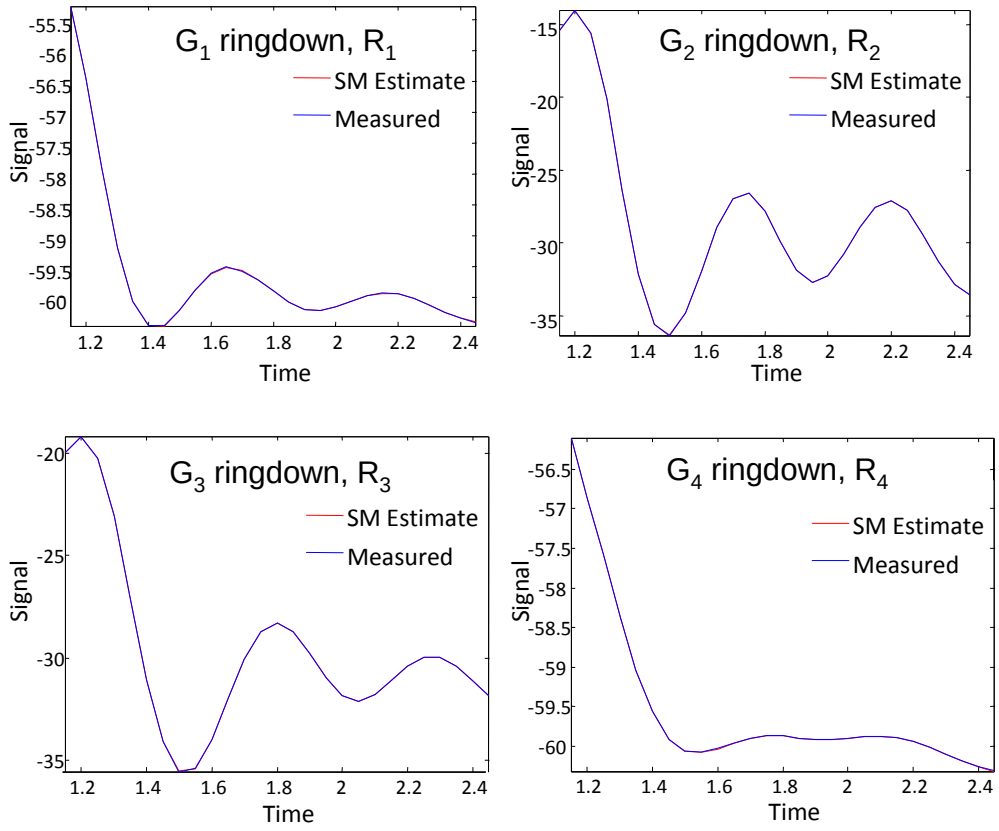

Figure 73. (Top left, top right, bottom left, bottom right) Measured ringdowns R_1, R_2, R_3, and R_4, shown with corresponding SM approximations, respectively

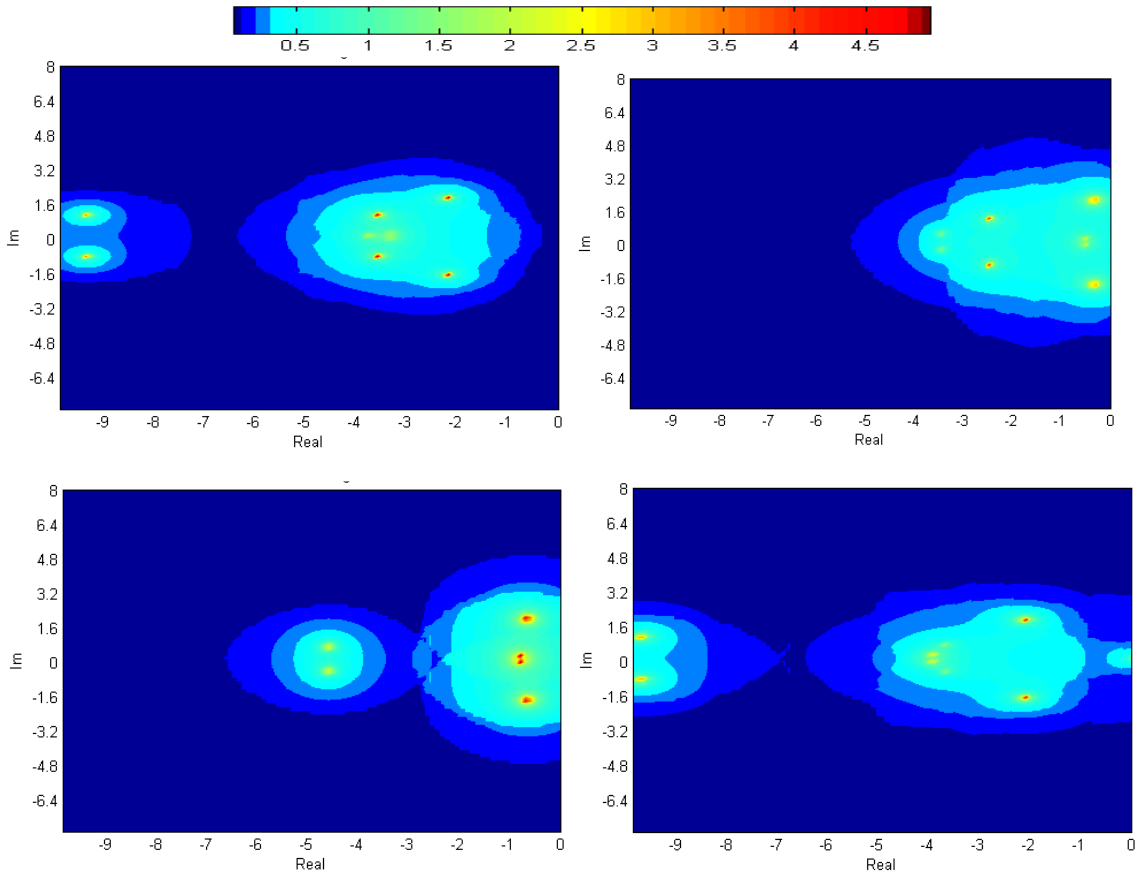

Figure 74. (Top left, top right, bottom left, bottom right) Modal images for G_1, G_2, G_3, and G_4, respectively

Results showing the estimated modes for each ringdown of Figure 73 are given in Table 32 through Table 35, obtained by applying the SM method with initial parameters, as previously described. Results and experience indicate that the electromechanical modes of the generator with the fault tend to be prominent in the ringdown for this type of event. Electromechanical modes, typically in the 1-3 Hz range, are associated with machine rotor angle and speed. The electromechanical modes can be identified using participation factor analysis [160]. Let participation factors $p_{\delta,\omega}$ relate how rotor angle and speed impact the mode. Then, the electromechanical modes of the four generators and the corresponding participation factors are given by the following:

$$G_1: \quad \lambda_{EM} = -2.304 \pm j11.042 \qquad p_{\delta,\omega} = 0.66, 0.61 \tag{219}$$

$$G_2: \quad \lambda_{EM} = -0.439 \pm j12.595 \qquad p_{\delta,\omega} = 0.69, 0.72 \tag{220}$$

$$G_2: \quad \lambda_{EM} = -28.041 \pm j6.846 \qquad p_{\delta,\omega} = 0.13 \tag{221}$$

$$G_3: \quad \lambda_{EM} = -0.765 \pm j11.656 \qquad p_{\delta,\omega} = 0.67, 0.74 \tag{222}$$

$$G_3: \quad \lambda_{EM} = -28.237 \pm j6.876 \qquad p_{\delta,\omega} = 0.10 \tag{223}$$

$$G_4: \quad \lambda_{EM} = -2.262 \pm j11.353 \qquad p_{\delta,\omega} = 0.65, 0.53 \tag{224}$$

As (221) and (223) indicate, participation factors for generators 2 and 3 indicate the presence of a second mode which contributes to the speed state variable. However, its participation factor is still significantly less than that of the other mode.

Table 32. G_1 ringdown, SM method results

A_i	σ_i	ω_i	f_i	ϕ_i
59.538	0.009333	0	0	3.1416
0.001071	-0.63631	38.171	6.0751	1.8971
0.001071	-0.63631	-38.171	-6.0751	-1.8971
0.36997	-1.3121	11.41	1.816	0.64625
0.36997	-1.3121	-11.41	-1.816	-0.64625
0.009564	-1.5249	28.16	4.4817	-0.07392
0.009564	-1.5249	-28.16	-4.4817	0.073921
1.967	-6.1478	9.3534	1.4886	-0.41702
1.967	-6.1478	-9.3534	-1.4886	0.41702

Table 33. G_2 ringdown, SM method results

A_i	σ_i	ω_i	f_i	ϕ_i
27.808	0.085692	0	0	3.1416
3.6972	-0.90208	12.869	2.0481	-1.1021
3.6972	-0.90208	-12.869	-2.0481	1.1021
4.7489	-2.5911	6.4759	1.0307	0.04304
4.7489	-2.5911	-6.4759	-1.0307	-0.04304
0.19369	-2.6439	24.789	3.9453	2.9495
0.19369	-2.6439	-24.789	-3.9453	-2.9495
0.034903	-3.5326	36.608	5.8263	0.18261
0.034903	-3.5326	-36.608	-5.8263	-0.18261

Table 34. G_3 ringdown, SM method results

A_i	σ_i	ω_i	f_i	ϕ_i
30.207	0.022992	0	0	3.1416
1.155	-0.40206	12.297	1.957	-1.258
1.155	-0.40206	-12.297	-1.957	1.258
0.037089	-1.2903	23.355	3.7171	2.3166
0.037089	-1.2903	-23.355	-3.7171	-2.3166
5.4866	-3.0222	9.1124	1.4503	-0.67972
5.4866	-3.0222	-9.1124	-1.4503	0.67972
0.099438	-7.4538	44.578	7.0949	-2.0917
0.099438	-7.4538	-44.578	-7.0949	2.0917

Table 35. G_4 ringdown, SM method results

A_i	σ_i	ω_i	f_i	ϕ_i
59.093	0.015505	0	0	3.1416
0.153	-1.7292	13.083	2.0823	-1.0678
0.153	-1.7292	-13.083	-2.0823	1.0678
1.3745	-3.4335	5.4603	0.86903	0.30096
1.3745	-3.4335	-5.4603	-0.86903	-0.30096
0.045315	-4.1202	26.548	4.2252	1.702
0.045315	-4.1202	-26.548	-4.2252	-1.702
0.15645	-10.379	27.326	4.3491	-1.9781
0.15645	-10.379	-27.326	-4.3491	1.9781

When examining the above tables, the electromechanical modes are clearly present but are not an exact match. Other modes are also present in the signals. Already an exclusivity problem begins to emerge: the electromechanical modes may not be sufficiently distinct to allow one to confidently distinguish between the generators based on ringdown data.

Using the proposed modal image comparison approach, it is possible to take advantage of constructed modal images to perform comparisons quickly and automatically. The modal image approach extends the manual electromechanical mode-matching concept to a more general procedure. Modal images may include eigenvalues, uncertainty, and expected eigenvalue movement. The automatic event recognition approach is applied with the modal images in Figure 74. In Table 36, each column represents one ringdown. For each ringdown, the value in (216) is computed from each of the modal images in Figure 74.

Table 36. WECC modal image matching results

	R_1	R_2	R_3	R_4
G_1	1.2902	1.7164	1.9082	1.5172
G_2	0.81342	1.7525	1.2115	1.9025
G_3	1.2566	1.3864	1.9564	1.5296
G_4	1.1335	0.44082	1.8571	1.9862

Each ringdown is correctly classified to its originating generator, as shown by the highlighted values in Table 36. As Table 36 indicates, there is a distinction between the ringdowns in a way that is analyzed and related back to the key contributing generators.

From analysis of this system, it is apparent that ringdowns can exhibit relatively small differences, and these need to be accentuated. Generators often have similar models and modes, so it becomes necessary to sharpen their distinguishing features. Even small differences can provide important identifying information.

6.5. Generator Exciter Event Classification

In addition to considering events which are generator faults, other event types may also be studied based on the same approach previously described. In particular, in this section we consider events caused by a change in exciter setpoint. The reason for choosing this type of event to study is that while the fault event recognition approach is illustrative of the basic concepts of automatic pattern recognition encouraged in this thesis, there are a number of reasons why a fault might not be the most "interesting" type of event to study using this method. As alluded to previously, determining the existence and location of faults is typically accomplished quite readily using existing methods and technology; hence, using the proposed methods specifically for event diagnostics consisting of identifying generator involvement in faults has limited applicability. Additionally, faults are physically occurring phenomena, and their effects cannot (easily or directly) be caused by human actions. Comparatively, there are a number of other event types in power systems which one can conceive of as being a result of something other than the normal operation of the power system. For example, an exciter setpoint can be changed simply via communications. Generally, when it is changed, there is a reason or a need for the change. In fact, a change in exciter setpoint that is *not* intended would be cause for further investigation and possibly alarm. In such a case, using the described methodology to

accomplish the identification of a signature indicative of an exciter setpoint change event would be advantageous. The SMIB eigenvalues of the generator used in this section are given in Table 37.

Table 37. Exciter example, generator SMIB eigenvalues

A_i	σ_i	ω_i	f_i
19.0621	-19.0621	0	0
11.1046	-0.4889	11.0939	1.7656
11.1046	-0.4889	-11.0939	-1.7656
4.2084	-4.2084	0	0
0.906	-0.6059	0.6736	0.1072
0.906	-0.6059	-0.6736	-0.1072

A few differences between the event types are noted. The time constant involved for the exciter events is slower. To view the entire ringdown, one must look at a window of approximately 50 seconds, compared to a window of 10 seconds for the faults. If too short of a time window is used, the information captured will be inadequate. Based on the ringdowns over 10 seconds, high-frequency oscillations appear to be visible in the ringdown of the bus angle, as shown in Figure 75. These plots show respectively the rotor angle, the voltage magnitude at the bus, and the voltage angle at the bus.

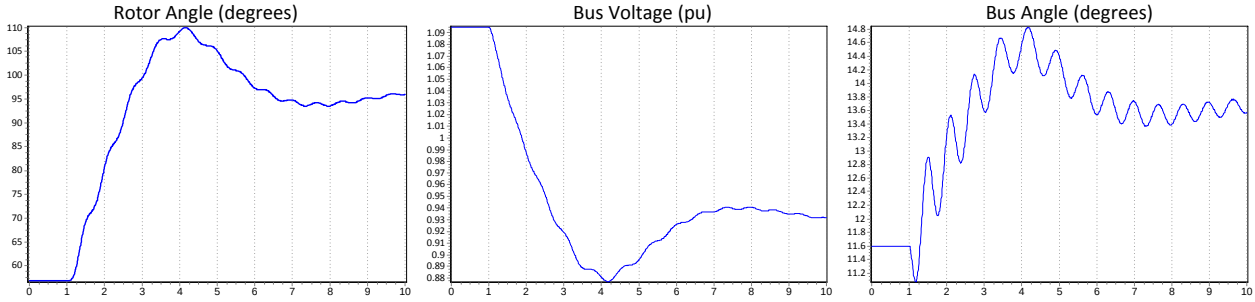

Figure 75. Rotor angle, bus voltage, and bus angle, due to exciter event, 10-s time window

Observing the same event over 50 seconds in Figure 76 reveals that indeed there is a high-frequency component present in the bus angle signal.

