The University of Maine [DigitalCommons@UMaine](http://digitalcommons.library.umaine.edu?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages)

[Electronic Theses and Dissertations](http://digitalcommons.library.umaine.edu/etd?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages) [Fogler Library](http://digitalcommons.library.umaine.edu/fogler?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages)

Summer 8-11-2016

Identification of Critical Locations and Reduced Model State Estimation for Power System Analysis

Amamihe Onwuachumba *University of Maine, Orono*, amamihe.onwuachumba@maine.edu

Follow this and additional works at: [http://digitalcommons.library.umaine.edu/etd](http://digitalcommons.library.umaine.edu/etd?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages) Part of the [Power and Energy Commons](http://network.bepress.com/hgg/discipline/274?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages)

Recommended Citation

Onwuachumba, Amamihe, "Identification of Critical Locations and Reduced Model State Estimation for Power System Analysis" (2016). *Electronic Theses and Dissertations*. 2503. [http://digitalcommons.library.umaine.edu/etd/2503](http://digitalcommons.library.umaine.edu/etd/2503?utm_source=digitalcommons.library.umaine.edu%2Fetd%2F2503&utm_medium=PDF&utm_campaign=PDFCoverPages)

This Open-Access Dissertation is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine.

IDENTIFICATION OF CRITICAL LOCATIONS AND REDUCED MODEL STATE ESTIMATION FOR POWER SYSTEM ANALYSIS

By

Amamihe Stanislaus Akobundu Onwuachumba

B.S. Moscow Power Engineering Institute (Technical University), 2008

M.S. Moscow Power Engineering Institute (Technical University), 2010

M.S. University of Maine, 2013

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Electrical and Computer Engineering)

The Graduate School

The University of Maine

August 2016

Advisory Committee:

Mohamad T. Musavi, Professor of Electrical and Computer Engineering, Advisor Carroll Lee, Former President of Bangor Hydro Electric (now Emera Maine) Paul Lerley, Senior Engineer of RLC-Engineering, LLC Bruce Segee, Professor of Electrical and Computer Engineering Paul Villeneuve, Professor of Electrical Engineering Technology Yifeng Zhu, Professor of Electrical and Computer Engineering

DISSERTATION ACCEPTANCE STATEMENT

On behalf of the Graduate Committee for Amamihe Onwuachumba I affirm that this manuscript is the final and accepted dissertation. Signatures of all committee members are on file with the Graduate School at the University of Maine, 42 Stodder Hall, Orono, Maine.

__

Mohamad T. Musavi, Professor of Electrical and Computer Engineering Date

© 2016 Amamihe Onwuachumba

All Rights Reserved

LIBRARY RIGHTS STATEMENT

In presenting this dissertation in partial fulfillment of the requirements for an advanced degree at The University of Maine, I agree that the Library shall make it freely available for inspection. I further agree that permission for "fair use" copying of this dissertation for scholarly purposes may be granted by the Librarian. It is understood that any copying or publication of this dissertation for financial gain shall not be allowed without my written permission.

Signature:

Date:

IDENTIFICATION OF CRITICAL LOCATIONS AND REDUCED MODEL STATE ESTIMATION FOR POWER SYSTEM ANALYSIS

By Amamihe Stanislaus Akobundu Onwuachumba

Dissertation Advisor: Dr. Mohamad T. Musavi

An Abstract of the Dissertation Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (in Electrical and Computer Engineering) August 2016

In order to reduce the carbon footprint and the cost of electric energy, the owners of electric power utilities today are faced with the task of reducing the use of expensive and carbon intensive fossil fuels and significantly increasing the amount of energy from renewable sources in their grids while meeting an increase in electricity demand. To deal with increase in demand, electric utilities operate very close to their maximum capacities and this sometimes results in violating security limits. Therefore, the integration of intermittent renewable energy into the utility grids poses serious concerns that must be addressed to ensure grid stability.

In order to improve monitoring of their system, utilities are increasing the number of measurement devices in the system. However, not all collectible data contain important, necessary or unique information about the system, so storing and analyzing them comes at a considerable financial cost to the company. Therefore, identifying parts of the system whose measurements provide information that reflects the general state of the system would help utilities smartly utilize resources.

In this dissertation, a methodology for the identification of critical variables of power systems and their locations using eigenvalue analysis of the measurements of the system variables is developed. This analysis is based on principal component analysis (PCA). The effectiveness of monitoring critical locations of a power system in ensuring steady state system security is demonstrated.

Also, an artificial neural network-based state estimator that utilizes data from regular measurement units and phasor measurement units (PMUs) placed at the critical locations is developed. A technique called state estimation is used to estimate measured and unmeasured electrical quantities. Conventional state estimation techniques require availability of many measurements. The proposed state estimator utilizes fewer measurements, leading to a reduction in the number of expensive PMUs needed and reduction in the cost of electric grid operation. Thus, electric power utilities would be able to assess the state of their grid efficiently and improve their ability to integrate renewable energy without violating the grid's security constraints.

DEDICATION

TO MY FAMILY AND MY LOVELY WIFE.

I LOVE YOU ALL DEARLY.

ACKNOWLEDGMENTS

I would like to thank all the individuals who have contributed to my research at the University of Maine. Firstly, I want to thank my advisor Prof. Mohamad Musavi for making my PhD experience productive and inspiring through his priceless contributions of time, ideas, fatherly advice and financial support.

I would also like to thank my committee members, Mr. Carroll Lee, Mr. Paul Lerley, Professor Bruce Segee, Professor Paul Villeneuve and Professor Yifeng Zhu for their brilliant comments, ideas and suggestions. I greatly appreciate other faculty and staff in the Electrical and Computer Engineering Department and the Graduate School for their support and encouragement during my study.

I am very thankful to the entities that provided financial and technical support that made my research feasible. Many thanks to Iberdrola USA-Central Maine Power for providing an avenue for feedback on the quality and value of my research, and to Underground Systems Inc and the US Department of Energy for funding the initial stages of my PhD work. I gratefully acknowledge RLC Engineering, LLC, for helping me develop a sound knowledge and understanding of the power system, for providing logistic support and funding for most part of my research. Special thanks to Paul Lerley, Mike Poulin and Rick Conant of RLC Engineering for their generous contribution of their time and wealth of experience. My sincere thanks to the Graduate Student Government (GSG) for providing travel funds to present my research at conferences.

Past and present Maine Smart Grid Lab members at the University of Maine were very supportive during the course of my study. I am grateful to Shengen Chen, Yunhui

Wu, Aseem Rambani, Matthew Edwards and Qi Li for sharing their knowledge and experience. I also want to thank my friends for providing me with the emotional support I needed to get through the difficult periods of my research. Special thanks to Matthew Valles, Tesfahiwet Zerayesus, Mussie Beyene, Fidel Odunze, Amma Amponsah, Tega Dibie and Ruona Dibie.

My time at the University of Maine was greatly enriched by the Newman Center, National Society of Black Engineers (NSBE), GSG, International Student Association (ISA) and African Student Association (AFSA). I will be forever grateful for the opportunities provided by these organizations to acquire and develop strong leadership and professional skills.

I want to express my warmest appreciation to my family for all their love and encouragement, especially my mom Rosemary for her unwavering support and encouragement all through my study. I am grateful in a special way to my loving, supportive and patient wife Chinonye for her support and encouragement during the final stages of my PhD.

Finally, I am immensely grateful to Almighty God for his faithfulness, for guiding me through all my difficulties and for making my dream come true.

TABLE OF CONTENTS

Chapter

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Motivation

Electrical power systems comprise of a network of electrical components designed for supplying, transmitting and using electric energy. These components include electric power generators, transmission lines and loads. Over the years, smaller systems have found it more beneficial to interconnect with neighboring systems. Some of the benefits of interconnection of neighboring utilities are improvement in system security and economy of operation [1]. Improved security stems from the mutual emergency assistance that the utilities can communally provide, while improved economy comes from the need to have less generating reserve capacity on each system. As the number of interconnections continues to grow, the size of the interconnected system expands. The power system becomes more complex and the monitoring and control of such a system becomes more challenging.

In order to rise up to the challenge of monitoring and controlling complex power systems, analytical tools such as power system state estimation and observability analysis were developed starting in the 1970's. These tools are highly valuable today, especially given that some of the major blackouts in recent history, such as the New York power outage of 1987, might have been prevented had state estimation been employed in those systems at that time [2]. Today, state estimation is the foundation on which modern power system control centers are built. Besides, it is the basis for the creation and operations of all markets, real time and otherwise, in electric power systems. The

theoretical background on Power System State Estimation and Observability Analysis will be provided in Section [2.1](#page-26-0) below.

Power systems encounter outages quite frequently. Some of these disturbances are initiated locally in one area and cascade over to large geographic areas far away from the initial starting point of the disturbance. Although system-wide disturbances that affect numerous customers in a large geographic area rarely take place, they are more common than a normal distribution of probabilities would predict. Approximately 10 significant outages have occurred in North America since 1965 and with the benefit of hindsight many of the blackouts could have been prevented. One of the factors recognized to be prevalent in these major outages is the inability of system operators to visualize events on the entire system [3].

In order to improve visualization of the system, utility companies increase the number of measurement devices such as conventional metering devices and phasor measurement units (PMUs) in their systems. In fact, PMU deployment has received unprecedented momentum due to recent high profile blackouts [4]. Installing more measurement devices in the system means more financial commitment for procurement, calibration and maintenance of devices; more data for a system operator to decipher, especially during an emergency; and more long term storage resources requirements.

However, not all collectible data contain important, necessary or unique information about the system, so storing and analyzing them comes at a considerable financial cost to the utility in the long run. This problem can be mitigated by identifying parts of the system whose measurements provide information that reflect the general state of the system. Identifying these critical locations of the power system will enable the utility to smartly utilize resources. This includes: prioritization of the measurement units in these areas for maintenance and calibration; procurement of backup units for these locations in case of failure; prioritization of measurements from the units in these locations for steady state monitoring and control of the system; and prioritization of these locations for PMU deployment (for systems without PMUs).

It is noteworthy that a number of factors could influence the citing of PMUs [5] but many PMU placement methods focus on a specific power system application. In reference [6] a method for line parameter estimation was discussed, references [7] and [8] examined approaches tailored for dynamic vulnerability assessment, whereas references [9] and [10] reported techniques focused on state estimation. Algorithms based on economic concerns were considered in references [11] and [12]. Deese et al [12] compared several optimal PMU placement algorithms designed to minimize implementation cost. They project a continued combined use of PMUs and other measurement devices (smart meters), so long as the cost of PMUs remains considerably higher than that of other meters. The authors in reference [13] discussed an approach useful in the identification of multiple power line outages.

However, some PMU placement algorithms in recent literature consider multiple objectives. For example, reference [14] proposed a multi-criteria approach considering fault analysis, voltage control and state estimation. This approach was implemented using an integer linear programming method. The authors used "a fixed and exclusive" method to create multiple solutions in order to accommodate several applications with a minimal

number of PMUs. Another approach in reference [15] is tailored to observability analysis and bad data detection.

Although some of these algorithms have similar objectives to the proposed methodologies in this dissertation, they do not consider that measurements in certain (critical) locations in the system are more reflective of the changes occurring in the whole system.

Experienced power systems engineers or system operators might be able to identify the critical parts of a particular system due to their experience and knowledge of the system. However, the addition / retirement of certain elements of the system could cause previously critical parts of the system to be less critical or even unimportant, and vice versa. So relying solely on experience to identify critical parts of a system could undermine the accuracy of such an exercise. Furthermore, an engineer who has no prior knowledge of a power system will be unable to identify the critical locations.