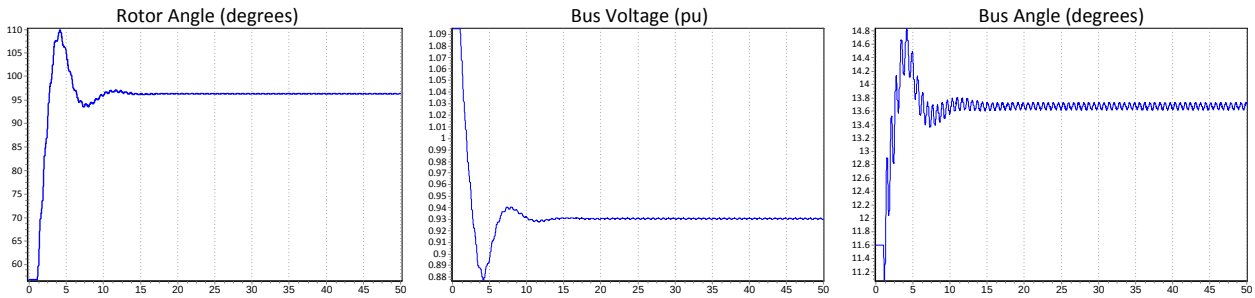

Figure 76. Rotor angle, bus voltage, and bus angle, due to exciter event, 50-s time window

It is not known what is causing this oscillation. The high frequency oscillations can present a difficulty for the modal estimation algorithms to approximate the signal. One possible solution is to apply a low-pass filter first, and experimentation shows that a low-pass filter does indeed help remedy this problem. If we first apply a low-pass filter to the signal, it is now possible to approximate the bus angle ringdown using a lower order (six instead of 20 or more). Figure 77 shows the original and the filtered signal, estimated using the Steiglitz Mc-Bride algorithm.

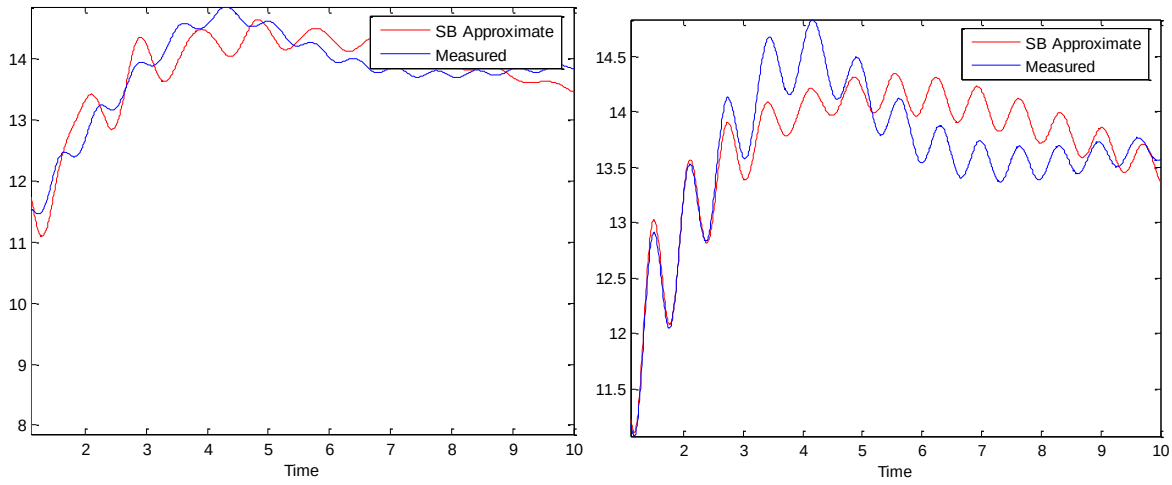

Figure 77. Bus angle ringdown with (left) and without (right) low-pass filtering

Table 38 shows the estimation, with the low-pass filter applied to the original signal first.

Table 38. Bus angle signal, modal estimation, after low-pass filter

A_i	σ_i	ω_i	f_i	ϕ_i
16.208	-0.01885	0	0	0
0.37154	-0.27821	6.0087	0.95631	0.91344
0.37154	-0.27821	-6.0087	-0.95631	-0.91344
4.7636	-0.49225	0	0	3.1416
0.23662	-0.63961	8.9137	1.4187	1.8014
0.23662	-0.63961	-8.9137	-1.4187	-1.8014

Comparatively, with no low-pass filtering, the signals are estimated according to Table 39. As evident in the table, the low-pass filtering leads to a better identification of the key modes:

Table 39. Bus angle signal, modal estimation, without low-pass filter

A_i	σ_i	ω_i	f_i	ϕ_i
15.863	-0.01642	0.000553	8.79E-05	-0.00582
0.10367	-0.02581	8.34	1.3274	-2.2586
0.10502	-0.02697	-8.3451	-1.3282	2.2787
4.0909	-0.43738	-0.0124	-0.00197	3.12
0.44023	-0.85646	-9.7346	-1.5493	-2.0023
0.44481	-0.85948	9.7248	1.5477	2.0054

The ringdown results for SM method for the voltage magnitude signal are given in Table 40. This oscillation is a much lower frequency than those from the three-phase faults looked at earlier.

Table 40. Exciter example, voltage signal, modal estimates

A_i	σ_i	ω_i	f_i	ϕ_i
0.93779	-0.00104	-8.26E-07	-1.31E-07	8.60E-07
0.076554	-0.33436	0.72877	0.11599	0.3879
0.076554	-0.33436	-0.72877	-0.11599	-0.38791
0.003219	-0.55351	-8.6731	-1.3804	0.50131
0.003219	-0.55351	8.6731	1.3804	-0.50131
0.001996	-3.3483	-10.63	-1.6918	-1.5171
0.001996	-3.3483	10.63	1.6918	1.5171

Comparing Table 40 with Table 37, the same dominant modes are present. These values are highlighted in the respective tables. Interestingly, rather than the response containing one mode in the 1-2 Hz range, it contains two, one at 1.38 Hz and another at 1.69 Hz. The dampings of these two modes do not match the damping of the 1.76 Hz SMIB eigenvalue.

Further investigation is made into cause of the high-frequency oscillations. The above examples correspond to a generator with an exciter of type SEXS_PTI illustrated in Figure 78:

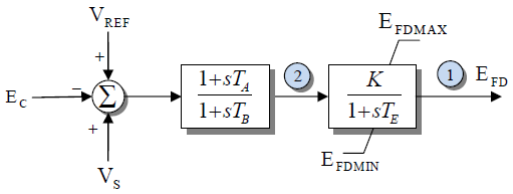

Figure 78. Original exciter model - SEXS_PTI

Then, this exciter was changed to IEEET1, as shown in Figure 66. The default parameters were used. The change in exciter type impacted the ringdowns, as we can see. This was done to try to see if the same high-frequency oscillation is observed with other exciters. With no other changes in the system except this exciter model, the ringdowns are given in Figure 79:

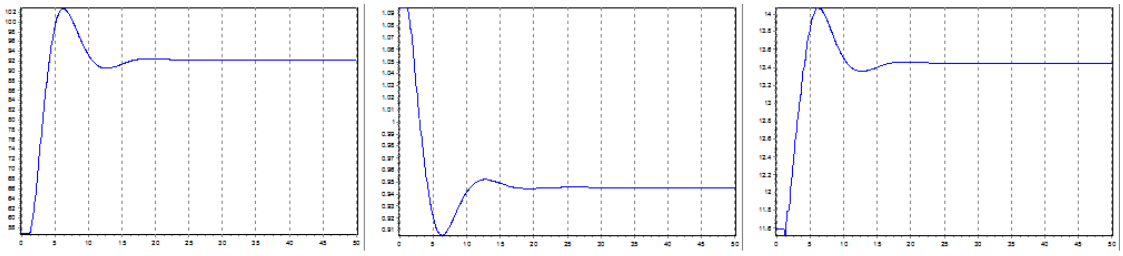

Figure 79. Rotor angle, bus voltage, and bus angle, due to exciter event, 50-s time window, different exciter mdoel

Interestingly, the high frequency oscillations observed in the bus angle signal earlier are not present when this exciter is used instead.

As can be seen in the preceding examples, a distinct ringdown signature is present for exciter setpoint modifications, which is related to the exciter SMIB eigenvalues of the precipitating generator. As described previously, once such signatures have been identified, the part of the work which is unique to each event type is essentially complete. What remains once a characterization of the event is obtained is to develop the online tools needed to perform this analysis automatically and quickly on incoming ringdowns. The previous section described how

this can be done using mode maps. For this type of event, rather than mode maps containing a heavy weighting or representation of generator electromechanical modes, it should contain the modes associated with the exciter. As we have seen, these tend to be lower in frequency and so require a larger time window to capture.

We have also seen that the use of filters can be a beneficial step to ensuring we do not obtain "noise" in our results, where "noise" here refers to signals outside of our range of interest. In fact, the noise itself may be indicative of a problem or an event. High frequency oscillations indicate that something may be switching quickly – so perhaps that exciter model was experiencing problems due to its output hitting a limit and then possibly reacting and hitting a different limit. There is further evidence in this application of the value of data for diagnostics. In addition to looking for the presence of certain events (event signatures) there is also the possibility of inducing an event and *not* seeing the presence of the expected signature, as occurred in this scenario. So, while this can be used to see if an exciter setpoint has changed without our knowledge or our authorization, it can also be used as a verification tool. While verifying what we expect, we can simultaneously increase our knowledge base (and mode maps) and improve the accuracy (or at least the trustworthiness) of the tool.

6.6. Conclusions

In this chapter, estimated modal content from oscillations is used to identify event signatures. The ability to accurately estimate modes is a valuable starting point for the further study of oscillations based on data, as described in Section 6.1. Section 6.2 examines the relationship between events in the system and the measured modal content of the oscillations. The links between system identification and data mining become apparent when analyzing oscillations.

Our results indicate that events impact modal estimates in distinct, predicable ways. With knowledge of the nominal modes of certain generators, it is often possible to distinguish an event at one of these generators. Regardless of which modal estimation algorithm is used, the feasibility of distinguishing different events from each other in the resulting ringdown data is clearly illustrated. While we conclude that model types and parameters have a distinguishable effect on modes, it is not an easy concept to quantify. Participation factors help by describing

how state variables relate to the eigenvalues, and we can quantify how certain modes may be expected in measurements.

A useful result regarding the SM method is obtained from this chapter. It is found that with the SM approximation, when the initial transfer function poles are chosen, a noteworthy improvement in results is observed when compared to a Prony approximation of the same order.

An important aspect of the problem is the automatic classification of the ringdowns, which is also addressed. Automated processing of modal estimates from OMS results can facilitate online decision making. An approach is introduced in Sections 6.3 and 6.4 for the systematic and automatic classification of ringdowns to events based on modal content. Modal images for each generator are constructed, where the values in a modal image reflect the likelihood associated with a mode belonging to a particular generator. These modal images are useful for comparing event ringdowns and can be generalized to use any choice of distance function. The method is demonstrated using examples from a WECC case, and results show that the ringdowns can be correctly classified. The ringdown comparison approach involves constructing *mode maps* or modal images for each generator, representative of a particular generator or event. Modal images are an attempt to deal with the problem of ringdown comparisons while accounting for the fact that modal estimates have some intrinsic uncertainty and error. A potentially rich opportunity is present for further research on a more general and optimal way of constructing these modal images.

By building up a knowledge base of these mode maps, a framework is in place to further discover and quantify associations between measured data and events. There is a wide spectrum of starting knowledge one might possess when performing this analysis. The framework is extensible to incorporate other events and various levels of known information. The available information will obviously influence the course of action one takes. In the event that *no* information is known, it can still be possible to *build* a knowledge base from actual measured ringdown data. Extra information can be incorporated into the mode maps as well as into the distance function choice. For example, it will be useful to collect and incorporate information about how modes move in the complex plane. The framework can be extended to other events

such as exciter setpoint change, considered in Section 6.5, and is a forerunner to performing more in-depth analysis based on event patterns or signatures in estimated modal content. This chapter highlights the fact that it is possible to build upon this work to ultimately be able to identify a wide variety of signatures for different types of events.

In conclusion, measurement data can be used to characterize relationships and facilitate event detection and classification. A relationship exists between the system eigenvalues, which may be associated with certain generators, and the modes in a ringdown. Examining the contributions of generator electromechanical modes to ringdowns lays the foundation for continued work in detecting and distinguishing event characteristics. Continuing work is needed to develop methods for signature recognition and classification in OMS data. The applications in this chapter are supplemental to existing OMS analysis, leading towards an improved and automated root-cause analysis of oscillations.

There is opportunity for further investigation into a number of questions, including the following: What circumstances are necessary to assure that certain events will present distinct signatures? How can we obtain more information by simultaneously considering the steady-state changes that occur?

A key challenge which is fundamental for this work is to develop metrics for comparing similarity of signals based on their modal content. A more basic question is how best to quantify similarity of modes. For example, the value of the frequency term may be of greater significance than the value of the damping term since frequency is based on physical characteristics of the generator whereas damping may more readily change based on controls in the system.

The work in this chapter may be additionally useful for aiding in transient stability response comparison, such as to validate dynamic characteristics of equivalents in Chapter 4. The applications in this chapter also provide further evidence of the value of data for diagnostics, examined in Chapter 5.

Another challenge is dealing with cases where the strong modes excited by an event are not easily identifiable as belonging to a particular generator's local modes. Rather than attempting to associate an event with one particular generator, it may be useful to determine sets of coherent generators, as in Chapter 4, and find the coherent group to which an event is likely to belong. Finding these groups or clusters can help make pattern-matching more tractable since rather than searching the large set of all generators, it is only necessary to search a small set of groups.

These results in this chapter are promising, especially considering the fact that the SMIB eigenvalues are essentially two orders of approximations; firstly from the non-linear system dynamics to a linearized representation of the system, and secondly from a linearized representation of the system to generator models which are decoupled from the dynamics of other generators. The connection between SMIB models and the full system is further examined in [161].

Difficulty arises from the fact that the concept of eigenvalues does not truly extend to nonlinear systems. The only exception is linearization around an equilibrium point. Thus, measured modes in power systems are not truly defined by the eigenvalues of linear system theory. Measured modes of the system are changing, especially during an event. It would be better to have an algorithm designed to account for the expected changes over time. Studying the trajectory of eigenvalues over a window of time may provide more information about the events. Modal sensitivities calculated from data [178], as part of the Modal Analysis for Grid Operation (MANGO) framework [179], provide insight into the relationships between measured modes and events or control actions in the system. Model types and parameters have a distinguishable effect on modes, but this is not an easy concept to quantify. Participation factors help by describing how state variables relate to the eigenvalues. Work remains to better represent the information contained in measured modes.

As a suggested extension, one may consider doing the reverse of the MANGO framework. That is, if modal sensitivities are known, perhaps from MANGO, and then one *observes* a mode changing in a certain way, the question is whether it is possible to go the other direction and map that mode change to possible injection changes. That is, it may be possible to determine what

generator output changes (or at least what *space* of generator changes) may have occurred to cause the change. In the MANGO framework, this amounts to needing the inverse of the modal sensitivity matrix and multiplying that matrix by the observed mode changes. Such an event detection application would be different from the current work since it is also makes use of steady state information.

There are other aspects and ongoing advancements in OMS technology which are beyond the scope of this work. Some issues include signal selection, filtering, model order, and linearity assessment [155], [156]. The number of modes, *n*, which should be estimated is often unknown. A primary limitation to obtaining accurate modal estimates is noise. Measurement location can also impact the estimates. It may be beneficial to consider the impact of location using relevant concepts in electromagnetic wave propagation theory [180], [181]. Signals at different locations can be correlated to reflect how events move through the system. Also, utilizing a multi-signal approach in [155] provides more accurate modal estimates in the presence of noise and system nonlinearities. Another useful area for further investigation is system identification. The ideal method would allow certain modes to be specified as algorithm parameters while allowing the remaining modes to be selected such that a best-fit of the signal is obtained by the algorithm. More research is needed to seek direct ways to accomplish this. Including such ongoing advances will further enhance this work.