1.2. Goal

The main goal of this research is to identify critical locations of power systems in order to assess the steady state security status of the systems using synchrophasor and / or regular measurements placed at the critical locations. To achieve this, the specific objectives are:

a) The development of a systematic methodology for identifying the critical variables, and hence, the critical locations, of any given power system.

- b) The demonstration of the effectiveness of monitoring the critical locations of the power system in ensuring system security.
- c) The development of an artificial neural network (ANN) based reduced model state estimation tool for power system analysis using the identified critical variables.

1.3. Major Contributions

The main contributions of this dissertation are:

- a) A scientific method for the identification of critical variables and their locations in a power system. The only existing option in the power system industry for determination of critical locations of power systems is the intuition of experienced power system engineers or system operators. However, when elements are added to or retired from the system previously critical parts of the system may become less critical or even unimportant, and vice versa. The proposed method introduces a systematic methodology and eliminates the errors that could result from guess work, especially by inexperienced engineers. This approach is based on eigenvalue analysis of the power system.
- b) The effectiveness of monitoring the critical locations of the power system. The main aim of identifying critical locations of a system is to provide the system operator a concise number of locations that reflect the security status of the entire power system. This dissertation proposes monitoring just the identified set of critical locations and demonstrates its effectiveness. This

makes it easier for the system operator to focus on important information, especially during an emergency.

- c) Identification of locations for phasor measurement unit (PMU) placement for steady state monitoring and control. PMUs are state of the art measurement devices used in the power system industry. Given the effectiveness of monitoring the critical locations described in [b\)](#page-21-0) above this dissertation proposes prioritizing the identified critical locations of a system for PMU placement (for utilities that have not yet installed PMUs in their systems), prioritization of the measurement units in these areas for maintenance and calibration, and procurement of backup units for these locations in case of failure.
- d) An ANN-based state estimation tool. Conventional state estimators require a lot of measurements to be made, and hence, a lot of computation resources, and the unavailability of data can have a significant impact on accuracy of the state estimation solution. This dissertation proposes an ANN-based state estimator that uses measurements from the critical variables of the system to estimate the rest of the power system variables. This state estimator uses fewer measurements, and hence requires much less computational resources.

1.4. Power Systems in the Dissertation

The power systems used in this dissertation are a General Electric (GE) 6-bus system and IEEE test systems. IEEE test systems are standardized systems that provide a benchmark for comparing results of different methodologies or algorithms. The IEEE systems used in this dissertation are the 14-bus and 118-bus systems.

1.5. Definition of Terms used in the Dissertation

Base case: this is the original mathematical model of a given system developed for a specific study. Load flow cases for scenarios investigated in the study are derived from this base case.

Dispatch: a dispatch is a variation of the combination of generators that are online or offline in a given load flow case. Every dispatch must observe the law of conservation of energy, that is, if a generator or group of generators' output changes, another generator or group of generators must be adjusted to accommodate this change provided load is constant. For instance, if a generator is taken offline, another generator or group of generators must supply the total amount of power output previously supplied by the offline generator.

Mathematical model: a mathematical model of a system is a numerical representation of the system. It is the set of data comprising of the values of elements of the system such as the impedance of transmission lines, the voltage magnitudes and phase angles of each bus in the system, the real and reactive power of generators and loads. It also contains information on the connections within the system, the subdivisions (areas) in the system, and an equivalent representation of neighboring systems, if applicable.

Regional system coordinating body: this is usually a non-profit organization that coordinates the activities of utility companies within a defined territory. It is responsible

for facilitating the buying and selling of electric power and ensuring the reliability of electric power supply in its territory. Examples are the Independent System Operator of New England (ISO-NE) and the New York Independent System Operator (NYISO). Regional system coordinating bodies in North America are often called regional transmission organizations (RTOs) or independent system operators (ISOs).

Substation: a substation is a basic part of the power system, where voltage is transformed from high to low, or vice-versa. At distribution substations, voltage is transformed from high to low and distributed to consumers. At sub-transmission and transmission substations, voltage is transformed from low to high and transmitted to other areas where the energy may be needed.

System security: a system is secure if there are no violations on any of its operational constraints. The operational constraints of an electric power system are upper and lower limits of bus voltage magnitudes, and limits on transmission line flows.

Utility company: utility companies may generate, transmit or distribute electricity. Most utilities perform only one of the above functions due to deregulation of the electric utility industry. Examples are Bangor Hydro-Electric Company (BHE) and Central Maine Power (CMP).

1.6. Dissertation Organization

Chapter 2 introduces the concepts used in this dissertation. These include state estimation and artificial neural networks.

Chapter 3 offers a proof of concept of state estimation using underdetermined system of equations, and provides preliminary results for the presented aspects of the methodology.

In Chapter 4, the methodology for identification of critical variables and their locations is presented.

Chapter 5 contains the implementation of the proposed methodology on a small power system.

Chapter 6 features the implementation of the proposed methodologies on a largescale power system; it illustrates the effectiveness of monitoring critical locations of power systems and the capability of the ANN-based state estimator to estimate the values of the voltage magnitude and phase angles of all the buses of the system using the identified critical variables.

Chapter 7 draws conclusions to the major contributions of this dissertation.

CHAPTER 2

BACKGROUND

2.1. State Estimation and Observability Analysis

Prior to the introduction of state estimation in the 1970's power systems were monitored only by supervisory control systems [2]. These systems monitored and controlled the status of circuit breakers at the substations. These systems were later upgraded with real-time system-wide data acquisition capabilities that allowed the control centers to obtain analog measurements and circuit breaker status data from the power system. These were called Supervisory Control and Data Acquisition (SCADA) Systems [2]. Knowledge of the real-time operating conditions of the power system facilitated the execution of application functions like contingency analysis, and corrective real and reactive power dispatch.

However, the information provided by SCADA systems was not always reliable as a result of errors in the measurements, telemetry failures, or communication noise. Also, the collected information may not allow for directly extracting the corresponding alternating current (AC) operating state of the system. In addition to the aforementioned issues, it was often cost prohibitive to telemeter all possible measurements even when they were available at the substations.

With the advent of state estimation the capabilities of SCADA system computers expanded and this led to the establishment of Energy Management Systems (EMS) [2]. State estimators enable accurate and efficient monitoring of operational constraints on quantities like bus voltage magnitude or transmission line power flow. They ensure a reliable real-time data base of the system including the existing state.

State estimators often include a range of functions some of which are briefly described here. A topology processor gathers status data about the circuit breakers and switches and configures the system diagram. Network observability analysis is used to determine if a state estimation solution for the entire system can be obtained using the available set of measurements. State estimation solution derives the optimal estimate of the system state based on the network model and obtained measurements from the system; the system state variables are the voltage magnitudes and phase angles. A bad data processor detects the existence of gross errors in the measurement set, identifies the bad measurement and may eliminate them given enough redundancy in the measurement set [2].

State estimation plays a very important role in enabling continuous monitoring of the power system. Its major function is to provide a clean set of data for use by various application functions such as contingency and power flow analysis. Traditionally, to ensure that a state estimation solution is obtainable observability analysis of the entire system needs to be done [2]. Network observability analysis relies heavily on the number, type and relative positions of the available measurements in the system.

Network observability analysis is a means of determining if a given set of available measurements is sufficient to obtain a unique estimate for the power system state [2]. This analysis is usually carried out during the planning and/or upgrade stages of the power system, or just before running the state estimator. It is largely dependent on the topology of the system and on the type and location of the available measurements. A power system is said to be observable if, given a set of measurements, the state estimator is able to determine all the system state variables, typically the voltage magnitudes and phase angles of all the buses in the system. Otherwise, the system is not observable.

Static state estimation constitutes the use of redundant measurements to minimize measurement errors thereby obtaining an optimum estimate of the system state. The most commonly used function for state estimation is the weighted least square method. Given that the measurement model is:

$$
z = h(x) + e
$$

\n
$$
\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} h_1(x_1, x_2, ..., x_N) \\ h_2(x_1, x_2, ..., x_N) \\ \vdots \\ h_m(x_1, x_2, ..., x_N) \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}
$$
 (1)

$$
x = [V, \theta] \tag{2}
$$

where *z* is the measurement vector, $h(x)$ is the vector of the non-linear relationship between measurements and state variables, *x* is the vector of the state variables, *e* is the vector of measurement errors, *m* is the number of measurements, *N* is the number of buses in the system and *V* and θ represent the voltage magnitudes and phase angles of all the buses in the system. In the most general case, the measurement vector comprises the following types of measurements:

$$
z = \left[P_{inj}, P_{flow}, Q_{inj}, Q_{flow}, I_{mag}, V \right]^T = v(x) \tag{3}
$$

where P_{inj} is the real power injection at a given bus, Q_{inj} is the reactive power injection at a given bus, P_{flow} is the real power flow between two buses, Q_{flow} is the reactive power flow between two buses, I_{mag} is the line current flow magnitude between two buses, *V* is the voltage magnitude at a given bus, and $v(x)$ is a matrix of nonlinear functions (comprising an over-determined system of equations) mapping measurements to state variables. A sample measurement set for a 3-bus system is shown in [Figure 2.1.](#page-29-0) Note that measurements are placed at the buses and transmission lines that guarantee observability of the system, that is, not at every bus and transmission line in the system. Oftentimes variables in the measurement vector (z) are also a subset of the state variables (*x*), for instance, the bus voltage magnitude (*V*).

Figure 2.1. A sample measurement set

The objective of the Weighted Least Square (WLS) estimation is to find *x* such that the cost function below is minimized [2].

$$
E(x) = (z - h(x))^{T} W (z - h(x))
$$
 (4)

where *W* is a matrix of weights consisting of the reciprocal of the covariance matrix of measurement errors. This objective is achieved by expanding the derivative of the cost function into a Taylor series with the vector of the state variables as the operation point [2].

$$
g(x) = \frac{\partial E(x)}{\partial x} = \frac{\partial h(x)}{\partial x}^T \mathbf{W}\left(z - h(x)\right) = 0 \tag{5}
$$

$$
g(x) = g(xo) + \frac{\partial g(xo)}{\partial x}(x - xo) + \dots = 0
$$
 (6)

The optimal state estimate is then obtained by solving equation [\(6\)](#page-30-0) through the iterative Gauss-Newton method [2]. A suitable result is found when the change in the state variable values ∆*x* is within a preset tolerance. Equation [\(7\)](#page-30-1) shows the iterative process.

$$
\Delta x^{k+1} = [H(x^k)]^{-1} J^T(x^k) W[z - h(x^k)] = [H(x^k)]^{-1} g(x^k)
$$
\n(7)

where $J(x)$ is the measurement Jacobian matrix, which is composed of an overdetermined set of equations and in the most general case can be represented thus:

$$
J(x) = \frac{\partial h(x)}{\partial x}
$$

$$
J(x) = \begin{bmatrix} \frac{\partial P_{inj}}{\partial v} & \frac{\partial P_{inj}}{\partial \theta} \\ \frac{\partial P_{flow}}{\partial v} & \frac{\partial P_{flow}}{\partial \theta} \\ \frac{\partial Q_{inj}}{\partial v} & \frac{\partial Q_{inj}}{\partial \theta} \\ \frac{\partial Q_{flow}}{\partial v} & \frac{\partial Q_{flow}}{\partial \theta} \\ \frac{\partial I_{mag}}{\partial v} & \frac{\partial I_{mag}}{\partial \theta} \\ \frac{\partial V_{mag}}{\partial v} & 0 \end{bmatrix}
$$
 (8),

 $H(x)$ is the gain matrix defined by:

$$
H(x) = \frac{\partial g(x)}{\partial x} = J^T(x)WJ(x)
$$
 (9),

$$
\Delta x^{k+1} = x^{k+1} - x^k \tag{10}
$$

and *k* is the iteration index. More details on this method are given in [2].

For over four decades much research work has been done in state estimation, notable among which are references [16], [17], [18], [19], [20] and [21]. In what shall be called the "2N approach" in this dissertation, two of the four variables $(P, Q, V \text{ and } \theta)$ of a bus in a power system with *N* buses are used to find the remaining two. The disadvantages of using this approach are 1) cost-intensiveness since many measurements need to be made and 2) significant impact that potential data unavailability can have on accuracy of the state estimation solution [2]. As a result of the large number of measurements required for state estimation, which is often based on the iterative weighted least square method, significant computation resources, including time and memory, are needed. This is one reason state estimation is run in intervals of roughly 15 minutes [22], hence, making it unsuitable for smart grid applications where it is desirable to have an estimate of the system state at intervals in the order of a few seconds.

In this dissertation an approach that reduces the number of measurements used in estimating the state of a power system shall be introduced. The number of measurements is less than the minimum 2*N*-1 variables that are required in a conventional state estimation approach. For the IEEE 14-bus system in [CHAPTER 5](#page-66-0) the number of measurements used in the proposed approach is 78% of the minimum 27 measurements required by conventional state estimators, whereas for the IEEE 118-bus system in [CHAPTER 6](#page-75-0) the number of measurements used in the proposed approach is 64% of the minimum 235 measurements required by conventional state estimators.

2.2. Artificial Neural Networks

The concept of artificial neural networks (ANNs) is modeled after the information-processing system of the human brain, which has been described as a complex, nonlinear and parallel computer [23]. ANNs are roughly analogous to a brain in many ways. Just like a brain has neurons, ANNs have processing units – a massive interconnection of simple computing cells – that can store experiential knowledge and make this knowledge available for later use. This knowledge is acquired by the network from its environment through a learning process. Interneuron connection strengths, known as synaptic weights are used to store the acquired knowledge.

The learning process for a neural network is performed using a procedure known as a learning algorithm. The function of a learning algorithm is to orderly modify the synaptic weights of the network for the purpose of achieving a desired objective. One such approach is analogous to linear adaptive filtering.

ANNs capable of generalization, which is the ability of the network to produce reasonable results for inputs not encountered during the training (learning) stage. These qualities help ANNs to solve complex and intractable problems.

The use of ANNs provides numerous benefits. Given that neural networks could be nonlinear and this nonlinearity is distributed throughout the network, they are able to operate with nonlinear functions. They are also able to adapt their synaptic weights to changes in the surrounding environment, i.e., a neural network trained to operate in a specific environment can be easily retrained to handle minor changes in the operating environmental conditions.

ANNs have been applied to state estimation in the past: [24], [25] and [26]. However, these approaches are essentially similar to the *2N* approach. ANNs provide an excellent tool for inexpensively – in terms of memory and time – implementing this minimized state estimation procedure. This makes ANN an ideal candidate for state estimation since it can accurately map the relationship between the measured variable and other state variables of the power system.

A multilayer ANN was used in this application in order to accurately capture the nonlinearity in power system parameters data. As shown in [Figure 2.2,](#page-34-0) an ANN comprises of input variables, hidden layers and an output layer. The most commonly used algorithm employed to train this network is the error back-propagation (BP) algorithm, which is an adaptive filtering algorithm [23]. This algorithm consists of two data passes, 1) the forward pass during which the synaptic weights of the network are fixed, and 2) the backward pass when the weights are changed according to an error-correction rule. The weights of the network are updated in such a way as to follow the negative of the gradient of the error between the actual output (target values) and the calculated output of the network with respect to the network weights and biases.

Figure 2.2. Multilayer ANN

The output error signal of the *i*th neuron during the *k*th iteration is given by:

$$
e_i(k) = d_i(k) - x_i(k) \tag{11}
$$

where $d_i(k)$ is the desired value for the output signal and $x_i(k)$ is the output value computed by the ANN. The instantaneous value of the error energy for each neuron is one-half of the square of the error value. The system total error energy is therefore defined as:

$$
E(k) = \frac{1}{2} \sum e_i^2(k)
$$
 (12)

$$
E_{av} = \frac{1}{N} \sum_{k=1}^{N} E(k)
$$
 (13)

where *N* is the total number of patterns in the training data set. Since the error surface of a multilayer perceptron is a highly non-linear function of the synaptic weight vector *w*, the cost function $E_{av}(w)$ is expanded about the operating point $w(k)$ using the Taylor series:

$$
E_{av}(w(k) + \Delta w(k)) = E_{av}(w(k)) + g^T(k)\Delta w(k) + \frac{1}{2}\Delta w^T(k)H(k)\Delta w(k) + \cdots
$$
\n(14)

where $g(k)$ is the local gradient vector given as:

$$
g(k) = \frac{\partial E_{av}(w)}{\partial w}\Big|_{w = w(k)}\tag{15}
$$

and $H(k)$ is the local Hessian matrix defined as:

$$
H(k) = \frac{\partial g(k)}{\partial w} = \frac{\partial^2 E_{av}(w)}{\partial w^2}\Big|_{w = w(k)}
$$
(16)

The optimum value of the increment $\Delta w(k)$ added to the weight vector $w(k)$ can be found using

$$
\Delta w(k) = H^{-1}(k)g(k) \tag{17}
$$

More details on this method can be found in [23].

2.3. Comparison of Conventional State Estimation and Artificial Neural Networks

Equation [\(17\)](#page-35-0) is very similar to equation [\(7\)](#page-30-1); this is attributable to the fact that both approaches are optimization methods that make use of the first and second derivatives of the cost function to minimize errors. However, the conventional state estimation (CSE) is only applicable to systems whose measurement Jacobians are fully
determined or over-determined, whereas ANNs are applicable to fully determined, overdetermined and more importantly, under-determined systems, with a very high accuracy level. State estimation using underdetermined system of equations shall be presented in [CHAPTER 3.](#page-39-0)

In the particular problem of determining the estimates of a power system's states for the preliminary case studies presented in Section [3.1,](#page-43-0) inputs to the neural network consist of only the load buses' real and reactive powers while the outputs are the voltage magnitude and phase angles of all the buses. The ANN maps the nonlinear function described by equation [\(25\)](#page-41-0) and directly finds state variables as functions of load real and reactive powers. However, CSE must have as inputs a measurement vector (*z*) composed of at least 2*N*-1 measurements located at specific points in the system, and the measurement Jacobian matrix $(J(x))$. These measurement locations for the CSE are determined using observability analysis, a process described in Section [2.1.](#page-26-0) The measurement Jacobian matrix is composed of the first derivative of the vector of the nonlinear relationship between measurements and state variables (*h*(*x*)).

In addition to using a lower number of variables, there are other significant differences between the approach proposed here and the conventional weighted least square (WLS) approach. First, WLS is based on the minimization of the measured variables (z) to arrive at an accurate approximation of the states (x) , hence, the reason for requiring many observations (an over-determined system) to achieve a reasonable accuracy. While in the ANN, the minimization is over the states (x) for developing the mapping functions between the required minimum number of variables and states, hence, contributing to an under-determined case.

Secondly, WLS is an iterative process for each time that the states have to be estimated. While ANN is iterative only in the training phase, during the testing it is a feed-forward straight operation, hence, providing a much faster response time.

Finally, WLS is static, while given sufficient training data ANN can provide a time trajectory of the states, and hence provide a dynamic response.

2.4. Load Variation

The amount of energy consumed (load) varies at different periods of the day. This follows a certain daily cycle and varies from customer to customer and from area to area. Even though the amount of energy consumed by individual customers varies, the trend of the total energy consumed by customers at specific substations is fairly predictable. Therefore, the trend of the load cycle for geographical areas, large or small, can be observed and assessed. It has been observed that the load cycle trends in the different areas that make up large geographical regions like New England tend to be similar. This is shown in [Figure 2.3,](#page-38-0) which is a snapshot of the Independent System Operator of New England (ISO-NE) report submitted to Federal Energy Regulatory Commission (FERC). This report was obtained from reference [27]. In this dissertation the load cycle trends in all the areas in each of the power systems used are assumed to be similar.

Figure 2.3. Typical daily load cycle in ISO-NE territory

CHAPTER 3

STATE ESTIMATION USING UNDERDETERMINED SYSTEM OF EQUATIONS

The changes in power generations are driven by the changes in load power demand. In other words, power systems are structured in such a way as to track changes in load in order to provide sufficient generation to balance this demand, such that the system is stable. Thus knowledge of the load behavior (parameters) at any given point in time should be sufficient to obtain an accurate estimate of the system state at that given point, assuming no generation outages and topology changes.

The real power injection $P_{Gi}(t)$ of the *i*th bus at time *t* as a function of changes in load real powers can be expressed as:

$$
P_{Gi}(t) = P_{Gi}(t-1) + \lambda_i \sum \Delta P_{Lj}(t)
$$
\n(18)

$$
\Delta P_L(t) = P_L(t) - P_L(t-1)
$$
\n(19)

where P_{Gi} and P_{Li} present the real power at the generator and load buses respectively and λ_i is the fraction of the load change picked up by the *i*th generator ($0 \leq \lambda_i \leq 1$).

Consider the simple system shown in [Figure 3.1.](#page-40-0) In the following analysis, *P* and *Q* represent real and reactive power, *V* and θ represent voltage and phase angle, *G* and *B* represent admittance parameters of a transmission line, and subscripts G_i and L_i refer to the *j*th generator bus and *i*th load bus. From equations [\(18\)](#page-39-1) and [\(19\)](#page-39-2), it is obvious that using the real and reactive power measurements of the load bus $(P_3 \text{ and } Q_3)$, respectively)

it is possible to find an estimate of the system state. P_3 and Q_3 can be expressed as follows:

$$
P_3 = V_3 V_1 (G_{31} \cos(\theta_3 - \theta_1) + B_{31} \sin(\theta_3 - \theta_1)) + V_3 V_2 (G_{32} \cos(\theta_3 - \theta_2) + B_{32} \sin(\theta_3 - \theta_2))
$$
\n(20)

$$
Q_3 = V_3 V_1 (G_{31} sin(\theta_3 - \theta_1) - B_{31} cos(\theta_3 - \theta_1)) + V_3 V_2 (G_{32} sin(\theta_3 - \theta_2) - B_{32} cos(\theta_3 - \theta_2))
$$
\n(21)

Figure 3.1. A simple 3-bus system

Considering bus 1 to be the slack bus ($\theta_1 = 0$), equations [\(20\)](#page-40-1) and [\(21\)](#page-40-2) can be written in the following shortened form:

$$
\begin{pmatrix} P_3 \\ Q_3 \end{pmatrix} = f(x) \quad \text{where} \quad x = [V_1, V_2, V_3, \theta_2, \theta_3]^T \tag{22}.
$$

Generally, for an *N*-bus system, assuming bus 1 to be the slack bus, the load real and reactive power measurements in equations [\(20\)](#page-40-1) and [\(21\)](#page-40-2) can be expressed as:

$$
P_{Li} = V_{Li} \sum_{j \neq i}^{N} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij})
$$
\n(23)

$$
Q_{Li} = V_{Li} \sum_{j \neq i}^{N} V_j (G_{ij} sin \theta_{ij} - B_{ij} cos \theta_{ij})
$$
 (24)

and in a shortened form,

$$
\begin{bmatrix} P_L \\ Q_L \end{bmatrix} = f(x)
$$

\n
$$
P_L = [P_{L1}, P_{L2}, ..., P_{LM}]^T
$$

\n
$$
Q_L = [Q_{L1}, Q_{L2}, ..., Q_{LM}]^T
$$

\n
$$
x = [V_1, V_2, ..., V_N, \theta_2, \theta_3, ..., \theta_N]^T
$$
\n(25)

where $M =$ number of load buses, $N =$ total number of buses, and $f(x)$ is a vector of 2M nonlinear functions, comprising an under-determined system of equations (*M<N*), mapping load power measurements to state variables.