7. Protecting Data and Applications

The focus in this chapter is on protecting the data itself as well as the applications which use the data. From this viewpoint, the goal is to make the system secure so that it will behave in the ways that are intended and not in the ways that are not intended. To leverage power system protection terminology, "dependability" and "security" must be balanced; this analogy is equally valid from an overall system perspective. There is a balance between acting every time that an action is required (dependable) and refraining from taking any actions that are unnecessary (secure). The protections of data and applications must also strive to reach this balance.

Accounting for the impact of false or "bad" data is crucial to defining the relationship between data and power systems. The goal of bad data protection is to improve the robustness or resilience of algorithms in the presence of wrong, malicious, or missing data. Data availability and integrity are critical for operational data. Sensors need to be protected from inappropriate access, both physical and remote. As new applications are created, they must be designed to operate safely and correctly, as much as possible, in the presence of any conceivable received data.

Given the criticality of the power grid, it can be an attractive attack target. Adversaries may attempt to manipulate sensor measurements, insert fake control commands, introduce delays into measurements and commands, and take other malicious actions. It is crucial to protect these systems against such malicious activity to ensure safe and reliable operation of the power grid.

Another component of protecting data and applications is the ability to detect and respond to problems caused by false data. This chapter also focuses on how to detect false or "bad" data in power system applications, while alerting operators to its presence and recommending actions, without disrupting the normal operation of the power grid. Regardless of how good the protections of a system are designed to be, they cannot be perfect; that is the emphasis of the second half of this chapter. As an engineer, one must walk the line between the role of designing applications which are maximally robust against false data and the role of identifying and handling false data in a system which necessarily has an imperfect defense.

The emphasis in the next two sections is on designing secure applications with respect to data quality. Some key considerations relating data to all the applications in this thesis are presented. Then, solutions along two different avenues of bad data detection and response are presented in the third and fourth sections. In Section 7.1, the framework detailed in [42] for intelligently coordinating and controlling devices is outlined. This framework was developed in parallel with the voltage control application in Chapter 3, in particular. In Section 7.2, the problem of protecting DC state estimation from false data injection attacks is solved by measurement selection. Then, the CPIDS framework we proposed in [182] is described in Section 7.3, which combines cyber and physical state information to provide a better awareness of the overall combined state of the system. Then, the last application, described in Section 7.4 and presented in [187], is a perturbation-based bad data detection system.

7.1. Intelligent Control and Communications Framework

An intelligent, secure communications and control framework is absolutely essential for the voltage control application in Chapter 3. Thus, it is imperative to mention some of its requirements here, as one cannot omit this discussion when describing control applications. With powerful wide-area control schemes comes the responsibility to understand and describe the requirements for a secure communications infrastructure to control these resources. Only then can a solution be carried out to satisfy these requirements, and only then can such controls, made available via smart-grid technologies, be actuated in the system.

In order to coordinate the control of the possibly vast amounts of distributed resources via the voltage control method, a communication structure which recognizes the authority and responsibility of the participating entities is required. An intelligent control framework is proposed in [41], [42] to coordinate many resources in order to achieve a common goal during an event. An intelligent control framework allows the control of end-user reactive-power-capable devices for mitigation of low voltage problems at a system-wide level. The framework involves requirements to control these resources over a secure communications infrastructure, outlined below. Such controls can be actuated in the system to maintain a desired voltage profile.

In [42], a communications and control hierarchy is proposed to address some of these concerns. This scheme is a part of the voltage control framework introduced in Chapter 3. The framework mimics a chain of command structure such as the Incident Command System (ICS), where members follow a line of authority and responsibility. Such an authority structure is the key idea of the proposed framework. The ICS is a "systematic tool used for the command, control, and coordination of an emergency response" [183]. This system is in widespread use by firefighters and other emergency personnel for efficiently handling the emergency scenarios they face daily. In particular, the ICS has proven that it is successful for dealing with emergencies and for coordinating large numbers of responders. Responders under its command may not all work together normally but have the same goals for the incident.

There are notable parallels between the ICS and the requirements of a framework needed for the intelligent control of power system devices. In particular, the reactive power control devices in Chapter 3 can be controlled to respond efficiently when the power system is in crisis. In the ICS, each individual reports to only one supervisor. The individuals work in groups, and the members report to a particular supervisor or officer who in turn reports to another designated officer. The functional unit with the highest authority is called *command*. Beneath command, there may exist different sections, branches, functional groups, and geographical divisions [183]. The resources which perform the task and have no responsibility to command other resources are at the lowest level in this structure.

The individual end-user reactive-power-capable devices in Chapter 3 are the resources, which are then aggregated. Similarly to the personnel resources in the ICS, while these end-user devices do not normally work together, they have the same goal in a crisis. For devices which are a part of the framework, by following a chain of command during an event, they are able to provide the help that is needed. As described in Chapter 3, reactive support groups [41] are a key enabler of the voltage control application. Since reactive power support is local in nature and must be generated close to its point of consumption, such reactive power support must come from within the region with the voltage problems. Each reactive support group would include an *officer*, who is responsible for that group. This officer may be a particular device within the group. Likewise, that officer reports to a *supervisor* who may be the central Energy Management System (EMS).

Following the same structure, the individual members of reactive support groups (devices at the transmission system buses) are lower-level officers responsible for assigning and coordinating the responses of those under their command. Ultimately the distribution system buses and end-user devices fall under the command of a transmission system level reactive support group member. Figure 80 illustrates this hierarchy in terms of power system components, where C0-C9 represent controllers which regulate the loads L0-L9. Such controllable loads may be able to provide services back to the grid during an event. The relays A-D control the delivery of power to the loads.

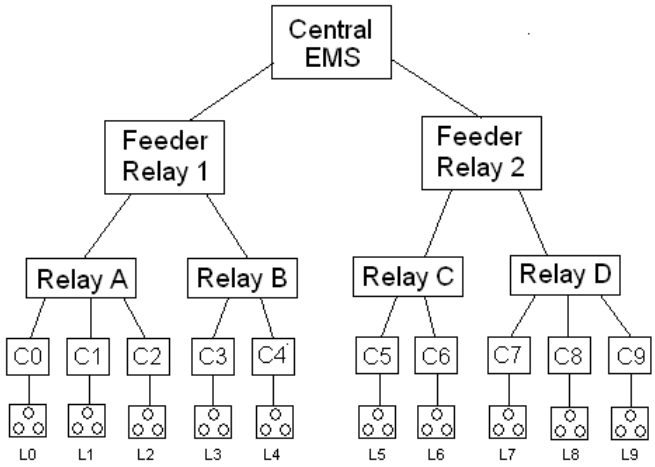

Figure 80. Transmission-distribution block diagram

Figure 80 shows officer-member groupings called *realms* which consist of a top layer and a bottom layer. Each device in the top layer of a realm can supervise and control the activity of a set of devices in its bottom layer. The top-level devices in each realm do not communicate directly with any devices lower in the hierarchy than the bottom-level devices in their realm. Instead, if control actions need to be taken further down the hierarchy, the bottom-level devices of the realm, which are also the top-level devices of the next lower realm, will send the appropriate control signals downstream. This pattern of delegation is at the heart of the ICS model, and it provides a convenient way to segregate and secure communications on the smart grid. Support regions can be used to establish webs of trust ahead of time, which will greatly simplify the task of distributing the keys to perform authentication.

To understand this idea of realms and layers, consider how a top-down detection-and-response pattern would work for the voltage control application of Chapter 3.

(1) The Central EMS detects a voltage problem somewhere on the system. It computes a response that would mitigate the problem, using the approach described in Chapter 3. It formulates action requests and sends them through the hierarchy where they are received by the feeder relays.
(2) Each feeder relay receives a reactive support request which originated from the EMS. The feeder relay agent computes a set of response actions that would allow it to fulfill the request, formulates the corresponding action requests, and sends them to the relays to which they are connected.
(3) Each relay receives the reactive support request from its feeder, computes a response action, formulates the corresponding action requests, and sends them to the load controllers it servers.
(4) Each controller then controls the loads under its supervision to meet the requests.

These communications each occur within a distinct realm. A device's realm and layer determine the actions and communications it is allowed to perform. Consider a simple example with only two realms, R_1 and R_2. The devices in these realms can be subdivided into three sets: A, B, and C (Figure 81).

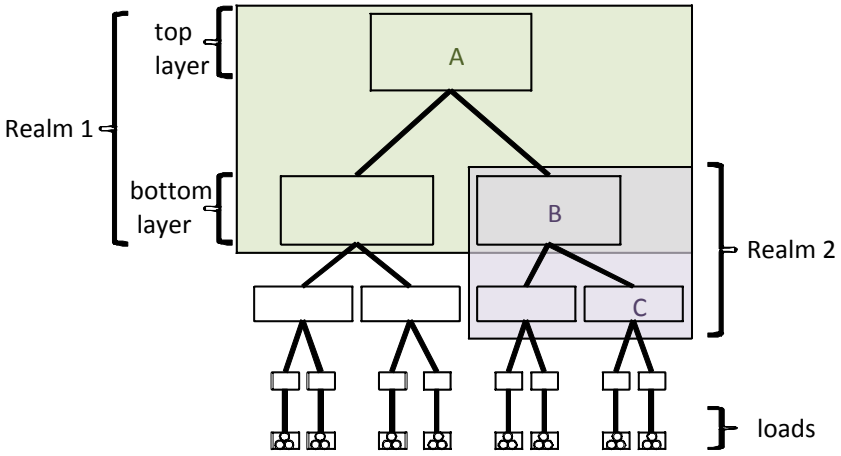

Figure 81. Example grouping of devices to help segregate and secure communications

In Figure 81, set A devices can communicate directly with set B devices. Likewise, set B devices can communicate directly with set C devices. However, set A devices cannot communicate directly with set C devices. Each realm is intended to have its own security parameters which it uses for communication exchanges. An exchange within this proposed secure communications infrastructure is detailed in an example in [41].

The number of devices involved in formulating and implementing a response action may be quite large. Their communications must also be immune to a variety of attacks. In considering the security of such communications, particularly important are authentication and the integrity of messages. These requirements are briefly described below; more detail is found in [41] and [42]. Specifically, an unauthorized device must not be able to send a control message to another device, and it must not be possible for a message to be replayed inappropriately. Any communication between devices must traverse unmodified from sender to receiver. If devices are not who they claim to be, or if devices receive messages that are not what the sending device intended, then the resulting action may actually worsen system conditions rather than improve them. Strategies for ensuring the authentication and integrity of the communications framework are critical to the overall success of such a wide-area measurement and control framework.

A description of the security requirements are found in [41] and [42]. While of critical importance for supporting the applications described, it is not part of the main focus of the thesis, so only a sketch is provided. Some of the key requirements are timeliness, authentication, integrity, and availability. Authentication mechanisms are used to "corroborate that an entity is the one that is claimed" according to the international standard ISO/IEC 9798-1 [184]. Confidentiality is less of a concern, at least for the basic functionality of this application. It is most important to ensure that only authenticated messages are delivered, and that their delivery is not disrupted. Authentication is critically important since incorrect responses of the controllers can have disastrous consequences for the system. If distributed reactive power resources can restore an unsolvable system to solvability, they can also do the opposite. End-to-end authentication is needed since intermediate routing nodes may not always be trustworthy. Design principles for authentication developed specifically for power grid cyber protocols are identified in [185]. Myriad tools exist for providing authentication and integrity. Availability is

also crucial because otherwise the controllers may not be able to reach devices to provide control, or worse, they may think that their messages are being delivered when they are not. To ensure availability, the system must be efficient in its use of computation and communication resources and must have sufficient error management built in to ensure proper handling of failures. Furthermore, the error management functions must not lead to resource exhaustion even in the face of adversarial action. Redundancy is also an important factor in maintaining availability. These issues, and several ideas to address them in the context of the voltage control framework, are explored in more detail in [41] and [42].

7.2. Measurement Protection for Application Security

The focus of this section is the protection of measurements to ensure application security. First, the subject of bad data detection is briefly reviewed. Then, an application is presented for protecting DC state estimation from false data injection. The motivation for this work is largely a response to an application proposing such attacks, explained below. A fundamental purpose of power system state estimation, as outlined in Chapter 2, is essentially to process raw data and turn it into meaningful information in the form of state variables. State estimation is a way to filter data into something humans and other applications can utilize. Due to this role of state estimation, much attention has been paid to the impact of bad data on its results.

7.2.1. Bad Data Detection

Bad data detection often accompanies power system state estimation. The purpose of bad data detection is to detect, identify, and correct measurement data which is inconsistent with the rest of the information about the system. Conventional bad data detection schemes are described in more detail as background for applications presented in [186], [182], and [187]. Measurement errors are a common source of bad data, also discussed in Chapter 5. Typical encountered errors include device misconfiguration and device failures, in addition to measurement inaccuracy. The estimation of state variables can be adversely affected by these errors. In fact, all data which is used to compute the estimate has at least some impact on its result, so it is important to find and eliminate bad data sources. Many proposed schemes exist for the purpose of detecting, identifying, and correcting bad measurements [28, 32].

Most bad data detection approaches are based on residuals. If the state estimate is denoted $\hat{\mathbf{x}}$, then the measurement residual $\mathbf{z}-\mathbf{H}\hat{\mathbf{x}}$ is the difference between the true measurements $\mathbf{z}$ and estimated values of the measurements. A common approach [28, 32] for detecting the presence of bad data is by examining the *L*2 −*norm* of measurement residual:

$$\left\|\mathbf{z}-\mathbf{H}\hat{\mathbf{x}}\right\| \tag{225}$$

Then, if the L2 −norm of the measurement residual in (225) is greater than a certain threshold τ, it is assumed that bad data is present in the observed measurements.

Until recently, it was generally assumed the bad data detection systems accompanying state estimators are sufficient to detect and recover from sensor measurement manipulation. The failure of typical bad data detection methods for malicious attacks can be suspected from the analysis in [188], but the issue has been largely brought to the attention of the power system cyber security community more recently by [32].

False data injection attacks on DC state estimation against this criterion, demonstrated in [32], illustrate the focus of existing bad data detection techniques on dealing with errors rather than interacting or malicious bad data. Bad data detection approaches which are based on the residual inherently require an initial estimate to be computed which may be affected by the bad data. That is, it is only after the estimate is first computed that the residual can be evaluated.

The work for this particular application, protecting DC state estimation, is inspired by our interest in defending against the proposed attacks in [32]. In summary, Liu et al. [32, 189] demonstrates that an adversary, armed with the knowledge of system topology, can inject false data into DC state estimators without being detected by current bad data detection techniques. The core of the problem is that most bad data detection techniques are designed to deal with certain types of errors, but not coordinated malicious changes. Under such circumstances, an adversary can fool the bad data detection scheme and cause various manipulations to state variables.

A large part of the problem is the *detection* of bad data, especially when produced by a coordinated, malicious attack. Recent work has proposed solutions to address detection problems [186, 190, 191, 192]. Cyber-based approaches typically involve deploying cryptographic mechanisms on the communications between the sensors in the field and the control center or deploying tamper resistant or tamper evident sensor hardware. Other approaches involve revisiting bad data detection schemes with the adversary in mind. Bad data detection for power systems, particularly with respect to state estimation, is a common topic in the literature [193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205]. More background on the approaches described in the literature are provided in [187].

7.2.2. Protection Strategy

We propose a protection strategy in [186] which addresses the underlying question of whether some measurements are better to protect than others. This examination is made primarily as a response to [32], where if an adversary is able to alter measurements, the adversary may be able to affect state estimation results without being detected. Conceivably, errors may be injected to manipulate the state variables to achieve specific conditions; details are omitted here.

The conclusion of this work is that we can identify a set of sensors such that if those sensors' measurements are protected, false data injection attacks on DC state estimation would be detected. Network observability plays a crucial role in this application. In state estimation, *basic measurements* are a minimum set of measurements sufficient to ensure system observability. A network is observable if the measurement matrix **H** is full rank. In [186], it is proven that protecting a set of basic measurements is a necessary and sufficient condition for detecting the false data injection attacks of [32] on DC state estimation. Multiple sets of basic measurements are straightforward to compute. We also show that a basic measurement is exchangeable for a verifiable state variable, such as a PMU angle. Thus, PMUs provide redundancy to critical measurements for further protection against such attacks. For details concerning the analysis, results, and other directions in this area, refer to [186].

A set of measurements that causes a loss of observability is called a *critical set*. In [206], a method to compute the LU factorization is described which is applied in [207] to the topology

matrix **H**. This LU factorization provides the necessary information to identify critical measurements [207], [208]. The basic idea is to do a change of basis on the matrix **H** to obtain a mapping from measurements to equivalent states. The following decomposition of **H** is then obtained as

$$\mathbf{H} = \mathbf{P}^{-1}\mathbf{L}\mathbf{U}_{\mathbf{b}} \tag{226}$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{\mathbf{b}} \\ \mathbf{M} \end{bmatrix} \tag{227}$$

where the new basis has the following structure:

$$\mathbf{L}' = \mathbf{L}\mathbf{L}_{\mathbf{b}}^{-1} = \begin{bmatrix} \mathbf{I}_{\mathbf{n}} \\ \mathbf{R} \end{bmatrix} \tag{228}$$

where *n* is the number of states. Then, the rows of $\mathbf{I_n}$ correspond to the essential measurements or basic measurements we seek. Rows of **R** correspond to redundant measurements. Columns correspond to the equivalent states which can easily be mapped back to the original states using the permutation matrix **P** obtained from the LU decomposition. If a column in **R** has no non-zero entry, then there is only one entry for that state variable and it is in $\mathbf{I_n}$. Thus, the measurement corresponding to that row in $\mathbf{I_n}$ is critical, since no information about that state variable can be obtained from any of the other measurements. Removing that measurement would cause a loss of observability. Thus, this factorization can be used to examine issues of observability relating to measurement redundancy.

As proven in [186], there is a clear relationship between identifying good meters to protect against false data injection attacks and ensuring network observability. Thus, existing techniques in power system observability analysis, such as the above matrix decomposition, provide insight into the nature of how applications which rely on measurement data can be made robust. There are a number of factors related to observability which can now be optimized over when determining which measurements to protect. Since sets of essential measurements are non-

unique, greater choices in measurement sets provide flexibility when designing a protection scheme.

7.3. Cyber-Physical Intrusion Detection Systems (CPIDS)

In this application, key aspects of the cyber-physical intrusion detection system (CPIDS) introduced in [182] are presented. Detection is an important part of protecting the power grid, and specifically designed intrusion detection techniques for large-scale cyber-physical systems are needed. While cyber-event detection is not the focus of this thesis, the cyber aspects of the problem are very much interwoven with the power aspects. A sketch of the cyber-security contributions of this work is necessarily included to convey the benefit of CPIDS to both domains. By taking advantage of network structures and data, both on the communications (cyber) side and the power side, the CPIDS is able to utilize cyber security information to make the power system applications such as state estimation more robust, an insight that no existing techniques seem to offer. A particular application of the CPIDS has been implemented for use with power system AC state estimation, and is described in this chapter. It is important to point out that although state estimation is the focus of this application, the concepts presented are not limited to state estimation. Valuable information is made available about potentially compromised hosts, which is translated into information about suspicious measurements.

At each time instant, the CPIDS identifies the compromised set of computers in the cyber network and their corresponding sets of measurements (power system sensors) which may have been maliciously modified. The main contribution is to combine information from distributed sensors of different types to detect the malicious activities within the cyber-physical system. A working prototype of the CPIDS has been implemented and evaluated. In [182], experimental results are presented for the IEEE 24-bus RTS (Figure 3), showing that the CPIDS significantly improves scalability of the tradition intrusion detection techniques by using information from both cyber and power sensors.

A key characteristic of the power grid, which is exploited by this application, is the interdependency between the cyber and physical systems. The power grid is a cyber-physical system whose *electrical* infrastructure's reliable operation depends largely on its *cyber*

infrastructure. In such an interdependent system, the requirements of both components should be considered simultaneously. Data is a common point of interdependence. Data acquisition occurs at the measurement devices. Then, data is transmitted through a communication system to a control center. In traversing the information path from the field to end-point applications for the electric power grid, it is clear that these communications and data handling alone form a complex infrastructure. The measurement devices and the communication system possess properties, cyber and physical, which may affect the electrical state of the power system though accidental and/or malicious impact on the data. This data is a fundamental requirement for power system applications. Therefore, the identification of potentially compromised measurements is valuable to any data-dependent power system applications such as cyber-physical state estimation.

While measurements of the power system are useful to verify what occurred, the conclusions are only as valid as the data is trustworthy. If one allows for the possibility that the data may be corrupted or maliciously altered, then it is no longer obvious whether the data alone can provide a clear indication of whether a perceived incident in the power grid was the appropriate response of the physical system. A concerning problem, which we aim to help address, is that power system applications which rely upon this data are usually considered in isolation from the computer and communications systems which provide the data.

Numerous examples exist of proposed detection mechanisms, specific either to the cyber domain or to the physical domain, but not both. Examples of work specific to the cyber side are listed in [182]. On the power system side, typical bad data detection has some major limitations, as described in Section 7.2.1. In particular, most initially consider *all* the measurements *good*, such as in least-square error based algorithms [209], and most ignore information from cyber sensors.

The CPIDS makes use of sensory information from both power-side measurement data and cyber-side intrusion detectors to identify cyber attacks and potential compromises of power system measurement data. It is hoped that this work will incentivize further development of cyber-physical detection systems as well as motivate real-world implementation, specifically including the technology transfer of the solution presented here.

The steps taken in the CPIDS application are summarized as follows. First in an offline preprocess phase, the CPIDS automatically generates the attack graph of the network based on the enforced access control policies. During the operational mode, the CPIDS monitors the power and the communication networks, detects and analyzes attacks based on the attack graph, and then probabilistically determines the set of computer systems and power system measurements which are likely to have maliciously been compromised. The CPIDS then uses this probabilistic information to make decisions regarding the suspicious measurements in order to protect the power system state estimator (or other applications) from the potentially malicious data. At each time instant, the CPIDS is able to provide situational awareness to the system operators by presenting them with a clear and complete cyber-physical state of the power grid.

The rest of this section is organized as follows. First, an overview is presented of the cyber-physical state notion used for the CPIDS. A short background on the cyber detection aspects, needed to fully understand this system, is included for completeness. Then, we discuss how suspicious activities are identified using a joint cyber and power state estimation algorithm. The experimental setup and evaluation are also presented.

7.3.1. Cyber-Physical System State

The purpose of the CPIDS is to link together information from the cyber network with information from the power network to better reconstruct the system state. Figure 82 [210] shows the combined cyber-physical state, which consists of compromised hosts (top) and power system consequences (bottom), determined from cyber IDS alerts and power system measurements.

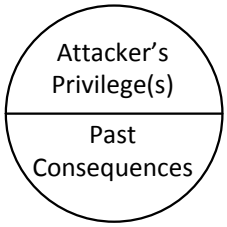

Figure 82. Cyber-physical state [210]

The compromised hosts, consequences, and associated set of probabilities are collectively called the "information state." The system information state is based on measured power system and cyber system data. The cyber data is used to determine the likely compromised hosts. Then, a mapping of the hosts to power system measurement devices is used to consolidate information about the network, both cyber and power.

Estimating the cyber-physical state involves the reconstruction of the path the attacker took to cause the event. This path is represented by an attack graph. Given the firewall rules and the communication network topology, a pessimistic attack graph is automatically generated using the proposed approach. The purpose of the firewall rules is to construct a network accessibility matrix which defines which devices can communicate. Figure 83 illustrates this concept for two hosts (*A* and *B*) and one protective relay (*R1*). The associated network connectivity map is shown in Figure 84.

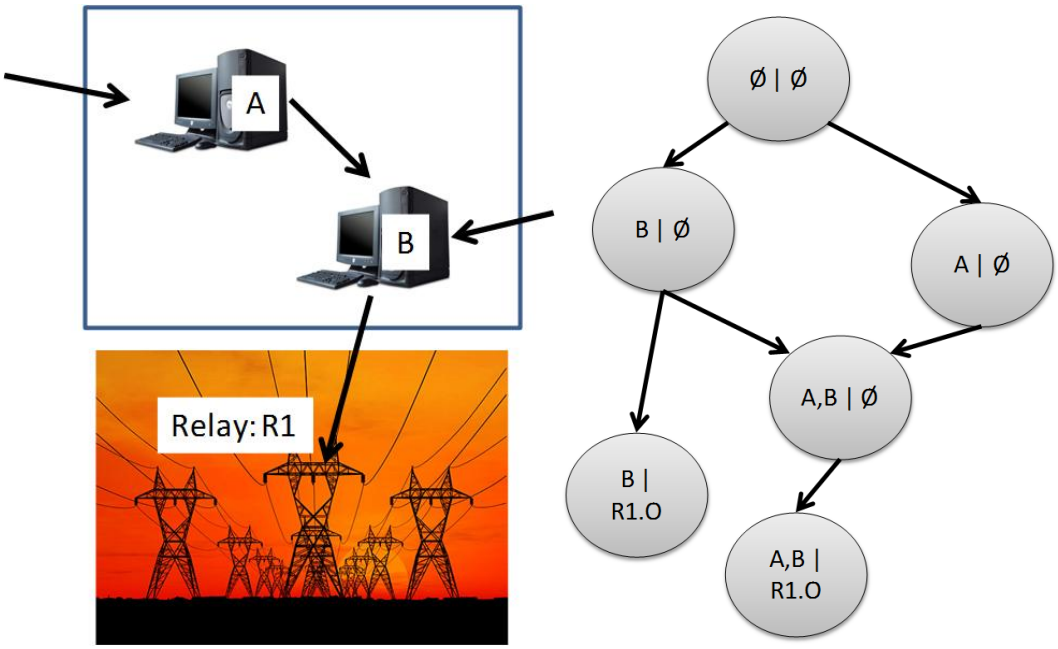

Figure 83. Privilege escalation (left) and attack graph (right) [210]

$$\begin{array}{c} \\ Rmt \\ A \\ B \\ R1 \end{array} \begin{array}{c} \begin{array}{cccc} Rmt & A & B & R1 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Figure 84. Network connectivity map for Figure 83 example [210]

The attack graph is pessimistic since it assumes that every host has a vulnerability. The firewall rules give a mapping of paths that can be used to penetrate the system. A privilege escalation graph gives the state space in which an attacker can move by exploiting vulnerabilities in the system.

7.3.2. IDSs and Attack Graphs as Hidden Markov Models

The cyber state estimation aspect of the solution is described in [182]. It is not a focus of this thesis. Only its key points are highlighted here to provide necessary background. The main inputs are the communication network topology and the security policy (e.g., firewall rulesets). This information defines the global access policy of the network. The CPIDS analyzes this input and automatically creates a network connectivity matrix as in Figure 84. The [*i*, *j*] entry of the matrix is true if the traffic from host h_i to host h_j is allowed, and false otherwise. The connectivity matrix always includes an internet node representing groups of hosts outside of the network. Attackers are assumed to initially reside in the internet node before launching the attacks. Based on the state-space notion introduced above, the approach and experimental results are outlined in the next sections.

The first step of the approach is attack graph generation for the cyber network. The starting state of the attack graph is the empty state, where the attacker has no privilege over the network (no hosts have been compromised). Starting with the attack graph's initial state and the network connectivity map, the set of next states is automatically deduced. State transitions represent vulnerability exploitations to escalate attacker's privileges. Applying this reasoning in a recursive manner, the entire attack graph is automatically generated. The generated attack graph is then available to use while the system is operating. To improve scalability and deployment ease, CPIDS generates attack graphs at the network level. Details are provided in [182] regarding

how to generate these attack graphs for real systems. By design, the completed AGT includes all possible attack paths from a remote (internet) host against the network.

In the next step of the approach, the CPIDS uses the generated AGT and online alerts from the distributed IDSs to do forensics analysis. This is done by modeling the AGT as a hidden Markov model (HMM). An HMM [211] assumes the system to be a Markov process with unobserved (hidden) states. However, the output which is dependent on the state is visible. Each state has a probability distribution over the possible outputs, so observing the output sequence gives some information about the sequence of states. Note that "hidden" refers to the state sequence through which the model passes, not to the parameters of the model; even if the model parameters are known exactly, the model is still hidden.

All state transitions of the HMM (network links of the AGT) are monitored by IDSs. A list of IDSs and corresponding links can be defined within the network topology file. The IDSs provide probabilistic locations of intrusion in the system. The observation probabilities of the HMM are IDS false negative and false positive rates. While the system is operating, at each time instant, the engine receives a set of IDS alerts and uses the generated attack graph to find the most likely attack path traversed by the attacker. The HMM allows for the fact that we have incomplete information and that the observed state might not be the true state. The result is probabilistic information of the attacker's privileges. To obtain this information, it is necessary to predict the most likely sequence of states in the HMM from a given observation sequence. To find the most likely hidden sequence of states, a dynamic programming algorithm, called Iturbi [212], is applied to the above problem. The solution gives the most likely path of the attacker.

7.3.3. Malicious Data Detection and Handling

Ultimately, the results reveal which measurements are potentially compromised and with what probabilities. Information about the attacker's current privileges can then be used by power system applications to make decisions about the data. Given the attacker's privileges, the CPIDS automatically determines a set of power measurements which may be compromised. These measurements may have been modified in a malicious way. Power system applications are then able to account for this *before* the measurements corrupt applications.

Bad data detection for power system state estimation, using the non-linear AC model described in Chapter 2, is implemented and evaluated with the CPIDS for this work. A description of this evaluation is given in [182]. Comparatively, a simplified DC model (23) is often used in papers addressing false data injection attacks on state estimation [32]. To prudently account for maliciously altered data, the proposed approach circumvents the need to compute the first estimate from the full data set.

Consider that if there are *M* sensors in the power network, any combination of the sensors may produce corrupted, altered, or "bad" data. This yields 2^M possible combinations of sensors which may be simultaneously compromised. The possibilities range from $\mathbf{m} = (0,0, \ldots ,0)^T$ if all measurements are good to $\mathbf{m}= (1,1, \ldots ,1)^T$ if all measurements are bad. The approach proceeds as follows. Assume each one of the 2^M possibilities, and omit the measurements at the indices which correspond to a value of "1." Using the remaining measurements, proceed with the normal state estimation routine. As long as there are enough good measurements which are available in order to ensure observability (basic measurements), the other measurements can theoretically be omitted from the analysis. From the state estimation routine, obtain (1) the state estimate and (2) the estimates of the measurements calculated from the state estimate. Using (2), determine if the estimate of the measurements which were assumed to be bad are significantly different from the true measured value. If so, then conclude that a particular set of measurements is bad.

Without information from cyber sensors, the method above is combinatorial [213] and should not be used, since the search space to detect bad measurements grows exponentially in the number of measurements. We take advantage of the IDS alerts and HMM results to refine the above method and address the scalability issue. The hosts and corresponding measurements which pose a potential threat may be segregated into networks or control areas (consisting of a group of measurement devices). When particular measurements are identified as suspicious, those sensors should be evaluated to confirm whether they are in fact producing malicious data. The set of flagged sensors will (typically) be much smaller than the total number of sensors in the network, so their subsequent evaluation is scalable.

7.3.4. CPIDS Implementation and Evaluation

The simulated power system used in this work is the IEEE 24-bus RTS, shown in Figure 3. The system is divided into two control networks. The first control network consists of buses 1-12 and the second control consists of buses 13-24. For each control network, there is one host corresponding to each bus. In general, each system will have different numbers and configurations of control networks and hosts. The Network Access Policy Tool (NetAPT) is used to obtain a network connectivity map listing all the host pairs and ports between which traffic is allowed given a topology and firewall rules.