The function of a state estimator is to find the estimate of the state (*x*) of the system given certain measurements. This process is described in Section [2.1.](#page-26-0) From equation [\(25\)](#page-41-0), assuming f^{-1} exists, we have:

$$
x = [V_1, V_2, \dots, V_N, \theta_2, \theta_3, \dots, \theta_N]^T = f^{-1}(P_L, Q_L)
$$
\n(26)

where f^{-1} is a nonlinear function mapping the load power measurements to the state variables.

Applying the conventional state estimation model outlined in equation [\(1\)](#page-28-0) to the above example, the measurement vector z , and the vector of the non-linear relationship between measurements and state variables *h(x)* become

$$
z = \begin{bmatrix} P_L \\ Q_L \end{bmatrix}
$$

$$
h(x) = f(x)
$$
 (27)

and the error vector *e* is the difference between the actual values and the measurements of P_L and Q_L .

$$
z = h(x) + e \leftrightarrow \begin{bmatrix} P_L \\ Q_L \end{bmatrix} = f(V, \theta) + e \tag{28}
$$

The measurement Jacobian $J(x)$ matrix in equation [\(8\)](#page-30-0) becomes

$$
J(x) = \frac{\partial f(x)}{\partial x}
$$

$$
J(x) = \begin{bmatrix} \frac{\partial P_L}{\partial V} & \frac{\partial P_L}{\partial \theta} \\ \frac{\partial Q_L}{\partial V} & \frac{\partial Q_L}{\partial \theta} \end{bmatrix}
$$
 (29),

Note that when the measurement set is composed of only load bus measurements the Jacobian is underdetermined, for instance, for the system in [Figure 3.1](#page-40-0) it is composed of two sets of equations with five unknown variables. In state estimation problems, the only known quantities are the measurements and the admittance. Admittance comprises of the conductance (G) and susceptance (B) . The process of state estimation in alternating current (AC) circuits is iterative because of the non-linear nature of the equations, so initial guesses need to be made for the state vector *x*. The typical initial guess is a flat start [2], which means the voltage magnitudes *V* are given a value of 1 and the phase angles θ a value of 0.

In all the case studies undertaken in this research, it was impossible to determine the states of the systems using conventional state estimation. In all cases the iterative state estimation process was unable to converge. This is due to the fact that the measurement Jacobian was underdetermined in all the cases. Conventional state estimators require at least a fully determined measurement Jacobian, which means that a minimum of 2*N*-1 measurements must be available. But typically, an overdetermined system of equations is used.

As a result, no existing approach for power systems state estimation is able to accurately model the f^{-1} function. However, in this research, the ability of neural networks to accurately map non-linear patterns is exploited to model it. A brief introduction of artificial neural networks (ANNs) is presented in Section [2.2.](#page-32-0) The ability of ANNs to accurately model the f^{-1} function is demonstrated in the following sections.

3.1. ANN-Based State Estimation

This section shows the application of ANN as a state estimator. The GE 6-bus power system and IEEE 14-bus have been successfully trained using the BP neural network, for state estimation analysis. The bus injection powers and voltages (magnitudes and phase angles) observations were generated from the base cases using the GE Power System Load Flow (PSLF) software and divided into testing and training patterns.

3.1.1. ANN-Based State Estimation for 6-bus System

The diagram of the 6-bus power system is shown in [Figure 3.2.](#page-44-0) For ANN-based state estimation application on the 6 bus power system, there are 4 inputs as the measurement set representing the real and reactive power of the two load buses and 12 outputs as the variable set representing the magnitude and phase angle of voltages of all the buses. Forty (40) observations (patterns) were generated with PLSF software. Each observation comprises 4 data points for the inputs and 12 data points for the outputs, a total of 16 data points. The variation ranges of the 4 inputs are around $\pm 40\%$ of the base case (see Section [1.5](#page-23-0) for definition of base case) values of the loads. This is within the normal (stable) range of the power system. It is assumed that the load cycle at all the load

buses is similar. Out of the 40 patterns, 20 patterns are used for training and the other 20 are used for testing. A neural network with three processing layers is used: two hidden layers with 3 neurons and 6 neurons, respectively and one output layer with 12 neurons. The BP network is designed using MATLAB Neural Network Toolbox.

Figure 3.2. Diagram of the GE 6-bus power system

The performance of the proposed ANN state estimator on the 6-bus system is measured by using the Mean Square Error (MSE) of the training and testing results, which is 1.46 \times 10⁻⁵ for the training data and is 6.81 \times 10⁻⁵ for the testing data. The plots of the ANN calculated voltage magnitudes (shown in dots) with respect to the actual voltage magnitudes (shown in connected lines) for each bus of the 6 bus system are shown in [Figure 3.3](#page-46-0) and the plots of the ANN calculated voltage phase angles (shown in dots) with respect to the actual voltage phase angles (shown in connected lines) for each bus of 6 bus system is shown in [Figure 3.4.](#page-47-0) Note that the actual values were obtained from GE PSLF software.

In the voltage magnitude plots, it might seem that the state estimator has a low accuracy, because the calculated values seem far from the actual values. This is due to the high resolution of the plot values. This resolution is necessary because the voltage magnitude values are in per unit (pu) and so do not have a wide range. Nevertheless, the error in predicting the values is very low. For instance, the farthest value from the diagonal line for the Vmag Bus 6 plot has the ANN (calculated) value of 1.025 pu, while the actual (target) value is 1.0325 pu. This translates to an error of 0.0075 pu, which is about 0.7% error. Note that the R-squared calculation takes into consideration all the plotted data points for the variable.

Figure 3.3. Plots of the actual voltage magnitudes (vertical axes) vs. the calculated voltage magnitudes for the 6-bus system (horizontal axes)

Figure 3.4. Plots of the actual voltage phase angles (vertical axes) vs. the calculated voltage phase angle for the 6-bus system (horizontal axes)

The performance of the ANN-based state estimator on the GE 6-bus system was also measured using the coefficient of determination (R-squared values) of the voltage magnitudes and phase angles from the testing results. The R-squared values ranged from between 98.19% to 100%. 4 buses (1, 2, 3 and 4) showed the highest R-squared value (100%) whereas bus 6 showed the lowest R-squared value (98.19%) for the voltage magnitudes, and bus 3 showed the highest R-squared value (100%) whereas bus 4 showed the lowest R-squared value (99.9997%) for phase angles. The R-squared values are shown in [Table 3.1.](#page-48-0) The voltage magnitudes of buses 1, 2 and 4 were fairly constant and so ANN could estimate their values with very high accuracy. The constancy of these values is due to the generators on these buses. Generators maintain the voltage magnitudes of their buses at a certain value as long as they have sufficient reactive power. Bus 3 is the slack bus, hence its voltage magnitude and phase angle are constant at 1.04 per unit (pu) and 0 degrees, respectively. This is why ANN could estimate their values with very high accuracy.

Bus number	Voltage magnitude	Phase angle R-	
	R-squared value	squared value	
		0.999999972	
		0.999999983	
		0.999997293	
	0.999867919	0.999999906	
	0.981884377	0.999997454	

Table 3.1. R-Squared Values for GE 6-Bus System

3.1.2. ANN-Based State Estimation for IEEE 14-Bus System

The diagram of the IEEE 14-bus power system is shown in [Figure 3.5.](#page-49-0) There are 16 inputs as the measurement set representing the real and reactive powers of the 8 load buses (4, 5, 9, 10, 11, 12, 13 and 14) and 28 outputs representing the voltage magnitudes and phase angles of all the buses in the system. 112 observations (patterns) were generated using GE's PSLF software and divided into 56 patterns for training and the rest for testing. Each observation comprises 16 data points for the inputs and 28 data points for the outputs, a total of 44 data points. The variation ranges of the 16 inputs are $\pm 60\%$ of around the base case (see definition in Section [1.5\)](#page-23-0) values of the load buses. This is within the normal (stable) range of the power system. A neural network with three layers is used for training: two hidden layers with 10 neurons and 8 neurons, respectively, and one output layer with 28 neurons.

Figure 3.5. Diagram of the IEEE 14-bus system

The performance of the ANN-based state estimator on the IEEE 14-bus system is measured using the Mean Square Error (MSE) of the training and testing results, which is 2.49×10^{-5} for the training data and 2.33×10^{-3} for the testing data. The plots of the ANN calculated voltage magnitudes (shown in dots) with respect to the actual voltage magnitudes (shown in connected lines) for buses 2, 3, 6, 9, 11 and 14 are shown in [Figure](#page-52-0) [3.6](#page-52-0) and the plots of the ANN calculated voltage phase angles (shown in dots) with respect to the actual voltage phase angles (shown in connected lines) for the above-listed buses is shown in [Figure 3.7.](#page-53-0) Note that the actual values were obtained from GE PSLF software.

The performance of the ANN-based state estimator on the IEEE 14-bus system was also measured using the coefficient of determination (R-squared values) of the voltage magnitudes and phase angles from the testing results. The R-squared values ranged from between 94.54% to 100%. 3 buses (1, 2 and 3) showed the highest Rsquared value (100%) whereas bus 8 showed the lowest R-squared value (99.74%) for the voltage magnitudes, and bus 1 showed the highest R-squared value (100%) whereas bus 2 showed the lowest R-squared value (94.54%) for phase angles. The R-squared values are presented in [Table 3.2.](#page-51-0) The voltage magnitudes of buses 1, 2 and 3 were fairly constant and so ANN could estimate their values with very high accuracy. The constancy of these values is due to their remoteness with respect to the load center – the farther a bus is from the load center, the lesser the variation in its voltage magnitude.

Bus number	Voltage magnitude	Phase angle R-	
	R-squared value	squared value	
$\overline{2}$		0.945405	
3		0.982143	
4	0.999673	0.988412	
5	0.999986	0.984879	
6	0.998921	0.994684	
	0.998891	0.993981	
8	0.997405	0.993981	
9	0.998309	0.995899	
10	0.998216	0.996002	
11	0.998385	0.995947	
12	0.998257	0.996084	
13	0.999246	0.99539	
14	0.998176	0.996939	

Table 3.2. R-Squared Values for IEEE 14-Bus System Using Only Load Measurements

In the voltage magnitude plots, it might seem that the state estimator has a low accuracy, because the calculated values seem far from the actual values. This is due to the high resolution of the plot values. This resolution is necessary because the voltage magnitude values are in per unit (pu) and so do not have a wide range. Nevertheless, the error in predicting the values is very low. For instance, the farthest value from the diagonal line for the Vmag Bus 14 plot has the ANN (calculated) value of 1.009 pu, while the actual (target) value is 0.991 pu. This translates to an error of 0.018 pu, which is about 1.8% error. Note that the R-squared calculation takes into consideration all the plotted data points for the variable.

Figure 3.6. Plots of the actual voltage magnitudes (vertical axes) vs. the calculated voltage magnitudes for the IEEE 14-bus system (horizontal axes)

Figure 3.7. Plots of the actual voltage phase angles (vertical axes) vs. the calculated voltage phase angle for the IEEE 14-bus system (horizontal axes)

The preliminary results shown above serve as a proof of concept that it is possible to implement a state estimator having a Jacobian with an underdetermined system of equations.