PowerWorld Simulator is used via SimAuto in MATLAB to simulate the power system. The MATLAB program acts as an interpreter, taking the simulated SCADA measurements of the system as well as inputs from the IDSs in the cyber network and interfacing with the state estimator. The MATLAB toolbox MATPOWER [214] is used as the state estimator. The implementation occurs as follows:

(1) The program first reads in the HMM result file. This file contains the information state, which reveals the host or hosts that are likely to be compromised.
(2) The program internally maintains a Host Measurement Map, which keeps track of which hosts have control over which power system measurements. Based on which hosts are determined to be compromised from step (1), the Host Measurement Map is used to determine which corresponding measurements are likely to be compromised by an attacker.
(3) The program removes the potentially malicious measurements and does an observability check.
(4) If the system is still observable, the potentially malicious measurements are ignored. Then, the state estimation procedure proceeds with the remaining measurements.
(5) Based on the computed state estimate, the system real and reactive power flows are computed.
(6) The values of the ignored potentially malicious measurements from step (5) are compared with the computed measurements in step (6) to determine which measurements in (5) are

significantly different. This provides identification of whether a measurement has actually been modified.

The implementation of the CPIDS and results are described in [182]. An overview of the combined framework is shown in Figure 85.

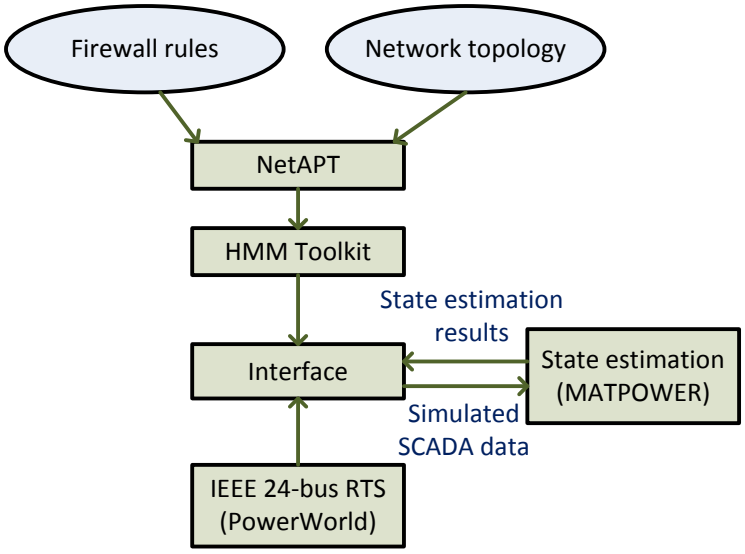

Figure 85. System implementation overview

Based on the AGT for the sample network, the CPIDS reveals the compromised hosts and associated probabilities. From the results of the HMM file, highlighted in Figure 86, the hosts at buses 1 and 2 are determined to be compromised with probability 0.953. This information is obtained during step (1) of the above procedure.

Prob.	Compromised hosts		
0.032141	28.1.1.1		
0.953099	28.1.1.1	101.10.0.3	
0.001001	28.1.1.1	101.10.0.3	101.11.0.3
0.016759	28.1.1.1	101.11.0.3	

Figure 86. HMM result file

The structure of the host measurement map for the system is illustrated in Figure 87. The host measurement map is compared with the file in Figure 86 during steps (2) and (3). As revealed in

Figure 87, the hosts at buses 1 and 2 correspond to the hosts identified in Figure 86, 28.1.1. and 101.10.0.3. In this example, each bus controls measurement data for its adjacent lines. The mapping is shown only for the suspected compromised hosts, although the entire system mapping is maintained.

```
    Host: Bus1: 28.1.1.1
PF  1 2;1 3;1 5;
PT  ;
QF  1 2; 1 3; 1 5;
QT  ;
    Host: Bus2: 101.10.0.3
PF  2 4; 2 6;
PT  1 2;
QF  2 4; 2 6;
QT  1 2;
```

Figure 87. Measurement map for potentially compromised hosts

In Figure 87, labels *PF*, *PT*, *QF*, and *QT* correspond to measurement types. Real power flows are denoted *PF* and *PT*, measured respectively at the *from* and *to* buses of the lines. Similarly, *Q* denotes reactive power flow measurements. As an example interpretation of one of these labels, Bus 1 monitors the real and reactive power flow on line (1,2) at the line's designated *from* bus. The format of the data identifiers in Figure 87 can be modified as necessary to match the convention used by the utility.

Once the measurements corresponding to bus 1 and bus 2 are flagged, they are removed according to step (4), and state estimation then proceeds successfully. Note that choices exist for how to handle flagged measurements. In addition to removing some or all of the flagged measurements, one may simply reduce the weightings of the suspicious measurements. Removing the flagged measurements is the simplest but not necessarily the optimal action. If the observability check in step (5) fails, removing all measurements is not an option. One may take the reduced set of selected sensors and perform the refined bad data verification and removal process. The subset may be expanded to include sensors which share a network or control area with the suspicious sensors. Analysis can determine specifically which measurements are producing results inconsistent with the other data. In summary, appropriate post-processing schemes should be developed to deal with the flagged measurements.

7.3.5. Data Rejection vs. Observability

The ability to estimate the system state from the given measurements (observability) is an important property which must be preserved, and the CPIDS solution should be applied with this consideration in mind. In this work, observability is important because compromised measurements may only be discarded if those compromised measurements are all redundant. If too many measurements are compromised, they cannot be removed, or the system will no longer be observable. Section 7.2 provides further analysis of the relationship between measurement protection and observability.

When selecting measurements to remove, we must avoid the removal of critical sets. If the set of suspicious measurements includes a critical set, the system will lose observability if those measurements are removed. Hence, observability analysis must be performed before an attempt is made to compute the state estimate with a reduced measurement set.

There exists a tradeoff, which may be encountered in practice, regarding whether it is more important to discard a potentially malicious measurement, and lose observability, or to keep the potentially malicious measurement and know that the resulting state estimate may be incorrect. Fortunately, those are not the only two possibilities. If one knows that a critical measurement may be malicious, one can take action to remedy the situation. This may be as simple as inserting a pseudomeasurement. It is important to point out the existence and possible occurrence of this situation. However, recommendations on how to handle a critical malicious measurement set are beyond the scope of this work. The key point is that the situation should be monitored and may require a decision concerning what tradeoffs to make.

7.4. Perturbation-Based Data Quality Assessment

As the previous sections and chapters have indicated, data quality is a serious practical problem which in many ways presents a barrier which must be overcome before applying more advanced techniques to the refined data. Measurement data in power systems may be “bad” for any number of reasons. Malicious data injection attacks, which alter the values of measurements without being detected, are one potential cause of bad data and may have serious consequences.

The idea in this application is to use a probing technique as a way to detect bad data. In principle, the idea is simple. We apply some known controlled input to the system, and we measure the response. The basis of the approach is that we (the defenders of the grid) perform the probing, so we also know the injected probing signal. If the response does not match what we expect, then we should be suspicious that something is wrong. More specifically, the solution is proposed in [187] for bad data detection in power systems, and it is particularly designed to detect malicious data attacks.

By applying particular perturbations to the system and measuring the changes elsewhere, we probe the system for unexpected responses of measurement values. Evaluating measurement changes due to the probes provides an indication of both bad and malicious data. Using the "keyspace approach," developed in this work, the probes are intended to be unknown and unpredictable to the attacker.

Conceptually, the approach is closely related to other applications of perturbations in power systems. Here, the perturbations are used for a specific purpose, which is to gauge the response of data. The perturbation approach detects data which does not respond appropriately. Regardless of whether the data is incorrect because it has been altered maliciously or because a measurement device is simply misconfigured or malfunctioning, the system recognizes the discrepancy and indicates that further investigation is warranted.

In this section, we describe an example of using our probing approach for detecting false data injection attacks. Specifically, we hypothetically leverage D-FACTS devices, introduced in [215], to perturb the system by changing the effective impedance on a set of lines. The perturbations should be large enough to cause an observable change in the measurement values of interest, but at the same time the perturbation should be small enough not to move the system away from optimal operation. These limits are investigated in [187]. It is important to highlight that our approach is in no way limited to D-FACTS devices. Any controllable components may be used. For example, an analogous approach may be developed for reactive power injection or other input perturbation schemes. In fact, it may be the case that other means of perturbation are

simpler, less costly, or more effective to implement. Such a comparison is beyond the scope of the current work.

The motivation for this work comes from a desire to design a bad data detector that is general enough to work reliably whenever it is needed and is robust against attacks. Rather than design a bad data detector for specific applications, we purposefully develop an approach which is application-independent. A great advantage of this approach is that it is completely independent of end-point applications which ultimately use the data. The reason is simple: bad data is a problem for everyone, yet proposed solutions are often overly specific to a particular application. While most bad data detection methods are closely interwoven with state estimation and possibly other applications, this solution avoids that limitation, and can therefore help protect the entire range of applications.

Using probing for the cyber security for control systems is investigated by [216]. In particular, [216] perturbs the system to detect replay attacks on a control system. In power systems, probing to study oscillatory modes is common in the western interconnect of the US (WECC) [217, 218]. The probing tests in WECC and their results are reported in several NASPI documents [219, 220]. A study of probing signal design for power systems is found in [221]. The most common probing signals for power systems, from [222], are rectangular pulses, brief periodic waveforms, sustained sinusoid signals, and sustained noise signals. In [222], square waves are used to probe specific oscillatory modes.

Specific devices where the probe is to be injected must chosen. The devices used to perform the probing could be of any type, or any combination of types. Sensitivities are again useful for providing the ability to characterize the effect of a small change. The probing approach can utilize any pair of variables where it is possible to perturb the input and observe the output. The generality of the solution approach is a key point to highlight.

In [187], the system is probed by changing D-FACTS settings to pre-determined values in the neighborhood around the optimal point. The vector of new settings is termed a *key*. If the measurements do not change in the way we expect, there may be cause for alarm, and further

investigation is warranted. It is assumed that the attacker can compromise values received from the measuring devices in the field. Random selection of the key from a pool of acceptable keys provides security by preventing the attacker from simply spoofing his measurements to match the values we expect from our probe. Keys, keyspaces and key choice are explored further in [187]. Each key $\mathbf{k_i}$ is a vector of impedance settings for controlled lines, where $\mathbf{k_0}$ denotes optimal operating point. The sensitivities, the key, and the current set of D-FACTS settings allow us to determine the predicted set of values. The perturbation applied to the system is denoted by $\mathbf{\Delta x}$:

$$\mathbf{\Delta x} = \mathbf{k}_i - \mathbf{x} \tag{229}$$

The sensitivity matrix **A** describes how measurement variables such as voltages or line flows (depending on choice) are expected to change in relation to a change in impedance. Denote the initial measurements $\mathbf{m_0}$ and the measurements at the next observed time $\mathbf{m_{obs}}$. For voltage magnitude measurements, the expected changes are $\mathbf{\Delta m = m_{obs} - m_0}$ in (230), and the predicted perturbed voltages are $\mathbf{m_{pred}}$ in (231).

$$\mathbf{\Delta m} = \mathbf{A\Delta x} \tag{230}$$

$$\mathbf{m}_{\mathbf{pred}} = \mathbf{A\Delta x} + \mathbf{m}_t \tag{231}$$

Bad data is detected by subtracting $\mathbf{m_{pred}}$ from $\mathbf{m_{obs}}$, discarding values within the acceptable error range, and noting any values remaining. These flagged values are indicative of bad data, and provide a starting point for further investigation of compromise or other measurement problems. Further validation of the model can be provided by comparing the observed difference in measured voltages be $\mathbf{\Delta m_{meas} = m_{obs} - m_0}$ times the inverse of **A** (assuming **A** is invertible) with the known probe. That is, we compute an estimate of our probe from the actual data, denoted by $\mathbf{\Delta x_{pred}}$ in (232).

$$\mathbf{\Delta x}_{\mathbf{pred}} = \mathbf{A}^{-1}\mathbf{\Delta m}_{\mathbf{meas}} \tag{232}$$

An important attribute of our approach is that detection is done by comparing the expected changes due to our probe with the actual changes, and the approach is enhanced by the secret, private key, which conceals the nature of the chosen perturbation. The approach does not require inherent trustworthiness of the observations. Effectively, the approach separates detection from trust determination, which can begin after an observation has been deemed suspect. This process of trust determination remains to be fully described, but it begins with finding "anchor points" for which the data can be validated as accurate. Once the set of trustworthy anchors is established, it is possible to continue to build trust levels for the remaining sensors.

7.5. Conclusions

The focus in this chapter is on protecting data itself as well as any applications which use the data. Through this work we have determined that the protection of power system data and applications is a broad subject. The applications in this chapter provide some key steps towards solving these problems. A key conclusion drawn from this work is that, moving forward, there needs to be an increase in studies and applications which focus on how to consider these seemingly different systems (cyber and physical) simultaneously.

The first application in the chapter describes a hierarchical framework to facilitate distributed control such as for voltage support, as in Chapter 3. In general, such a scheme can be used to enact any corrective and preventative controls, and the potential applications of the framework extend beyond voltage control. The framework is designed to segment and secure communications and to be able to easily assign and manage duties, which is especially important as the number of devices involved in a response action may be quite large. The ICS structure is flexible enough to support the handling of problems in a decentralized way. Rather than involve the Central EMS, an event may potentially be handled within the individual regions where it occurs. In that case, a device at a lower layer in the hierarchy may assume the highest command position for that incident. If a top-level device on any of the realms detects a local problem, it can initiate correction of the problem by coordinating the devices beneath it. The approach has since been studied in the context of an agent-based framework [223] and decentralized control [224]. The incentive to pursue these alternative approaches is driven by the impact of such

control on the rest of the system and whether there is a perceived benefit over centralized control.

In the second application, presented in [186], the key result is to analytically relate power system observability to protection with respect to cyber attackers, and we prove that protecting a set of basic measurements is a necessary and sufficient condition for detecting the false data injection attacks on DC state estimation. It is recognized that techniques in power system observability analysis can provide insight into how applications which rely on measurement data can be made robust.

A discrepancy may exist between the precipitating event that the power system data indicates and that which the cyber information indicates. Such discrepancies need to be reconciled. The third application is the design and implementation of a cyber-physical intrusion detection system (CPIDS), presented in [182]. While the *cyber*-event detection aspects of this application are not the focus of this thesis, the cyber and power (physical) aspects of the problem highly interdependent. By taking advantage of both the communications and power network as well as their data, the CPIDS provides a solution to make power system applications more robust, an insight that no existing techniques seem to offer. The approach circumvents the need to compute the first estimate from the full data set. The approach is resistant to noise and does not depend on the existence of a linear power-system model. A key benefit is the ability provide feedback to operators for identifying data sources which are potentially compromised. The CPIDS can help operators evaluate the extent to which a perceived problem is caused by data which is wrong due to a cyber problem or due to a physical power system problem. A working prototype has been implemented and evaluated, and the results reported in [182]. The usefulness of this application is more general than bad data detection for state estimation; there are many ways information about potentially compromised hosts and measurements can be utilized on the power side.

Interesting applications lie ahead for implementations of the CPIDS on real systems, since real systems have real communications network with real IDSs which can be utilized. Future plans for this application include the TCIPG test bed and real utilities. Other directions for future enhancements include asking utilities about IDS domains and real compromises to give us

insight so we can better validate our framework with what is done in real life. Additionally, utilities may be able to provide information about the frequency of such events in their system. It is also possible to provide this as a tool to utilities to obtain the host measurement maps for their real systems. The host measurement maps are helpful in determining what to protect. For example, certain PMUs are recording more measurements, so those may be more critical. That is, we may need to think in terms of entire RTUs or even entire substations going down. The cost is essentially the same to compromise or protect measurements which all belong to one device (host, RTU, PMU). Thus, grouping measurements that belong to the same RTU, or sets of RTUs that do not cost extra to protect, may be a sensible defense strategy. Such groupings are enabled by the work in this thesis. Finally, it would be useful for the industry to develop a set of cyber-physical models for experimental and study purposes based on a tool like this, in particular for benchmark systems such as the IEEE 24-bus used in this work.

The final application considered in this thesis involves using measurement probing as a technique to detect bad data. This approach is purposefully designed to be an application-independent solution for detecting data which is in error. Malicious activity is considered as a special case of other data-source problems which could be detected using this technique. Analysis and results are presented in [187]. For bad data detection, one can also use data trending methods to characterize when measurements are behaving/changing in an inconsistent way.

8. Conclusion

There exist a wide variety of techniques which facilitate the enhancement of knowledge based on data. The commonality among such techniques, which are often referred to under the very broad category of data mining, is that they provide a way to discover and recognize patterns and relationships in the data in order that one may learn about the system and utilize the knowledge for improved power systems analysis, operations, and control. Thus, in this thesis, it is shown that several seemingly different aspects of power systems analysis are related through their ability to be enhanced by finding and utilizing patterns that exist in data.

The stated purpose of this work was to adopt an application-centric viewpoint, and study the use of data mining and data mining concepts in the context of certain important areas and applications of power systems analysis where data mining insight is deemed to be useful. This has been accomplished. Several areas with specific applications for which the benefits of these techniques are immediately apparent are described in this thesis.

In the power systems field, the types of data which need to be handled include varying granularities of models of the system, models of the generators, models of the loads, simulation results, and actual measurements. The existing data that is already being collected is underutilized, and its full value is still being understood. Even when considering data mining specifically for power systems, as in this work, a cross-disciplinary approach is required to leverage expertise from several complementary areas. The major conclusions from the applications in these areas are summarized below.

To summarize, several key areas have been identified and addressed in this work. These are the following: sensitivity-based pattern recognition, model estimation and enhancement based on data (focus was on network models, but applicable to other models as well), statistical analysis of time series data, event signature recognition (oscillation monitoring), and identifying data problems and building applications which are robust to those data problems.

There is considerable insight to be gained from sensitivities, which is valuable throughout this thesis. Sensitivities are a repository of expected patterns revealing how variables are coupled to each other and may be exploited by myriad machine learning, data mining, and pattern recognition tools. In Chapter 3, sensitivities are used for the selective grouping of reactive support devices for providing wide-area voltage support as well as for identifying generators with market power potential. Thus, Chapter 3 reveals that clustering sensitivities allows for the selection and aggregation of buses with the ability to "help each other." Offering a variety of algorithmic tools as choices to the user is important. The concepts developed in Chapter 3 are general and applicable to a wide range of applications.

The validity of power network models is important; the model needs to be consistent with the data which is obtained. Chapter 4 illustrates the benefit of data-over-time for obtaining information about the system, facilitating improved network modeling. From this chapter, it is apparent that data and data mining insight can be utilized to develop models, an important focus since such models underlie most power systems analysis. In Chapter 4, the need for well-defined metrics for acceptable network reductions is also made apparent.

When underlying models of the system are unavailable or incomplete, there is still a need to know what events are occurring. Section 4.2 explores the extent to which it is possible to develop such models from the data itself. In summary, although this estimation is possible, its accuracy depends on factors over which we have no control. Obtaining and analyzing real data in the context of this application will provide a better understanding of how severe the "external factors" are and how much they will truly affect the estimation. Then, Section 4.3 builds upon the clustering and data mining insight from Chapter 3 to group network elements and obtain reduced system models which preserve important structural features, where the important features may vary depending on the application. The tools created for this application provide a framework for evaluating any equivalent or model for which transient stability properties are of interest. The application in Section 4.4 is analogous to the first application except that rather than estimating an equivalent model, it estimates the parameters on specific lines. There is great potential to further develop and apply these methods, as discussed in Section 4.5.

The introduction of real data and error for this application is considered in Chapter 5. Real data provided by a North American utility is studied. Visualization is an important aspect of data analysis and pattern recognition. Tools such as correlation analysis and visualization are used to identify relationships in the measurement data at a basic level. These empirical relationships can be compared to what is expected and supported by the model, as in Chapter 4. From Chapter 5, we conclude *real data* is important for developing data mining applications. Especially for the type of analysis presented here, it is important to obtain data corresponding to *known events* for learning and verification purposes. It is difficult to represent real data distributions in simulation, and there is no substitute for validating applications with real data. Results with real data also indicate its usefulness for identifying sources of problems, which may be due to model problems, data problems, or both.

Chapter 6 examines the associations between measured modal content and events in the system. The question is whether this type of information is enough to distinguish between events. We find that certain changes impact modal estimates in distinct, predicable ways. Ringdowns differ from each other to an extent related to how different the events are from which they originate. The applications in Chapter 6 develop a methodology for using OMS data to classify ringdown events resulting from generator faults. Our results show that with knowledge of the nominal modes of certain generators, it is often possible to distinguish an event at one of these generators. Regardless of which modal estimation algorithm is used, the feasibility of distinguishing different events from each other in the resulting ringdown data is clearly illustrated. Thus, a relationship exists between the system eigenvalues, which may be associated with certain generators, and the modes in a ringdown. Measurement data can be used to characterize these relationships to facilitate event detection and classification. The framework can be extended to other events, and is a forerunner to performing more in-depth event analysis based on patterns or signatures in estimated modal content. An extension to exciter setpoint changes is considered in Section 6.5.

The modal estimates obtained from OMS results can be processed in an automated way which facilitates decision making. In Chapter 6, an approach is presented to automatically classify ringdowns to events based on modal content. Results in Section 6.4.3 show that the ringdowns in

the WECC case can be correctly classified with this approach. More detailed conclusions and results from the OMS applications are presented in Section 6.6. Examining the contributions of generator electromechanical modes to ringdowns lays the foundation for continued work in detecting and distinguishing event characteristics. Continuing work is needed to develop methods for signature recognition and classification in OMS data. It is possible to build upon the work presented here to ultimately be able to identify a wide variety of signatures for different types of events. A main contribution of Chapter 6 is essentially to highlight this fact.

The focus of Chapter 7 is on protecting data as well as applications which use the data. The cyber and power (physical) aspects of protecting data and applications are highly interdependent. Two general conclusions are drawn from the work in this chapter. First, we conclude that moving forward, there needs to be an increase in studies and applications which focus on how to consider these seemingly different systems (cyber and physical) simultaneously. Secondly, through this work, we have determined that the protection of power system data and applications is a broad subject, and the applications in this chapter provide key steps towards solving these problems. By no means are these solutions expected to provide full coverage.

There are numerous existing software tools and documented methods to perform well-known data mining tasks such as classification, so those have not been the focus of this work. What is the most pertinent for the power systems field right now is to analyze how power system problems can utilize data to solve problems and how data mining tools may be able to help automate the decision making process.

While there are benefits of using data, there are also barriers which have prevented us from utilizing data to its full potential. There is no reason to believe that these barriers will disappear or become less of a problem on their own. The barriers must be directly addressed and adequate solutions must be found. In summary, several of the major barriers fall under the category of data quality problems. The problem is data availability as well as integrity. Availability issues at the low level are related to storing, handling, and passing the raw data. First, making data available requires a communication network with specific characteristics, and there are technical

challenges associated with making the correct data available to the correct people at the correct times.

However, even if deployable technical solutions are developed so that the appropriate applications may obtain and share data as required, there still remain significant barriers for data dissemination, largely due to the proprietary nature of the data. A lot of the data resides in industry. Entities which own the data are often (but not always) reluctant to share it with others, even for research purposes. It is difficult to do advanced data analysis research without ready access to data. One needs data to do the research about what is useful within the data. These concerns are understandable from their perspective, especially when the goal is, in fact, to extract information about the system from the data. There is the potential concern that the relationships extracted from the data may reveal information that the owners do not wish for us to know (proprietary information). Thus, it may be said that their desire for confidentiality of their data actually causes a lack of availability and integrity. This problem is beginning to be more widely recognized, and there are a number of possible proposed solutions and work-arounds. The current approach is to sign legal agreements between entities who wish to share and use the data. These are bureaucratic challenges. A long-term solution to this problem is needed, and finding and implementing such a solution presents a challenge since the entities and organizations involved in the data-flow may have different or even contradictory goals. The policy and bureaucracy obstacles to data availability are not a conventional engineering problem, but they certainly present a barrier to solving engineering problems. Thus, it may be that a solution can be orchestrated just as well by engineers as by anyone else.

Assume the above problem is solved and data is made available. Then, the data one obtains often has errors of indeterminate types. These are integrity issues. The contents of each data point are intended to reflect physically based quantities, represented by specific identifiers. A data integrity failure is when an actual data value fails to represent the identifier with which it is associated. The identifiers reflect what the data is supposed to represent; typically these are based on physical measurements but may also involve constants or components of a model. If the data does not reflect the true value of the identifier, this is an integrity problem. In the course of doing this work, it was discovered that bad data exists in reality and truly is a source of a lot of

problems. Thus, finding effective ways to find, handle, and remove bad data and to build applications which are robust with respect to bad data has immeasurable value for future work. Bad data is seen as a serious impediment to rolling out large-scale data mining solutions, and applying algorithms blindly without solving the data quality problems first may be a foolish or at least an unfruitful pursuit. In summary, data quality is a practical problem that in many ways presents a barrier which must be overcome before applying more advanced techniques to the refined data. Strategies for detecting and dealing with data problems are presented in Chapter 5 and Chapter 7. The work on bad data in this thesis is motivated by the desire for future efforts to more easily proceed in this area, unhindered by the problems associated with data quality.

In summary, the challenges associated with the data may act either as a barrier or as an enabler to the smart grid. Data mining is a broad, large area, full of useful concepts and techniques, and attempting to apply all of it to power systems at once is not feasible. An analogy can be drawn between "data mining" and "smart grid"; both are terms which seem to encompass everything. It is certainly beyond the scope of this thesis to attempt to address the entire problem. The work in this thesis has required a broad background both on different areas and techniques related to data mining and advanced data processing as well as power systems areas. The recommendation from this work is to focus on particular aspects of the whole problem. The work in this thesis provides a coverage of data analysis and categories of its potential uses and application areas in power systems.

This thesis motivates further extensions in several seemingly different but related areas, where the central connection between these areas is their potential to be enhanced by data and data mining. It is hoped that this thesis will provide the foundation for more of such explorations. In fact, the work of this thesis is already accomplishing such a purpose. Most of the explored applications in this thesis have already found a following, armed with this knowledge, to continue work along related directions. It is hoped that subsequent researchers can take the data analysis principles and power system insights in this thesis and develop applications which have not yet been fathomed.

References

[1] M. McGranaghan, "Making connections: Asset management and the smart grid," *IEEE Power and Energy Magazine*, vol. 8, pp. 16-22, Nov./Dec. 2010.

[2] P. Myrda, "Optimizing assets: Smart grid-enabled asset management," *IEEE Power and Energy Magazine*, vol. 8, pp. 109-112, Nov./Dec. 2010.

[3] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10-15, Apr. 1993.

[4] D. W. Allan and M. A. Weiss, "Accurate time and frequency transfer during common-view of a GPS Satellite," in *Proceedings of the 34th Ann. Freq. Control Symposium, USAERADCOM*, 1980, pp. 334-346.

[5] A. G. Phadke, J. S. Thorp, and M. G. Adamiak, "A new measurement technique for tracking voltage phasors, local system frequency and rate of change of frequency," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, pp. 1025-1038. May 1983.

[6] A.G. Phadke, J.S. Thorp, and K.J. Karimi, "State estimation with phasor measurements," *IEEE Transactions on Power Systems*, vol. PWRS-1, pp. 233-238, Feb. 1986.

[7] *Instruction Manual, SEL-421 Relay, Protection and Automation System*, Schweitzer Engineering Laboratories, Inc., Pullman, WA, 2009.

[8] T. Overbye, P. Sauer, C. DeMarco, B. Lesieutre, and M. Venkatasubramanian, "Using PMU data to increase situational awareness," Power Systems Engineering Research Center (PSERC) final project report, pub. no. 10-16, Sept. 2010.

[9] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp.4-37, Jan. 2000.

[10] D. Roverso, "Multivariate temporal classification by windowed wavelet decomposition and recurrent neural networks," *International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies*, Washington, DC, Nov. 2000.

[11] B. B. Mandelbrot and J. R. Wallis, "Noah, Joseph, and operational hydrology," *Water Resources Research*, vol. 4, no. 5, pp. 909-918, Oct. 1968.

[12] S. K. Friedlander and L. Topper, *Turbulence: Classic Papers on Statistical Theory*. New York, NY: Interscience, 1961.

[13] K. Bae and J. S. Thorp, "An importance sampling application: 179 bus WSCC system under voltage based hidden failures and relay misoperations," in *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, 1998, pp. 39-46.

[14] C. L. DeMarco, "A phase transition model for cascading network failure," *IEEE Control Systems Magazine*, vol. 21, no. 6, pp. 40-51, Dec. 2001.

[15] B. A. Carreras, D. E. Newman, I. Dobson, and A. B. Poole, "Evidence for self-organized criticality in a time series of electrical power system blackouts," *IEEE Trans. on Circuits and Systems*, vol. 51, no. 9, Sept. 2004, pp. 1733-174.

[16] S. R. Kulkarni, G. Lugosi, and S. S. Venkatesh, "Learning pattern classification - a survey," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2178-2206, Oct. 1998.

[17] L. W. Couch II, *Digital and Analog Communication Systems*, 7th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

[18] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Amsterdam: Morgan-Kaufmann, 2005.

[19] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

[20] B. Scholkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[21] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264-323, Sept. 1999.

[22] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Boston, Mass.: Pearson Addison Wesley, 2006.

[23] J. Mandel, *The Statistical Analysis of Experimental Data*. Reprint, 1984, Mineola, NY: Dover Publications, Originally published New York: Interscience, 1964.

[24] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 1999.

[25] G. R. Krumpholz, K. A. Clements, and P. W. Davis, "Power system observability: a practical algorithm using network topology," *IEEE Transactions on Power Systems*, vol. PAS-99, no. 4, pp. 1534-1542, July/Aug. 1980.

[26] A. Monticelli, *State estimation in electric power systems: a generalized approach*. Boston, Mass.: Kluwer Academic Publishers, 1999.

[27] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part I: exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120-125, 1970.

[28] A. Wood and B. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. New York, NY: John Wiley and Sons, 1996.

[29] F. Wu and W. Liu, "Detection of topology errors by state estimations," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 176-183, Jan 1989.

[30] H-J. Koglin, Th. Neisius, G. Beißler, and K. D. Schmitt, "Bad data detection and identification," *International Journal of Electrical Power and Energy Systems*, vol. 12, no. 2, pp. 94-103, April 1990.

[31] R. L. Lugtu, D. F. Hackett, K. C. Liu, and D. D. Might, "Power system state estimation: Detection of topological errors," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 6, pp. 2406-2412, Nov. 1980.

[32] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *CCS '09: Proceedings of the 16th ACM conference on Computer and communications security*, 2009, pp. 21-32.

[33] J. J. Lee, (2006, Jan. 10). SVD (singular value decomposition) and its applications. [Online]. Available: http://mathnet.kaist.ac.kr/Video/etc/dongseo/1101_Lee.ppt

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Md.: The John Hopkins University Press, 1996.

[35] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA: Addison Wesley, 2006.

[36] C. Chen, *Linear System Theory and Design*, 3rd ed. New York, NY: Oxford University Press, 1999.

[37] R. Bo and F. Li, "Power flow studies using principle component analysis," in *Proceedings of the 40th North American Power Symposium*, 2008, pp. 1-6.

[38] O. Alsac, J. Bright, M. Prais, and B. Stott, "Further developments in LP-based optimal power flow," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 697-711, Aug. 1990.