CHAPTER 4

DETERMINATION OF CRITICAL LOCATIONS

Critical variables of a system are those variables that invariably capture the changes occurring in that system; in other words, they significantly reflect changes of other variables in the system. The methodology proposed in this dissertation is based on the eigenvalues of the measurable variables of the system, including the real and reactive power injections at buses (generators, synchronous condensers and loads), the real and reactive power flows on the lines and the voltage magnitudes and phase angles of all the buses. The eigenvalues of the given system are determined using principal component analysis (PCA). The reduced model state estimation tool is developed using artificial neural networks (ANNs). [Figure 4.1](#page-57-0) gives the flowchart for the identification of critical variables developed in this research. The flowchart is more specific to the threshold method discussed in Section [4.2.1.](#page-64-0)

The methodology comprises the following steps:

1) Data generation

Mathematical models (see definition in Section [1.5\)](#page-23-0) of the system under study are used to generate data samples by simulation as it is difficult to obtain historical data containing every desired measurable variable. Simulations covering the normal operating range of the system are run using a load flow program like Siemens Power System Simulator for Engineering (PSS/E). Typically, in system planning studies, the utility company or regional system operator provides the base cases (see definition in Section [1.5\)](#page-23-0) and advises on the typical generator dispatches and configurations of the system.

2) Selection of principal components

Eigenvalues of the variables of the system are calculated from the generated data and ranked in order of magnitude into principal components. The principal components contributing a certain percentage of the total sum of the eigenvalues are selected for further analysis.

3) Identification of critical variables and their locations

Critical variables of the system are identified using a predetermined threshold. Coefficients (elements) of the eigenvector matrix that correspond to the selected principal components are compared against the threshold. Variables with coefficients greater than the threshold are identified as critical. Buses corresponding to the identified variables are classified as the critical locations of the system.

4) State Estimation

Measurements of the identified critical variables of the system are used to train an ANN-based state estimator, as explained in [CHAPTER 3.](#page-39-0) This state estimator utilizes a fewer number of measurements than conventional state estimators.

These processes are described in more detail below.

Figure 4.1. Flowchart of the methodology

4.1. Principal Component Analysis

This dissertation proposes a method for identification of critical variables of a power system and their locations using eigenvalue analysis. The most readily available and easily understandable technique for eigenvalue analysis is the principal component analysis (PCA).

PCA is a fundamental aspect in the study of multivariate data and a standard tool in modern data analysis. It is a non-parametric statistical method for converting observations of possibly correlated variables into a set of uncorrelated variables through orthogonal transformation. The goal of PCA is to find the most meaningful basis to present a data set with the hope that the new basis will eliminate the noise and uncover hidden structure.

PCA operates using three major assumptions which include: linearity, large variances representing interesting and important structure and the orthogonality of principal components. Linearity vastly simplifies the problem of re-expressing the data set by restricting the number of potential bases; therefore PCA is limited to expressing the data as a linear combination of its basis vectors. The directions with the largest variances in the measurement space contain the dynamics of interest and are presumed to be directions with the highest signal-to-noise ratio (SNR), while those with lower variances represent noise. The assumption that the principal components are orthogonal provides an intuitive simplification that makes it possible for PCA to exploit linear algebra decomposition techniques [28]. Such techniques include eigenvector decomposition and singular value decomposition.

42

PCA is gradually gaining popularity in the analysis of power systems. Some works use an extension of PCA to solve peculiar problems, for instance, reference [29] combines radial basis function with PCA to handle non-Gaussian distributed variables, whereas in reference [30] PCA is used to eliminate colored measurement noise in order to improve the accuracy of the Kalman state estimator. Other works employ PCA directly. In references [31] and [32] PCA is used to reduce the dimensions of measurement data in order to speed up the computation process. PCA has likewise been used to detect and visualize power system disturbances [33], to identify coherent generators in large power systems [34], and to detect islands for distributed generation systems [35].

In this dissertation, PCA is used to generate the eigenvalues for data comprising observations of measurable system variables, such as real and reactive power flows and injections, and voltage magnitudes and phase angles. Further analysis is done on a subset of the data corresponding to the most prominent eigenvalues. This data subset is used to identify the system critical variables and their locations. The effectiveness of monitoring these critical locations is demonstrated on the IEEE 118-bus system.

The algorithm for PCA is simple and can be summarized in three steps for multidimensional data:

a) Choose a normalized direction in *m*-dimensional space along which the variance in the data *Y* is maximized. Save this as vector p_l . (*m* is the number of basis vectors.)

- b) Identify another direction along which the variance is maximized. Because of the orthonormality conditions, restrict the search to all directions orthogonal to all previous selected directions. Save this as vector *pⁱ* .
- c) Repeat steps a) and b) until *m* vectors are selected.

The resulting ordered set of the vectors in $P = \{p_1, ..., p_i, ..., p_m\}$ are the principal components of *Y*.

Assuming *n* observations are generated for a system with *m* measurable variables, the $n \times m$ data matrix *Y* has a symmetric $m \times m$ correlation matrix *C*.

$$
\mathbf{Y} = \begin{bmatrix} y_{11} & \dots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \dots & y_{nm} \end{bmatrix}
$$

$$
\mathbf{C} = \begin{bmatrix} c_{11} & \dots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{m1} & \dots & c_{mm} \end{bmatrix}
$$
 (30)

Premultiplying and postmultiplying *C* by a certain orthonormal matrix *O* converts it to a diagonal matrix $\boldsymbol{\Lambda}$ [36] such that

$$
\mathbf{O}^T \mathbf{C} \mathbf{O} = \mathbf{\Lambda} \tag{31}
$$

$$
\mathbf{O} = \begin{bmatrix} o_{11} & \dots & o_{1m} \\ \vdots & \ddots & \vdots \\ o_{m1} & \dots & o_{mm} \end{bmatrix}
$$

$$
\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \lambda_i & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda_m \end{bmatrix}
$$
 (32)

The eigenvalues (characteristic roots) λ of C form the diagonal elements of Λ and are calculated by finding the determinant of the characteristic equation for *C*:

$$
|\mathbf{C} - \lambda \mathbf{I}| = 0 \tag{33}
$$

where λ is an *m*th degree polynomial and *I* is the identity matrix of size *m*. Practically, the eigenvalues are obtained using iterative procedures, and they can be scaled such that they sum up to 1.

The eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$ are sample variances of the principal components (PCs), which are obtained through a principal transformation of the original *m* coordinate axes such that the new variables are uncorrelated and each new axis is selected to represent as much of the variance in *Y* as possible [37]. Therefore, the first PC accounts for more variability than the second, the second PC accounts for more variability than the third, and so on. Essentially, the $m \times n$ PCs matrix **P** is found by:

$$
P = F^T [Y - \overline{Y}]^T
$$
 (34)

where *F* is the matrix whose columns (f_1, f_2, \ldots, f_m) are the eigenvectors of *C* and \overline{Y} is the mean of *Y* [36].

$$
\boldsymbol{P} = \begin{bmatrix} P C_1 \\ \vdots \\ P C_m \end{bmatrix} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \dots & p_{mn} \end{bmatrix}
$$
(35)

$$
\mathbf{F} = [f_1 \cdots f_i \cdots f_m] = \begin{bmatrix} a_{11} & \cdots & \cdots & \cdots & a_{1m} \\ \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \cdots & a_{ki} & \cdots & \vdots \\ a_{m1} & \cdots & \cdots & \cdots & a_{mm} \end{bmatrix}
$$
(36)

The trace of $F(r(F))$ is equal to the trace of $C(r(C))$. The rows of P correspond to the PCs and are arranged in order of decreasing principality $(PC_1$ is more important than $PC₂$). The eigenvectors are determined by solving:

$$
\begin{cases}\n[\mathbf{C} - \lambda \mathbf{I}]g_i = 0 \\
f_i = \frac{g_i}{\sqrt{g_i^T g_i}}\n\end{cases}
$$
\n(37)

where *g* is an intermediate variable used for solving the set of equations in [\(37\)](#page-62-0).

The proportion of the variance explained by each PC is given by:

$$
S_i = \frac{\lambda_i}{tr(C)} = \frac{\lambda_i}{\sum_i^m \lambda_i} \tag{38}
$$

where $tr(C)$ is the trace of C . The selection of PCs retained for further computations was based on the values of the vector *S*. The *i*th PC is retained if

$$
S_i \ge \frac{1}{m} \tag{39}
$$

where 1 is the sum of the scaled values of the eigenvalues. This quick technique is derived from the Average Root technique described along with other PC significance tests in [36] and [37].

The *P* and *F* matrices can then be expressed thus:

$$
P = \begin{bmatrix} P_{ret} \\ - - \\ P_{dis} \end{bmatrix} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{r1} & \dots & p_{rn} \\ p_{r+1,1} & \dots & p_{r+1n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \dots & p_{mn} \end{bmatrix} \text{ discarded}
$$
 (40)

$$
\mathbf{F} = [\mathbf{F}_{ret} \mid \mathbf{F}_{dis}] = \begin{bmatrix} a_{11} & \cdots & a_{1r} & | & a_{1,r+1} & \vdots & a_{1m} \\ \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mr} & | & a_{m,r+1} & \cdots & a_{mm} \end{bmatrix}
$$
(41)

where P_{ret} and F_{ret} represent the matrices of the retained PCs and eigenvectors whereas *Pdis* and *Fdis* represent the matrices of the discarded PCs and eigenvectors.

4.2. Identification of Critical Locations

From equation [\(34\)](#page-61-0) the matrix *Y* can be reconstructed thus:

$$
Y = P^T F^T + \overline{Y}
$$
 (42)

and

$$
\widehat{\mathbf{Y}} = \boldsymbol{P}_{ret}^T \boldsymbol{F}_{ret}^T + \overline{\mathbf{Y}} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{r1} & \dots & p_{rn} \end{bmatrix}^T \begin{bmatrix} a_{11} & \dots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mr} \end{bmatrix}^T + \overline{\mathbf{Y}} \tag{43}
$$

where \hat{Y} is $n \times m$ matrix of the reconstructed data. However, from observation, some variables in the power system are more critical than others. The critical variables have more influence in the reconstruction of \hat{Y} than the others. The variable V_k is considered to be critical if it is significantly representative of the changes of other variables in the system:

$$
\forall V_i: \partial V_j \to \Delta V_k; i, j = 1, m; V_k \in critical variables
$$
\n
$$
(44)
$$

where ∂ represents any change in the network variables, Δ represents a substantial expression of the critical variable V_k . Critical variables are identified using the coefficients of the eigenvectors corresponding to the retained PCs. Two approaches (threshold and R-squared methods) are investigated in this dissertation for the identification of critical variables. The critical locations of the power system are determined by the locations of the critical variables. [Figure 4.1](#page-57-0) illustrates the process of the identification of critical variables and locations using the threshold method. In the diagram *x* represents 1/*m*.

4.2.1. Identification of Critical Variables using the Threshold Method

The criticality of each of the *m* variables of the system can be determined through the coefficients (elements) of the retained eigenvector matrix F_{ret} . The variable corresponding to the highest absolute coefficient value in the retained eigenvector matrix *Fret* is the most critical and the minimum absolute coefficient value relates to the least critical variable in the system. Although this variable selection technique is novel, it is similar in philosophy to the principal component methods presented in references [38] and [39].