[39] K. M. Rogers, "Power system control with distributed flexible AC transmission system devices," M.S. thesis, University of Illinois at Urbana-Champaign, Urbana, IL, May 2009.

[40] N. Sandell, Jr., P. Varaiya, M. Athans, and M. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Transactions on Automatic Control*, vol. 23, no. 2, pp. 108-128, Apr. 1978.

[41] K. M. Rogers, R. Klump, H. Khurana, and T.J. Overbye, "Smart-grid –enabled load and distributed generation as a reactive resource," in *Proceedings of the 2010 IEEE PES Conference on Innovative Smart Grid Technologies*, 2010, pp. 1-8.

[42] K. M. Rogers, R. Klump, H. Khurana, A. Aquino-Lugo, and T.J. Overbye, "An authenticated control framework for distributed voltage support on the smart grid," *IEEE Transactions on Smart Grid,* vol. 1, no. 1, pp. 40-47, June 2010.

[43] D. L. Logue and P. T. Krein, "Utility distributed reactive power control using correlation techniques," in *Proceedings of the Sixteenth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2001, vol. 2, pp.1294-1300.

[44] M. Begovic, A. Pregelj, A. Rohatgi, and D. Novosel, "Impact of renewable distributed generation on power systems," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS)*, 2001, pp. 654-663.

[45] D. Logue and P. T. Krein, "The power buffer concept for utility load decoupling," in *Procreedings of the IEEE 31st Annual Power Electronics Specialists Conference (PESC),* 2000, vol. 2, pp. 973-978.

[46] W. W. Weaver and P. T. Krein, "Mitigation of power system collapse through active dynamic buffers," in *Proceedings of the IEEE 35th Annual Power Electronics Specialists Conference (PESC)*, 2004, vol. 2, pp. 1080-1084.

[47] N. Balu et al., "On-line power system security analysis," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 262-282, Feb. 1992.

[48] North American Electric Reliability Corporation. (2008, July). Reliability standards for the bulk power systems of North America. [Online]. Available: http://www.nerc.com/files/Reliability_Standards_Complete_Set_21Jul08.pdf

[49] U.S.-Canada Power System Outage Task Force. (2004, April). Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations. [Online]. Available: https://reports.energy.gov/BlackoutFinal-Web.pdf

[50] T. J. Overbye and R. P. Klump, "Determination of emergency power system voltage control actions," *IEEE Transactions on Power Systems*, vol. 13, no. 1, pp. 205-210, Feb. 1998.

[51] Z. Feng, V. Ajjarapu, and D. J. Maratukulam, "A comprehensive approach for preventive and corrective control to mitigate voltage collapse," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 791-797, May 2000.

[52] T. J. Overbye, "A power flow measure for unsolvable cases," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1359-1365, Aug. 1994.

[53] T. J. Overbye, "Computation of a practical method to restore power flow solvability," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 280-287, Feb. 1995.

[54] J. G. Rolim and L. J. Machado, "A study of the use of corrective switching in transmission systems," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 336-341, Feb. 1999.

[55] K. W. Hedman, R. P. O'Neill, E. B. Fisher, and S. S. Oren, "Optimal transmission switching with contingency analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1577-1586, Aug. 2009.

[56] A. A. Mazi, B. F. Wollenberg, and M. H. Hesse, "Corrective control of power system flows by line and bus-bar switching," *IEEE Transactions on Power Systems*, vol. 1, no. 3, pp. 258-264, Aug. 1986.

[57] K. Ma and J. Mutale, "Risk and reliability worth assessment of power systems under corrective control," in *Proceedings of 41st North American Power Symposium*, 2009, pp. 1-6.

[58] W. Shao and V. Vittal, "LP-based OPF for corrective FACTS control to relieve overloads and voltage violations," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1832-1839, Nov. 2006.

[59] N. G. Hingorani, "Flexible AC transmission," *IEEE Spectrum*, vol. 30, no. 4, pp. 40-45, Apr. 1993.

[60] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Boston, MA: Kluwer Academic Publishers, 2003.

[61] R. A. Schlueter, J. C. Lo, T. Lie, T. Y. Guo, and I. Hu, "A fast accurate method for midterm transient stability simulation of voltage collapse in power systems," in *Proceedings of the 28th IEEE Conference on Decision and Control*, 1989, pp. 340-344.

[62] Y. Chen, A. Bose, "Security analysis for voltage problems using a reduced model," *IEEE Trans. on Power Systems*, vol. 5, no. 3, pp. 933-940, Aug. 1990.

[63] R. A. Schlueter, I.-P. Hu, M.-W. Chang, J. C. Lo, and A. Costi, "Methods for determining proximity to voltage collapse," *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 258-292, Feb. 1991.

[64] D. G. Taylor and L. J. Maahs, "A reactive contingency analysis algorithm using MW and MVAR distribution factors," *IEEE Transactions on Power Systems*, vol. 6, no. 1, pp. 349-355, Feb. 1991.

[65] R. A. Schlueter, M. W. Chang, and A. Costi, "Loss of voltage controllability as a cause for voltage collapse," in *Proceedings of the 27th IEEE Conference on Decision and Control*, 1988, pp. 2120-2126.

[66] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data: Identification and analysis of coexpressed genes," *Genome Research*, vol. 9, no. 11, pp. 1106-1115, Nov. 1999.

[67] O. Dan and H. Mocian, "Scalable web mining with Newistic," in *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, 2009, pp. 556-63.

[68] P. M. Subcommittee, "IEEE reliability test system," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 6, pp. 2047-2054, Nov. 1979.

[69] B. Lesieutre, K. M. Rogers, T. J. Overbye, and A. Borden, "A sensitivity approach to detection of local market power potential," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 1980-1988, Nov. 2011.

[70] K. M. Rogers and T. J. Overbye, "Clustering of power system data and its use in load pocket identification," in *Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS)*, 2011, pp. 1-10.

[71] A. K. David and F. Wen, "Market power in electricity supply," *IEEE Transactions on Energy Conversion*, vol. 16, no. 4, pp. 352-360, Dec. 2001.

[72] G. Rothwell, T. Gómez, *Regulation and Deregulation: Electricity Economics.* Hoboken, NJ: John Wiley, 2003.

[73] B. C. Lesieutre, O. HyungSeon, R. J. Thomas, and V. Donde, "Identification of market power in large-scale electric energy markets," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS)*, 2006, pp. 240b-240b.

[74] B. Stott and E. Hobson, "Power system security control calculations using linear programming, part I," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 5, pp. 1713-1720, Sept. 1978.

[75] B. Stott and E. Hobson, "Power system security control calculations using linear programming, part II," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 5, pp. 1721-1731, Sept. 1978.

[76] B. Stott and J. L. Marinho, "Linear programming for power-system network security applications," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 3, pp. 837-848, May 1979.

[77] D. Cheverez-Gonzalez and C. L. DeMarco, "Admissible locational marginal prices via Laplacian structure in network constraints," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 125-133, Feb. 2009.

[78] B. C. Lesieutre, R. J. Thomas, and T. D. Mount, "Identification of load pockets and market power in electric power systems," *Decision Support Systems*, vol. 40, no. 3-4, p 517-528, October 2005.

[79] B. C. Lesieutre, R. J. Thomas, and T. D. Mount, "A revenue sensitivity approach for the identification and quantification of market power in electric energy markets," in *Proceedings of the IEEE Power Engineering Society General Meeting*, 2003, pp. 838-842.

[80] J. Nicolaisen, V. Petrov, and L. Tesfatsion, "Market power and efficiency in a computational electricity market with discriminatory double-auction pricing," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 5, pp.504-523, Oct. 2001.

[81] C. E. Murillo-Sanchez, S. M. Ede, T. D. Mount, R. J. Thomas, and R. D. Zimmerman, "An engineering approach to monitoring market power in restructured markets for electricity," in *Proceedings of the 24th Annual International Conference, International Association for Energy Economics*, 2001.

[82] D. Arthur and S. Vassilvitskii, "On the worst case complexity of the K-means method," Stanford InfoLab, Tech. Rep. 2005-34, 2005.

[83] D. Arthur, B. Manthey, and H. Röglin, "K-means has polynomial smoothed complexity," in *Proceedings of 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2009, pp. 405-414.

[84] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509- 517, Sept. 1975.

[85] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 209-226, Sept. 1977.

[86] C. E. Reeg and T. J. Overbye, "Algorithm development for non-intrusive load monitoring for verification and diagnostics," in *Proceedings of the North American Power Symposium (NAPS)*, 2010, pp. 1-5.

[87] J. C. Gower and G. J. S. Ross, "Minimum spanning trees and single linkage cluster analysis," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, vol. 18, no. 1, pp. 54-64, 1969.

[88] A. Yao. "On constructing minimum spanning trees in k-dimensional space and related problems," *SIAM Journal on Computing*, vol. 11, no. 4, pp. 721–736, 1982.

[89] F. J. Rohlf, "Algorithm 76: Hierarchical clustering using the minimum spanning tree," *The Computer Journal*, vol. 16, pp. 93-95, 1973.

[90] J. L. Bentley and J. H. Friedman, "Fast algorithms for constructing minimal spanning trees in coordinate spaces," *IEEE Transactions on Computers*, vol. 27, no. 2, pp. 97-105, Feb. 1978.

[91] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *The Computer Journal*, vol. 26, no. 4, pp. 354-358, 1983.

[92] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Systems Technical Journal*, vol. 36, pp. 1389-1401, 1957.

[93] H. Oberman and A. Weinberger, "Formal procedures for connecting terminals with a minimum total wire length," *Journal of the ACM*, vol. 4, no. 4, pp. 428-437, Oct. 1957.

[94] M. I. Shamos and D. Hoey, "Closest-point problems," in *Proceedings of the 16th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 151-162, 1975.

[95] Midwest ISO (MISO). (2011, Nov.). Real-time binding constraints. [Online]. Available: https://www.midwestiso.org/MarketsOperations/RealTimeMarketData/Pages/RealTimeBindingConstraints.aspx

[96] J.P. Corfmat and A.S. Morse, "Decentralized control of linear multivariable systems," *Automatica*, vol. 12, no. 5, pp. 479-495, Sept. 1976.

[97] North American Synchrophasor Initiative (NASPI). [Online]. Available: http://www.naspi.org

[98] openPDC, Phasor Data Concentrator. [Online]. Available: http://openpdc.codeplex.com/

[99] R. Bobba, E. Heine, H. Khurana, and T. Yardley, "Exploring a tiered architecture for NASPInet," in *Proceedings of the 1st IEEE PES Conference on Innovative Smart Grid Technologies (ISGT)*, 2010, pp. 1-8.

[100] K. M. Rogers, R. D. Spadoni, and T. J. Overbye, "Identification of power system topology from synchrophasor data," in *Proceedings of the IEEE Power Systems Conference and Exposition (PSCE)*, 2011, pp. 1-8.

[101] B. Stott and O. Alsac, "Fast decoupled load flow," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 859-869, May 1974.

[102] B. Stott, J. Jardim, and O. Alsac, "DC power flow revisited," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1290-1300, Aug. 2009.

[103] J. F. Dopazo, M. H. Dwarakanath, J. J. Li, and A. M. Sasson, "An external system equivalent model using real-time measurements for system security evaluation," *IEEE Transactions on Power Apparatus and Systems*, vol. 96, no. 2, pp. 431- 446, Mar. 1977.

[104] F. L. Alvarado, "Determination of external system topology errors," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no.11, pp. 4553-4561, Nov. 1981.

[105] E. J. S. Pires de Souza and A. M. Leite da Silva, "An efficient methodology for coherency-based dynamic equivalents," in *Proceedings of IEE Generation, Transmission and Distribution*, vol. 139, no. 5, pp. 371-382, Sept. 1992.

[106] J. B. Ward, "Equivalent circuits for power flow," *AIEE Transactions on Power Apparatus and Systems*, vol. 68, pp. 373-382, Jan. 1949.

[107] H. Oh, "A new network reduction methodology for power system planning studies," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 677-684, May 2010.

[108] X. Cheng andT. J. Overbye, "PTDF-based power system equivalents," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1868-1876, Nov. 2005.

[109] A. Bose, "Modeling of external networks for on-line security analysis," *IEEE transactions on power apparatus and systems*, vol. PAS-103, no. 8, pp. 2117-2123, Aug. 1984.

[110] A. C. Z. de Souza, J. C. S. de Souza, and A. M. L. da Silva, "On-line voltage stability monitoring," *IEEE Transactions on Power Systems*, vol. 15, no. 4, pp. 1300-1305, Nov. 2000.

[111] J. Machowski, "Dynamic equivalents for transient stability studies of electrical power systems," *International Journal of Electrical Power and Energy System*, vol. 7, no. 4, pp. 215-224, Oct. 1985.

[112] H. E. Brown, R. B. Shipley, D. Coleman, and R. E. Nied, Jr. "A study of stability equivalents," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-88, no. 3, pp. 200-207, March 1969.

[113] R. Podmore, "Identification of coherent generators for dynamic equivalents," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 4, pp. 1344-1354, Jul-Aug 1978.

[114] R. Nath, S. S. Lamba, and K. S. Prakasa Rao, "Coherency based system decomposition into study and external areas using weak coupling," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 6, pp. 1443-1449, June 1985.

[115] J. S. Lawler and R. A. Schlueter, "Computational algorithms for constructing modal-coherent dynamic equivalents," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 5, pp. 1070-1080, May 1982.

[116] J. H. Chow, A. Chakrabortty, L. Vanfretti, and M. Arcak, "Estimation of radial power system transfer path dynamic parameters using synchronized phasor data," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 564-571, May 2008.

[117] J. R. Winkelman, J. H. Chow, B. C. Bowler, B. Avramovic, and P. V. Kokotovic, "An analysis of interarea dynamics of multi-machine systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 2, pp. 754-763, Feb. 1981.

[118] B. Yang, V. Vittal and G. T. Heydt, "Slow coherency based controlled islanding – a demonstration of the approach on the August 14, 2003 blackout scenario," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1840-1847, Nov 2006.

[119] L. Wang, M. Klein, S. Yirga, and P. Kundur, "Dynamic reduction of large power systems for stability studies," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 889-895, May 1997.

[120] A. Ishchenko, J. M. A. Myrzik, and W. L. Kling, "Dynamic equivalencing of distribution networks with dispersed generation using Hankel norm approximation," *IET Generation, Transmission and Distribution*, vol. 1, no. 5, pp. 818-825, Sept. 2007.

[121] K. Chongsrid and S. Hara, "Hankel norm of sampled-data systems," *IEEE Transactions on Automatic Control*, vol. 40, no. 11, pp. 1939-1942, Nov. 1995.

[122] D. J. Trudnowski, A. Gentile, J. M. Khan, and E. M. Petritz, "Fixed-speed wind-generator and wind-park modeling for transient stability studies," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1911-1917, Nov. 2004.

[123] X. Feng, Z. Lubosny, and J. W. Bialek, "Identification based dynamic equivalencing," *IEEE Lausanne Power Tech*, 2007, pp. 267-272.

[124] H. K. Khalil, *Nonlinear systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[125] G. Kron, *Tensor Analysis of Networks*. New York, NY: John Wiley and Sons, 1939.

[126] G. Kron, *Diakoptics- The Piecewise Solution of Large-Scale Systems*. London: Macdonald & Co., 1963.

[127] S. Dutta, K. M. Rogers, T. J. Overbye, and J. Gronquist, "Estimation of transmission line parameters from historical data," submitted to *IEEE Transactions on Power Systems*.

[128] G. L. Kusic and D. L. Garrison, "Measurement of transmission line parameters from SCADA data," in *Proceedings of the IEEE PES Power Systems Conference and Exposition*, 2004, pp. 440- 445.

[129] J. Zhu and A. Abur, "Identification of network parameter errors," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 586- 592, May 2006.