For practical purposes, a threshold is used to set a cut-off value for selecting critical variables. Therefore, V_k is considered critical if

$$
|a_{ki}| \ge \varepsilon; \quad i = 1, r; k = 1, m \tag{45}
$$

where a_{ki} is an element in \vec{F} as shown in equation [\(36\)](#page-61-1) and the *k*th coefficient of the *i*th column, ε is the specified threshold and *r* is the number of retained eigenvectors. Note that the number of retained eigenvectors equals the number of retained PCs. A suitable threshold ε for determining the criticality of a variable is given by:

$$
\varepsilon = \frac{2}{3} e \sum_{k=1}^{m} s_k \tag{46}
$$

$$
s = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (a_i - \bar{a})^2}
$$
 (47)

where *e* is Euler's number, a_i is the coefficient of the *i*th variable, \bar{a} is the mean of the coefficients for each variable. If multiple coefficients of the same variable are greater than the threshold, only one instance of the variable is recorded as critical.

The formulas in equations [\(46\)](#page-64-1) and [\(47\)](#page-64-2) are derived from observation of the eigenvector (F) matrices of the IEEE 14-bus system analyzed in [CHAPTER 5.](#page-66-0) The formulas are developed by judiciously analyzing the statistical composition (means, variances, standard deviations, etc.) of the matrices and correlating them to the threshold determined hitherto by trial and error. As observed in that chapter, 21 critical variables were observed for each of the three dispatches analyzed and these 21 variables were exactly the same for all three dispatches. These derived formulas will be used in subsequent determination of critical variables of power systems.

4.2.2. Identification of Critical Variables using the R-squared Method

This entails simulating the loss of the measurement for each variable individually and determining the effect of this loss on the system. To simulate the loss of the measurement for a variable, all the elements of the retained eigenvector matrix (*Fret*) related to that variable are set to zero. Then, a reconstruction of the data matrix (\hat{Y}) is done using equation [\(43\)](#page-63-0). The measure of the impact of the loss of each variable on the system is the coefficient of determination (R-squared):

$$
R^{2} = 1 - \frac{\Sigma(y_{nm} - \hat{y}_{nm})^{2}}{\Sigma(y_{nm} - \bar{y}_{nm})^{2}}
$$
(48)

$$
y_{nm} \in \mathbf{Y}; \ \hat{y}_{nm} \in \widehat{\mathbf{Y}}; \ \bar{y}_{nm} \in \overline{\mathbf{Y}} \tag{49}
$$

The variables are then ranked in order of increasing R-squared values. The most critical variable is the variable with the minimum R-squared value because a lower value means a loss of the variable's measurement would have a significant impact on the system.

CHAPTER 5

IMPLEMENTATION ON A SMALL SYSTEM

In this chapter the concept of the identification of critical variables of the IEEE 14-bus system and their locations and the use of the identified critical variables in estimating the state of the system is presented. The IEEE 14-bus system was a segment of the American Electric Power System (AEP) around February 1962 [40]. AEP served the Midwestern US. The original test case as obtained from reference [40] does not have line limits, and has low base voltages and too much voltage control capability compared to the power systems of the 1990's. The diagram of the IEEE 14-bus power system is shown in [Figure 3.5.](#page-49-0)

5.1. Principal Component Analysis on IEEE 14-Bus System

Firstly, bus injection powers, line flows and complex voltage observations were generated from three different generation dispatches (Table 5.1) of the system using the GE Power System Load Flow (PSLF) software. This was done in the range within which the system is in the normal (stable) state. The observations were created by varying loads at all load buses at $\pm 60\%$ of the base case value.

Dispatch #	Gen 1	Gen 2
	Slack bus	40 MW
	Slack bus	60 MW
	Slack bus	50 MW

Table 5.1. Three Dispatches Used

The measurable variables in the IEEE 14-bus system include 40 real and reactive power flow measurements, 32 real and reactive power injection (for loads, generators and synchronous condensers) measurements, and 28 voltage magnitude and phase angle measurements – a total of 100 measurements. After generating the observations as described above, PCA was run on each of the dispatches. From a total of 100 PCs the first six PCs were selected in each case. The total contribution of the six PCs in each case was about 96% of the system power. The values of the first 10 PCs for each of the three dispatches are presented in [Table 5.2.](#page-67-0) Using a threshold of $\varepsilon = 0.18$, 21 critical variables were identified for each of the three dispatches. The same exact 21 variables were identified for all the dispatches. The results are presented in Table 5.3. For power injections, p and q stand for real and reactive power injection, and the number behind them is the bus number. For power flows, the first number stands for the "from bus" bus number, whereas the second number stands for the "to bus" bus number.

	Dispatch 1		Dispatch 2		Dispatch 3	
PC	Contribution,	Cumulative	Contribution.	Cumulative	Contribution.	Cumulative
#	$\%$	Total, %	$\%$	Total, %	$\%$	Total, %
	52.6748		63.6569		59.7645	
2	20.9395	73.6143	13.8658	77.5227	16.9066	76.6711
3	10.9705	84.5848	8.4681	85.9908	8.1347	84.8058
$\overline{4}$	5.2002	89.7850	4.6442	90.6350	5.0801	89.8859
5	3.8218	93.6068	3.6778	94.3128	3.8967	93.7826
6	2.0643	95.6711	1.8318	96.1447	2.0043	95.7869
7	1.1539	N/A	1.0290	N/A	1.1242	N/A
8	0.8206	N/A	0.7313	N/A	0.7991	N/A
9	0.6450	N/A	0.5733	N/A	0.6263	N/A
10	0.4827	N/A	0.4284	N/A	0.4688	N/A

Table 5.2. First 10 Principal Components for the Three Dispatches

Power Injections
p1, p2, q2, p4, p9, q9, p13, p14
Power Flows
p1-2, p1-5, p2-4, p4-5, p4-7, q4-7, p5-6, q5-6, p6-13, p7-9, q7-9, p9-14, p13-14

Table 5.3. Critical Variables in the IEEE 14-bus System

5.1.1. Critical Locations of the IEEE 14-Bus System

Nine critical locations (buses) of the IEEE 14-bus system corresponding to the locations of the critical variables were identified. These locations are represented by the bus numbers: 1, 2, 4, 5, 6, 7, 9, 13 and 14. For example, critical variables $p1$, $p2$ and $p1-2$ identify buses 1 and 2 as critical locations. Other critical locations were identified in a similar manner. From the foregoing, the number of monitored nodes for the IEEE 14-bus system is about 64% of the total available nodes in the system.

5.1.2. ANN-Based State Estimation with Critical Variables

The IEEE 14-bus was successfully trained, using the BP neural network, for state estimation analysis. The inputs to the ANN are the 21 critical variables identified above. The outputs are the 28 voltage magnitudes and phase angles of all buses in the system. A total of 112 observations (patterns) were generated from Dispatch #3. Each observation comprises 21 data points for the inputs and 28 data points for the outputs, a total of 49 data points. From these, half of the patterns were used in training the ANN and the rest for testing. The neural network has three processing layers: two hidden layers with 10 neurons and 8 neurons, respectively and one output layer with 28 neurons. The BP

network is designed using MATLAB Neural Network Toolbox. The performance of the proposed ANN state estimator could be measured using the Mean Square Error (MSE). The MSE for the training data is 2.81×10^{-5} . The MSE for the testing data for this scenario is 1.38×10^{-4} . The plots of the ANN calculated voltage magnitudes (shown in dots) with respect to the actual voltage magnitudes (shown in connected lines) for 6 buses selected at random (buses 2, 3, 6, 9, 11 and 14) are shown in [Figure 5.1](#page-72-0) and the plots of the ANN calculated voltage phase angles with respect to the actual voltage phase angles for the same buses are shown in [Figure 5.2.](#page-73-0) Note that the actual values were obtained from GE PSLF software.

The performance of the ANN-based state estimator on the IEEE 14-bus system was also measured using the coefficient of determination (R-squared values) of the voltage magnitudes and phase angles from the testing results. The R-squared values ranged from between 99.66% to 100%. 3 buses (1, 2 and 3) showed the highest Rsquared value (100%) whereas bus 5 showed the lowest R-squared value (99.66) for the voltage magnitudes, and bus 1 showed the highest R-squared value (100%) whereas bus 12 showed the lowest R-squared value (99.98%) for phase angles. The R-squared values are presented in [Table 5.4.](#page-70-0) The voltage magnitudes of buses 1, 2 and 3 were fairly constant and so ANN could estimate their values with very high accuracy. The constancy of these values is due to their remoteness with respect to the load center – the farther a bus is from the load center, the lesser the variation in its voltage magnitude.

Table 5.4. R-Squared Values for IEEE 14-Bus System Using Critical Variables'

Measurements

In the voltage magnitude plots, it might seem that the state estimator has a low accuracy, because the calculated values seem far from the actual values. This is due to the high resolution of the plot values. This resolution is necessary because the voltage magnitude values are in per unit (pu) and so do not have a wide range. Nevertheless, the error in predicting the values is very low. For instance, the farthest value from the diagonal line for the Vmag Bus 14 plot has the ANN (calculated) value of 1.009 pu, while the actual (target) value is 0.991 pu. This translates to an error of 0.018 pu, which is about 1.8% error. Note that the R-squared calculation takes into consideration all the plotted data points for the variable.

The critical variables identified for this IEEE 14-bus system make up about 96% of the power of the system as shown in [Table 5.2.](#page-67-0) However, the minimum R-squared value from the ANN-based state estimation is 99.66%. This level of accuracy is

achievable because only the values of the voltage magnitudes and phase angles, which make up only 28% of the variables, were estimated.

The proposed ANN state estimator program was coded in MATLAB and run on Intel i7 64-bit Dell Precision T1500 machine running a Windows 7 operating system.

Figure 5.1. Plots of the actual voltage magnitudes (vertical axes) vs. the calculated voltage magnitudes for IEEE 14-bus system (horizontal axes)

Figure 5.2. Plots of the actual voltage phase angles (vertical axes) vs. the calculated voltage phase angles for the 14-bus system (horizontal axes)

Conventional state estimators require a minimum of 2*N*-1 input variables (measurements) in order to function. Here *N* refers to the number of buses in the system. The minimum number of measurements required for conventional state estimation for the IEEE 14-bus system is 27. However, by using only 21 measurements, the results of this methodology have proved that it is possible to estimate the state of a system with a number of input variables fewer than 2*N*-1. The difference in the number of input variables may not be much in this small 14-bus system but it is significant in larger systems as documented in [CHAPTER 6.](#page-75-0)

In this chapter the number of critical variables picked up by principal component analysis (PCA) was determined by trial and error. However, the results of the analysis on the IEEE 14-bus system in this chapter form the basis for the derivation of the threshold equation used in the threshold method for identification of critical variables of power systems presented in [CHAPTER 4.](#page-55-0)

CHAPTER 6

IMPLEMENTATION ON A LARGE SYSTEM

The results presented in [CHAPTER 5](#page-66-0) are the implementation of the methodology on relatively small systems. This chapter will be devoted to implementation of the methodology on IEEE 118-bus system. This is a fairly large standardized test system composed of 186 branches (including transformers), 91 loads and 54 generators. This system represents a portion of the American Electric Power System in the Midwestern US as of December 1962. Even though it has a lot of voltage control devices, and the base KV (kilovolt) levels and line MVA (Megavolt Amperes) limits were made up, the test case is quite robust and converges in about 5 iterations with a fast decoupled power flow [40]. Using a standardized system makes it easier to verify and compare results for different approaches and methodologies in power system analyses. The base case (see definition in Section [1.5\)](#page-23-0) of the system was obtained from reference [40] in the IEEE common data format (CDF).

6.1. Data Generation for IEEE 118-Bus System

The base case (see definition in Section [1.5\)](#page-23-0) of the IEEE 118-bus system was converted from the CDF format into a Power System Simulator for Engineering (PSS/E) case format. Given that the format in which the system data is available does not contain details about the system load level, it was assumed that the system load was at a shoulder level (70% of the peak value). Therefore, the minimum and maximum generation capacities (*Pmin* and *Pmax*) of the generators were adjusted to accommodate system loading at light load and peak load levels using the "Load Modeling Guide for ISO New England

Network Model" [41]. The light load level is 45% of the peak load level. Thus, the light load level used is 60% (\approx 45/70) of the base case loading, whereas the peak load level used is 140% (\approx 100/70). Also, since no utility company or regional system coordinating body can provide information on typical system dispatches for the IEEE 118-bus system, the following dispatches documented in [Table 6.1](#page-76-0) were assumed. Dispatch 1 is the base dispatch with all generators online as shown in [Table 6.2](#page-77-0) (motors and synchronous condensers are not included because they do not generate real power). In the remaining dispatches, one generator is switched offline (Generator Offline column) while another (Pickup Generator column) picks up the real power (MW Redispatched column) previously generated by the offline generator.

Dispatch	Generator Offline	Pickup Generator	MW Redispatched
ıа	Bus 12	Bus 10	85
1b	Bus 61	Bus 65	160
1c	Bus 49	Bus 89	204
1d	Bus 25	Bus 80	220
1e	Bus 103	Bus 100	

Table 6.1. Dispatches Used for the IEEE 118-bus system

Generator Bus	Real Power Output	Generator Bus	Real Power Output
Number	(MW)	Number	(MW)
10	450	65	391
12	85	66	392
25	220	69	513.4
26	314	80	697
31	7	87	
46	19	89	607
49	204	100	252
54	48	103	40
59	155	111	36
61	160		

Table 6.2. Real Power Output of Generators in Dispatch 1

Data samples were generated for each of the dispatches outlined above by varying all the loads simultaneously within the margins described at the beginning of this section $(\pm 40\%$ of the original value in the load flow case) and adjusting the online generators as necessary to accommodate the permutations. This is due to the assumption that the system loading level as obtained from reference [40] is at shoulder level. This permutation range covers the light load and peak load levels, which correspond to the normal operating range of power systems. Load flow simulations were run using Python and PSS/E, and the values of the measurable system variables, such as real and reactive power flows and injections, and voltage magnitudes and phase angles, were recorded. A total of 1506 observations were recorded.

6.2. Principal Component Analysis on IEEE 118-Bus System

The measurable variables considered for the IEEE 118-bus system include 372 real and reactive power flow measurements, 290 real and reactive power injection measurements (for loads, generators and synchronous condensers), and 236 voltage magnitude and phase angle measurements – a total of 898 measurements. Given that the number of observations is 1506, a 1506 \times 898 data matrix *Y* is obtained.

The 898×898 covariance matrix *C* has 898 eigenvalues, which are scaled such that they sum up to 1. The scaled values range from 0 to 0.7999. From equation [\(39\)](#page-62-0) the PCs retained must fulfill the condition: $S_i \geq \frac{1}{m}$ $\frac{1}{\text{m}} = \frac{1}{89}$ $\frac{1}{898}$ = 0.001113 = 0.1113%.

The PCA investigation for this experiment was done using MATLAB [42] and [43], and the singular value decomposition algorithm in MATLAB was employed for the eigenvalue calculation.

Out of the original 898 PCs, the first 11 PCs, with values ranging from 0.1160% to 79.9943% are greater than 0.1113%; thus they satisfied the PC retention condition in equation [\(39\)](#page-62-0) and were retained. The values of the first 15 PCs are presented in [Table 6.3.](#page-79-0) The sum of the contributions of the retained PCs equals 99.76%. The discarded 887 PCs had contributions ranging from 0 to 0.0614%.

PC#	Contribution, %	Cumulative Total, %
	79.9943	
$\overline{2}$	8.1245	88.1188
3	6.1586	94.2774
4	2.6551	96.9325
5	1.1696	98.1021
6	0.5996	98.7017
7	0.3552	99.0569
8	0.2342	99.2911
9	0.2220	99.5131
10	0.1292	99.6423
11	0.1161	99.7584
12	0.0614	N/A
13	0.0504	N/A
14	0.0411	N/A
15	0.0293	N/A

Table 6.3. First 15 Principal Components

6.3. Identification of Critical Locations using the Threshold Method

From equations [\(46\)](#page-64-0) and [\(47\)](#page-64-1) the threshold for determining the criticality of variables ε was calculated and found to be 0.06. Based on this threshold 151 unique variables were classified as critical because the absolute value of their coefficients is greater than the threshold. The coefficients are elements of the matrix of the retained eigenvectors (\mathbf{F}_{ret}). The critical variables include 54 real and reactive power injections and 97 real and reactive power flows. These variables are presented in [Table 6.4.](#page-80-0) For power injections, P_G and Q_G stand for generator real and reactive power, P_L and Q_L stand for load real and reactive powers, and the number behind them is the bus number. For power flows, the first number stands for the "from bus" bus number, whereas the second number stands for the "to bus" bus number. A plot of the coefficients of the identified

critical variables is presented in [Figure 6.1.](#page-81-0) The coefficient of the least critical variable

was 0.06, whereas the most critical variable's coefficient was 0.534.

Figure 6.1. Coefficients of the critical variables

In order to visualize the impact of the critical variables on the system, a simulation of the cumulative loss of all the critical variable measurements, starting from the most critical to the least, was carried out. A cumulative loss of all the critical variable measurements involves several iterations. The number of iterations equals 151, which is the number of critical variables in this experiment with the IEEE 118-bus system. At the first iteration all the elements of the retained eigenvector matrix (*Fret*) related to the most critical variable is set to zero. At the second iteration all the elements of the retained eigenvector matrix (*Fret*) related to the most critical and second most critical variables are set to zero, and so on. At the final iteration all the elements of the retained eigenvector

matrix (F_{ret}) relating to all the critical variables are set to zero. At each iteration a reconstruction of the data matrix (\hat{Y}) is done using equation [\(43\)](#page-63-0) and the R-squared value of matrix \hat{Y} with respect to matrix Y is used as a measure of the impact of the cumulative loss of the measurements involved in that iteration. The graph of the impact of the cumulative loss of all the critical variable measurements is presented in [Figure 6.2.](#page-83-0) From the graph it could be observed that before any critical variable was lost, the R-squared value was 99.76%. The reason why the R-squared value is not 100% is because only the retained PCs are used for the computations. With the loss of the most critical variable, the R-squared value decreased to 95.27%. As more measurements of critical variables are cumulatively lost the R-squared value progressively reduces. When all the critical variables are lost the R-squared value drops to 12.71% and this demonstrates the massive impact that critical variables have on the system. This exercise of examining the cumulative impact of losing all the identified critical variables was done to verify the criticality of the identified set.

Figure 6.2. Impact of a cumulative loss of measurements of the critical variables

Also, the impact of the loss of all non-critical variables' measurements was assessed. The loss of the measurements for all the non-critical variables was simulated by setting to zero all the elements of the retained eigenvector matrix (F_{ret}) related to these variables. Then a reconstruction of the data matrix (\hat{Y}) was done using equation [\(43\)](#page-63-0) and the R-squared value of matrix \hat{Y} with respect to matrix Y is used as a measure of the impact of the loss of the measurements of the non-critical variables. The R-squared value was found to be 87.05% and this demonstrates that the non-critical variables do not have as much impact on the system as the critical variables do.

76 critical locations of the IEEE 118-bus system corresponding to the locations of the critical variables were identified. These locations are represented by their bus numbers and are presented in [Table 6.5.](#page-84-0) Critical locations are candidate locations for PMU placement (for utilities that have not yet installed PMUs in their systems), prioritization of the measurement units in these areas for maintenance and calibration, and procurement of backup units for these locations in case of failure. The effectiveness of using critical locations for steady state monitoring and control of the IEEE 118-bus system is demonstrated in Section [6.6.](#page-92-0)

Table 6.5. Critical Locations Identified using the Threshold Method

1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 17, 18, 19, 23, 24, 25, 26, 27, 30, 32, 34, 35, 36, 37, 38, 40, 42, 46, 47, 49, 54, 56, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 85, 88, 89, 92, 93, 94, 95, 96, 97, 98, 99, 100, 103, 104, 105, 110, 113, 118

6.4. Identification of Critical Locations using the R-Squared Method

This is an alternative method for identifying critical variables and involves two rounds of simulation of the loss of measurements. The first round comprises simulating the impact of the loss of the measurement of each variable on the system, whereas the second round involves the simulation of a cumulative loss of measurements of all the variables.

To simulate the loss of the measurement for a variable, all the elements of the retained eigenvector matrix (F_{ret}) related to that variable is set to zero. Then a reconstruction of the data matrix \hat{Y} is done using equation [\(43\)](#page-63-0) and the R-squared value of matrix \hat{Y} with respect to matrix Y is used as a measure of the impact of the loss of the measurement of the variable.

After the first round of simulations, the variables are ranked in descending order of criticality based on the R-squared value corresponding to the loss of their measurement. The least R-squared value corresponds to the most critical variable, whereas the largest R-squared value corresponds to the least critical variable. A plot of the impact of each variable on the system is presented in [Figure 6.3.](#page-86-0) The R-square value for the loss of the most critical variable was 89.26%, whereas the R-square value for the loss of the least critical variable was 99.76%. This implies that the least critical variable has minimal impact on the system. For the 25 most critical variables a loss of their measurement individually yielded an R-square value less than 99% (ranged between 89.26% and 98.92%).

Figure 6.3. Impact of the loss of measurements of individual variables on the system

In the second round, the simulation of a cumulative loss of measurements of all the variables in the system was done. A cumulative loss of measurements of all the variables on the system involves several iterations. The number of iterations equals 898, which is the number of variables in this experiment with the IEEE 118-bus system. At the first iteration all the elements of the retained eigenvector matrix (F_{ret}) related to the first variable is set to zero. At the second iteration all the elements of the retained eigenvector matrix (*Fret*) related to the first and second variables are set to zero, and so on. At the final iteration all the elements of the retained eigenvector matrix (*Fret*) are set to zero, that is, the *Fret* matrix the becomes a zero matrix. At each iteration a reconstruction of the data matrix (\hat{Y}) was done using equation [\(43\)](#page-63-0) and the R-squared value of matrix \hat{Y} with respect to matrix *Y* is used as a measure of the impact of the cumulative loss of the measurements involved in that iteration. Starting from the least significant variables, the cumulative loss of the measurements of all the variables was simulated and the R-squared values were calculated. [Figure 6.4](#page-88-0) presents a graph of the impact of a cumulative loss of measurements of all the variables on the system.

This method allows a system planner to visually locate a cutoff position for the classification of critical variables. Also, if a confidence value, say 95%, is preferred for the selection of critical variables this method allows for easy identification of the cutoff point. In [Figure 6.4](#page-88-0) the straight horizontal line corresponds to the 95% R-square line. 214 variables lie below this line. This means that at least 214 most critical variables need to be retained to attain a 95% confidence value.