[130] J. Zhu and A. Abur, "Improvements in network parameter error identification via synchronized phasors," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 44-50, Feb. 2010.

[131] J. F. Dopazo, G. Irisarri, and A. M. Sasson, "Real-time external equivalent for on-line contingency analysis," *IEEE Transactions on Power Systems*, vol. PAS-98, no. 6, pp. 2153-2171, Nov. 1979.

[132] A. S. Debs, "Estimation of steady-state power system model parameters," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 5, pp. 1260-1268, Sept. 1974.

[133] A. K. Al-Othman and M. R. Irving, "Uncertainty modelling in power system state estimation," in *IEE Proceedings- Generation, Transmission and Distribution*, vol. 152, no. 2, pp. 233- 239, March 2005.

[134] R. Bergen and V. Vittal, *Power System Analysis*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2000.

[135] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Mass.: Athena Scientific, 1999.

[136] D. P. Bertsekas, "Incremental least squares methods and the extended Kalman filter," *SIAM Journal on Optimization*, vol. 6, no. 3, pp. 807–822, Aug. 1996.

[137] H. Moriyama, N. Yamashita, and M. Fukushima, "The incremental Gauss-Newton algorithm with adaptive stepsize rule," *Computational Optimization and Applications,*vol. 26, no. 2, pp. 107-141, Nov. 2003.

[138] S. Grijalva, "Direct utilization of planning applications in the real-time environment," in *Proceedings of the IEEE Transmission and Distribution Conference and Exposition*, 2008.

[139] C. M. Marzinzik, S. Grijalva, and J. D. Weber, "Experience using planning software to solve real-time systems," in *Proceedings of the 42nd International Hawaii International Conference on System Sciences (HICSS)*, 2009, pp. 1-7.

[140] E. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.

[141] K. R. Gabriel, "The biplot graphical display of matrices with application to principal component analysis," *Biometrika*, vol. 58, no. 3, pp. 453-467, 1971.

[142] C. DeMarco, "Situational awareness, singular value methods for PMU data interpretation," presented at the PSERC Teleseminar Series, Mar. 2, 2010.

[143] C. M. Davis, J. E. Tate, and T. J. Overbye, "Wide area phasor data visualization," in *Proceedings of the 39th Annual North. American Power Symposium*, 2007, pp. 246-252.

[144] M. V. Venkatasubramanian, M. Kezunovic, and V. Vittal, "Detection, prevention and mitigation of cascading events," Power Systems Engineering Research Center (PSERC) final project report, pub. no. 08-18, Sept. 2008.

[145] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Transactions on Power Systems*, vol. 5, no. 1, pp. 80-89, Feb. 1990.

[146] M. J. Gibbard, N. Martins, J. J. Sanchez-Gasca, N. Uchida, V. Vittal, and L. Wang "Recent applications of linear analysis techniques," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 154-62, Feb. 2001.

[147] B. P. Lathi, *Linear Systems and Signals*, 2nd ed. New York, NY: Oxford University Press, Inc., 2005.

[148] G. E. Franklin, J. D. Powell, and A. Emami-Naeni, *Feedback Control of Dynamic Systems*, 5th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.

[149] E. M. Grabbe, S. Ramo, and D. E. Wooldridge, *Handbook of Automation, Computation and Control*, vol. 1, *Control Fundamentals*. New York, NY: John Wiley & Sons, Inc., 1958.

[150] V. Arcidiacono, E. Ferrari, R. Marconato, J. Dos Ghali, and D. Grandez, "Evaluation and improvement of electromechanical oscillation damping by means of eigenvalue-eigenvector analysis. Practical results in the central Peru power system," *IEEE Transactions on Power Apparatus and Systems*, vol. 1, PAS-99, no. 2, pp. 769-778, Mar/April 1980.

[151] MATLAB Central, Prony Toolbox. (2003, Sept. 13). [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/3955

[152] S. Singh, "Application of Prony analysis to characterize pulsed corona reactor measurements," M.S. Thesis, ECE Department, University of Wyoming, Aug. 2003.

[153] K. E. Bollinger and W. Eric Norum, "Time series identification of interarea and local generator resonant modes," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 273-279, Feb. 1995.

[154] L. Qi, L. Qian, S. Woodruff, and D. Cartes, "Prony analysis for power system transient harmonics," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, article ID 48406, 12 pages, 2007.

[155] D. J. Trudnowski, J. M. Johnson, and J. F. Hauer, "Making Prony analysis more accurate using multiple signals," *IEEE Transactions on Power Systems*, vol. 14, no. 1, Feb. 1999, pp. 226-231.

[156] M. A. Johnson, I. P. Zarafonitis, and M. Calligaris, "Prony analysis and power system stability-some recent theoretical and applications research," in *Proceedings of the IEEE Power Engineering Society Summer Meeting*, 2000, pp.1918-1923, vol. 3.

[157] D. J. Tudnowski, M. K. Donnelly, and J. F. Hauer, "Estimating the damping effectiveness of BPA's thyristor controlled series capacitor by applying time and frequency domain methods to measured response," *IEEE Transactions on Power Systems*, vol. 11, no. 2, 1996, pp. 761-766.

[158] A. J. Allen, S. Santoso, and W. M. Grady, "Voltage phase angle variation in relation to wind power," in *Proceedings of the IEEE Power and Energy Society General Meeting*, 2010, pp. 1-7.

[159] R. Doraiswami and W. Liu, "Real-time estimation of the parameters of power system small signal oscillations," *IEEE Transactions on Power Systems*, vol. 8, no. 1, pp. 74-83, Feb. 1993.

[160] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Champaign, IL: Stipes Publishing L.L.C., 1997.

[161] R. H. Yeu, "Small signal analysis of power systems: Eigenvalue tracking method and eigenvalue estimation contingency screening for DSA," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, Dec. 2010.

[162] J. M. Undrill and T. F. Laskowski, "Model selection and data assembly for power system simulations," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 9, pp. 3333-3341, Sept. 1982.

[163] IEEE Working Group on Computer Modelling of Excitation Systems, "Excitation system models for power system stability studies," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 2, pp. 494-509, Feb. 1981.

[164] Energy Development and Power Generating Committee of the Power Engineering Society, *IEEE Recommended Practice for Excitation System Models for Power System Stability Studies*, IEEE Std 421.5-1992, Mar. 1992.

[165] IEEE Power Engineering Society, Energy Development and Power Generating Committee, *IEEE Recommended Practice for Excitation System Models for Power System Stability Studies*, IEEE Std 421.5-2005, Apr. 2006.

[166] PowerWorld Simulator 15 Help File. (2010, June 11).Transient Stability Block Diagrams. [Online]. Available: http://www.powerworld.com/downloads/general.asp

[167] K. M. Rogers and T. J. Overbye, "Application of oscillation monitoring data for the detection of event signatures," in *Proceedings of IEEE Trondheim PowerTech*, 2011, pp. 1-7.

[168] J. J. Sanchez-Gasca and J. H. Chow, "Performance comparison of three identification methods for the analysis of electromechanical oscillations," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 995-1002, Aug. 1999.

[169] K. Steiglitz and L. E. McBridge, "A technique for the identification of linear systems," *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 461- 464, Oct. 1965.

[170] P. Stoica and T. Soderstrom, "The Steiglitz-McBride identification algorithm revisited-convergence analysis and accuracy aspects," *IEEE Transactions on Automatic Control*, vol. AC-26, no. 3, pp. 712-717, June 1981.

[171] H. Okamoto, J. J. Sanchez-Gasca, K. Clark, C. A. Wegner, N. W. Miller, and J. H. Chow, "Application of the Steiglitz-McBride identification algorithm to measured data from a power system simulator," in *Proceedings of the American Control Conference*, 1998, pp. 2956-2962.

[172] T. P. Krauss, L. Shore, and J. N. Little, Signal Procesing Toolbox, The MathWorks, Inc., 1995.

[173] J. D. Glover, M. S. Sarma, and T. J. Overbye, *Power system analysis and design*, 5th ed. Stamford, CT: Cengage Learning, 2012.

[174] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. Bellmont, MA: Athena Scientific, 2002.

[175] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 23rd ACM National Conference*, 1969, pp. 517-523.

[176] J. E. Tate and T. J. Overbye, "Contouring for power systems using graphical processing units," in *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)*, 2008.

[177] R. J. Renka, "Multivariate interpolation of large sets of scattered data," *ACM Transactions on Mathematical Software*, vol. 14, no. 2, pp. 139-148, Jun. 1988.

[178] Z. Huang, N. Zhou, F. Tuffner , D. Trudnowski, "Use of modal sensitivity to operating conditions for damping control in power systems," in *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011, pp. 1-9.

[179] Z. Huang et al., "MANGO - modal analysis for grid operation: A method for damping improvement through operating point adjustment," Pacific Northwest National Laboratory (PNNL), Technical Report, Prepared for the US Department of Energy, pub. no. PNNL-19890, Oct. 2010.

[180] J.S. Thorp, C.E. Seyler, and A.G. Phadke, "Electromechanical wave propagation in large electric power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 6, pp. 614–622, June 1998.

[181] R. Snieder and K. Wapenaar. "Imaging with ambient noise," *Physics Today*, vol. 63, no. 9, pp. 44-49, Sept. 2010.

[182] S. A. Zonouz, K. M. Rogers, W. H. Sanders, and T. J. Overbye, "CPIDS: A cyber-physical intrusion detection system for power-grid critical infrastructures," submitted to *IEEE Transactions on Smart Grid*.

[183] U.S. Department of Transportation, Federal Highway Administration, Office of Operations. (2006, Feb.). Simplified Guide to the Incident Command System for Transportation Professionals. [Online]. Available: http://www.ops.fhwa.dot.gov/publications/ics_guide/

[184] *Information technology -- Security techniques -- Entity authentication -- Part 1: General*, International Organization for Standardization and International Electrotechnical Commission ISO/IEC 9798-1:2010.

[185] H. Khurana, R. Bobba, T. Yardley, P. Agarwal, and E. Heine. "Design principles for power grid cyber-infrastructure authentication protocols," in *Proceedings of the Forty-Third Annual Hawaii International Conference on System Sciences*, 2010.

[186] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on DC state estimation," in *First Workshop on Secure Control Systems (SCS 2010)*, 2010.

[187] K. Morrow, E. Heine, K. M. Rogers, R. Bobba, and T. J. Overbye, "Topology perturbation for detecting malicious data injection," accepted to Hawaii International Conference on System Sciences (HICSS), Jan. 2012.

[188] A. Monticelli, F. F. Wu, and M. Yen, "Multiple bad data identification for state estimation by combinatorial optimization," *IEEE Power Engineering Review*, vol. PER-6, no. 7, pp. 73-74, July 1986.

[189] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions in Information and Systems Security*, vol. 14, no. 1, pp. 13:1-13:33.

[190] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 214-219.

[191] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010, pp. 1-6.

[192] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 220-225.

[193] H. Merrill and F. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-90, no. 6, pp. 2718-2725, Nov. 1971.

[194] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329-337, Mar 1975.

[195] D. Falcao, P. Cooke, and A. Brameller, "Power system tracking state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol PAS-101, no. 2, pp. 325-333, Feb. 1982.

[196] W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weighted least absolute value techniques applied to static state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 4, pp. 844-853, April 1982.

[197] X. Nian-de, W. Shi-ying, and Y. Er-keng, "A new approach for detection and identification of multiple bad data in power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 2, pp. 454-462, Feb. 1982.

[198] A. Monticelli and A. Garcia, "Reliable bad data processing for real-time state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 5, pp. 1126-1139, May 1983.

[199] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS- 103, no. 11, pp. 3239-3252, Nov. 1984.

[200] X. Nian-de, W. Shi-ying, and Y. Ers-keng, "An application of estimation-identification approach of multiple bad data in power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 2, pp. 225-233, Feb. 1984.

[201] F. Wu, W.-H. Liu, and S.-M. Lun, "Observability analysis and bad data processing for state estimation with equality constraints," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 541–548, May 1988.

[202] I. Slutske, "Bad data identification in power system state estimation based on measurement compensation and linear residual calculation," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 53–60, Feb 1989.

[203] A. Abur, "A bad data identification method for linear programming state estimation," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 894–901, Aug 1990.

[204] B. Zhang, S. Wang, and N. Xiang, "A linear recursive bad data identification method with real-time application to power system state estimation," *IEEE Transactions on Power Systems*, vol. 7, no. 3, pp. 1378–1385, Aug 1992.

[205] E. Asada, A. Garcia, and R. Romero, "Identifying multiple interacting bad data in power system state estimation," in *Proceedings of the IEEE Power Engineering Society General Meeting*, 2005, pp. 571–577.

[206] G. Peters and J. H. Wilkinson, "The least squares problem and pseudo-inverses," *The Computer Journal*, vol. 13, no. 3, pp. 309-316, Aug. 1970.

[207] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *IEEE Transactions on Power Systems*, vol.21, no.4, pp.1608-1615, Nov. 2006

[208] J. B. A. London, L. F. C. Alberto, and N. G. Bretas, "Analysis of measurement-set qualitative characteristics for state-estimation purposes," *IET Generation, Transmission & Distribution*, vol. 1, no. 1, pp. 39-45, Jan. 2007.

[209] W. Peterson and A. Girgis, "Multiple bad data detection in power system state estimation using linear programming," in *Proceedings of the Twentieth Southeastern Symposium on System Theory*, 1988, pp. 405-409.

[210] S. A. Zonouz and K. M. Rogers, "Towards designing power control network intrusion detection systems," presented at TCIPG meeting, March 17, 2011.

[211] S. R. Eddy, "Hidden Markov models," *Current Opinion in Structural Biology*, vol. 6, no. 3, pp. 361-365, June 1996.

[212] G. D. Forney, Jr.; "The Viterbi algorithm," in *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268-278, 1973.

[213] A. Monticelli, F. F. Wu, and M. Yen, "Multiple bad data identification for state estimation by combinatorial optimization," *IEEE Power Engineering Review*, vol. PER-6, no. 7, pp. 73-74, July 1986.

[214] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,"*IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12 –19, Feb. 2011.

[215] D. Divan, W. Brumsickle, R. Schneider, B. Kranz, R. Gascoigne, D. Bradshaw, M. Ingram, and I. Grant, "A distributed static series compensator system for realizing active power flow control on existing power lines," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 642-649, Jan. 2007.

[216] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proceedings of the 47th Annual Allerton Conference on Communication Control and Computing*, pp. 911-918, 2009.

[217] J. Hauer, W. Mittelstadt, K. Martin, J. Burns, H. Lee, J. Pierre, and D. Trudnowski, "Use of the WECC WAMS in wide-area probing tests for validation of system performance and modeling," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 250-257, Feb. 2009.

[218] N. Zhou, J. Pierre, and J. Hauer, "Initial results in power system identification from injected probing signals using a subspace method," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1296–1302, Aug 2006.

[219] D. Trudnowski and T. Ferryman, "WECC modal baseline analysis results as of June 2010," NASPI Technical report, June 2010.

[220] D. Trudnowski, J. Pierre, N. Zhou, F. Tuffner, J. Hauer, and B. Mittelstadt, "WECC dynamic probing tests: Purpose and results," NASPI Technical report, October 2008.

[221] J. Pierre, N. Zhou, F. Tuffner, J. Hauer, D. Trudnowski, and W. Mittelstadt, "Probing signal design for power system identification," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 835-843, May 2010.

[222] J. Hauer and J. DeSteese, "A tutorial on detection and characterization of special behavior in large electric power systems," Pacific Northwest National Laboratory Technical report, pub. no. PNL-14655, July 2004.

[223] A. Aquino-Lugo and T. Overbye, "Agent technologies for control application in the power grid," in *Proceedings of the 43rd Hawaii International Conference on System Sciences*, 2009.

[224] A. A. Aquino-Lugo, "Distributed and decentralized control of the power grid," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, Dec. 2010.