Figure 6.4. Loss of all measurements starting from the least critical to the most critical

In line with the number of critical variables identified in Section [6.3](#page-79-1) the most critical 151 variables in the R-squared method were selected for further analysis. These variables include 68 real and reactive power injections and 83 real and reactive power flows. These are presented in [Table 6.6.](#page-89-0)

Power injections	$P_G89, P_G80, P_G10, P_G65, P_G66, P_G26, P_L59, P_G100, Q_G8, P_G25,$ $P_G49, P_G59, Q_G77, Q_G59, P_G61, Q_G80, Q_G54, P_L80, Q_G100,$ Q_G 49, Q_G 66, P _L 54, Q _L 59, Q _G 12, P _G 69, Q _G 4, Q _G 10, Q _G 46, Q _G 18, Q_G40 , P _L 15, Q _G 27, P _L 49, P _L 56, Q _G 26, P _G 12, Q _G 65, P _L 60, P _L 90, P_L 62, Q _G 90, Q _G 15, P _L 78, P _L 11, P _L 74, P _L 76, Q _G 70, Q _G 31, P _L 70, P_L 92, Q _G 42, P _L 55, P _L 27, Q _G 6, P _L 77, Q _G 69, P _L 18, P _L 32, P _L 34, Q_G 76, Q_G 85, Q_G 36, P _L 82, P _L 45, Q _G 89, Q _G 104, P _L 6, P _L 1
Power flows	P9-10, P8-9, P5-8, P64-65, P26-30, P89-92, P17-30, P65-68, P80-81, P68-81, P37-38, P38-65, P23-25, P25-26, P30-38, Q8-9, P61-64, P59-63, P63-64, P25-27, P49-66, P49-66, P77-80, P68- 69, P65-66, P88-89, P60-61, P82-83, P23-24, P15-17, P4-5, P34- 37, P8-30, P85-89, P100-103, P5-6, P77-82, P69-70, P92-93, P92-94, P5-11, P23-32, P79-80, P94-95, P17-18, P85-88, P93-94, P89-92, P83-85, P76-77, P69-77, P69-75, Q77-82, P4-11, P11- 12, Q11-12, P94-96, P3-5, P89-90, P49-51, P77-80, Q4-5, P84- 85, P66-67, P75-77, Q17-30, Q94-100, P92-102, P74-75, P98- 100, P59-61, P24-70, Q25-26, P70-71, P49-50, P71-72, Q37-38, P101-102, P47-69, P22-23, P24-72, P83-84, Q38-65

Table 6.6. Critical Variables Identified using the R-squared Method

 r

77 critical locations of the IEEE 118-bus system corresponding to the locations of the critical variables were identified. These locations are represented by their bus numbers and are presented in [Table 6.7.](#page-90-0) Critical locations are candidate locations for PMU placement (for utilities that have not yet installed PMUs in their systems), prioritization of the measurement units in these areas for maintenance and calibration, and procurement of backup units for these locations in case of failure.

1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 15, 17, 18, 22, 23, 24, 25, 26, 27, 30, 31, 32, 34, 36, 37, 38, 40, 42, 45, 46, 47, 49, 50, 51, 54, 55, 56, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 88, 89, 90, 92, 93, 94, 95, 96, 98, 100, 101, 102, 103, 104

6.5. Comparison of Threshold and R-Squared Methods

The threshold and R-squared methods offer two avenues for identifying critical variables and locations of a power system. They have their pros and cons. With the threshold method, the number of monitored nodes (76) is 64% of the total available nodes (118) in the system, whereas with the R-squared method it is 65% (77/118). The major differences between the two methods are outlined below.

The threshold method is fairly easy and fast to implement; it took 28.3 seconds to run in MATLAB. Whereas the R-squared method takes a longer time because it requires two rounds of simulations in order to identify the critical variables. The first round of simulations took a total of 201.5 seconds, while the second round was done in 201 seconds in MATLAB. This total of 402.5 seconds observed runtime of the R-squared method does not include the time needed to rank the variables in order of criticality.

The threshold method considers the values of individual coefficients of a variable in the matrix of the retained eigenvectors F_{ret} as in equation [\(45\)](#page-64-2), whereas the R-squared method involves all the coefficients of a variable in the matrix of the retained eigenvectors \mathbf{F}_{ret} (via matrix multiplication), as shown in equation [\(43\)](#page-63-0).

The R-squared method is more flexible to the system planner and allows the arbitrary selection of a cutoff point (confidence value or number of desired variables) in the identification of critical variables, whereas the threshold method does not allow the system planner that level of flexibility.

[Table 6.8](#page-91-0) presents a comparison of the results described in Sections [6.3](#page-79-1) and [6.4.](#page-84-1) the critical variables match row describes how many of the critical variables classified using that method are also found in the set classified using the other method. The same goes for the critical locations match row. About 64% of the variables classified using the threshold method are also found in the set classified using the R-squared method, and vice versa, whereas for critical locations the values are greater than 85%.

Table 6.8. Comparison of Results Obtained Using the Threshold and R-squared Methods

Method	Threshold	R-squared
Number of variables	151	
Critical variables match	63.58%	63.58%
Critical locations match	86.84%	85.71%

However, about 87% of the critical locations found using the threshold method belong to the set found using the R-squared method. The reason for this disparity is that the two methods may identify different variables from the same location as critical variables, for instance, the threshold method identified Q_G1 as critical whereas the Rsquared method identified P_L1 as critical. These two variables belong to bus 1. The reason why neither the critical variables nor the locations matched at 100% is because in determining the criticality of a variable the threshold method compares individual coefficients with the calculated threshold whereas the R-squared method uses all the coefficients (via matrix multiplication) in F_{ret} . This means that in the R-squared method, a variable with relatively smaller coefficients that add up to a larger value will be selected ahead of a variable with coefficients that add up to a smaller value even if the latter variable has one very large coefficient.

Given the comparable match ratios of the two methods presented above, they could be used interchangeably. The choice, ultimately, may be driven by the level of flexibility or the ease of implementation desired by the system planner.

6.6. Effectiveness of Monitoring the Critical Locations

One of the main objectives of this dissertation is the identification of critical locations of a power system so that these locations can be monitored more closely and used for real-time assessment of system security. These critical locations are necessary and should be priority for steady state monitoring because they are sensitive to changes happening in the system and indicate when the system is not secure in the steady state. A power system is secure in the steady state if the variables of the system are within the normal operation range of the system. In the steady state, the bus voltage magnitude is a necessary measure for determining the security status of a power system and it is used in this verification exercise because voltage level criteria are straightforward. Many utilities use 95% as the normal minimum voltage level and 105% as the normal maximum. The values of the bus voltage magnitudes in the base case of the IEEE 118-bus system were considered to be the nominal values for the respective buses and the minimum (95%) and maximum (105%) voltage levels were calculated based on these values.

The thermal limit of transmission and distribution lines is another measure of steady state system security but it varies with season and / or time of day. It depends on real time weather conditions such as temperature and wind speed. As a result, it was not considered in this assessment.

To demonstrate the effectiveness of monitoring the critical locations of a power system using the IEEE 118-bus system, 48 scenarios, eight for each dispatch, were investigated. The eight scenarios for each dispatch investigated were created by varying the loads outside the assumed normal load levels $(\pm 40\%$ of the shoulder load level) of the system. The generators were allowed to adjust their real power output to accommodate the changes until they reach their maximum or minimum limits. The reactive devices (generators, synchronous condensers and switched shunts) were also allowed to vary until they ran out of range. These load flows were run using PSS/E.

The bus voltage magnitudes of all the buses in the system were inspected to determine if they violated security limits (were less than the minimum voltage level or greater than the maximum level). It was observed that if no critical location bus voltage magnitude violates the security margin, the system was secure. That is, in all the scenarios where violations were detected through inspection of all the bus voltages in the load flow cases, buses of critical locations of the system were always present in the set of buses with the voltage violations. This illustrates the effectiveness of identifying and monitoring the critical locations of power systems. The results of this investigation are presented in [Table 6.9](#page-95-0) – [Table 6.14.](#page-97-0) The # Violations column represents the total number of violations detected through the inspection of all the load flow case voltages for the given scenario, the # CLT column represents the number of violations at critical locations

present in the total set identified using the threshold method, whereas the # CLR column represents the number of violations at critical locations present in the total set identified using the R-squared method.

In Dispatch 1, violations at critical locations identified using the threshold method (# CLT) consisted 57%, 100%, 100%, 100%, 78% and 56% of the total violations for the 0.2, 0.3, 0.4, 1.5, 1.6, 1.7 and 1.8 load levels, respectively. Violations at critical locations identified using the R-squared method (# CLR) comprised 43%, 67%, 50%, 50%, 67% and 67% of the total violations for the 0.2, 0.3, 0.4, 1.5, 1.6, 1.7 and 1.8 load levels, respectively. No violations were observed for the 0.5 and 1.5 load levels for both methods in Dispatch 1. The average percentage value for critical locations identified using the threshold method for all the dispatches is 78% whereas for the R-squared method the average is 60%.

Essentially, while monitoring only the critical locations of its system, once the utility company observes a voltage violation at a critical location, it could use a load flow program to identify all the remaining locations with violations and use appropriate system operation procedures to correct the violations. Also, all the non-critical locations in the set with violations are connected to the critical locations: most are one bus away (in Dispatch 1, 6 unique locations out of 8 for # CLT and 3 unique locations out of 5 for # CLR) and a few are two buses away (in Dispatch 1, 2 unique locations out of 8 for # CLT and 2 unique locations out of 5 for # CLR) from the nearest critical location.

	Load Level	# Violations	# CLT	# CLR
	0.2	7		3
	0.3	3	3	$\mathcal{D}_{\mathcal{L}}$
High Voltage Violations	0.4	\mathcal{D}	\mathcal{D}	
	0.5	0		
	$0.6 - 1.4$		Normal Load Levels	
	1.5	0		Ω
	1.6	\mathcal{D}	$\mathcal{D}_{\mathcal{L}}$	
Low Voltage Violations	1.7	9		6
	1.8	18		12

Table 6.9. Voltage Violations in the IEEE 118-Bus System: Dispatch 1

Table 6.10. Voltage Violations in the IEEE 118-Bus System: Dispatch 1a

	Load Level	# Violations	# CLT	# CLR
	0.2	8	$\overline{4}$	3
	0.3	3	3	$\overline{2}$
High Voltage Violations	0.4	$\overline{2}$	$\overline{2}$	
	0.5	$\left(\right)$		
	$0.6 - 1.4$		Normal Load Levels	
	1.5	$\overline{4}$	2	2
Low Voltage Violations	1.6	11	5	$\overline{\mathbf{5}}$
	1.7	23	11	11
	1.8	28	14	16

	Load Level	# Violations	# CLT	# CLR
	0.2	┑		3
	0.3	3	3	2
High Voltage Violations	0.4	◠	↑	
	0.5			
	$0.6 - 1.4$		Normal Load Levels	
	1.5			
	1.6	ာ	\mathcal{D}	
Low Voltage Violations	1.7	Q		6
	1.8	18		12

Table 6.11. Voltage Violations in the IEEE 118-Bus System: Dispatch 1b

Table 6.12. Voltage Violations in the IEEE 118-Bus System: Dispatch 1c

	Load Level	# Violations	# CLT	# CLR
	0.2	6	4	3
High Voltage Violations	0.3	3	3	$\overline{2}$
	0.4	$\mathcal{D}_{\mathcal{L}}$	$\overline{2}$	
	0.5	\mathcal{O}		
	$0.6 - 1.4$		Normal Load Levels	
	1.5	10	8	10
Low Voltage Violations	1.6	21	14	17
	1.7	37	23	27
	1.8	49	31	34

	Load Level	# Violations	# CLT	# CLR
	0.2	6		3
	0.3	3	3	$\mathcal{D}_{\mathcal{L}}$
High Voltage Violations	0.4			
	0.5			
	$0.6 - 1.4$		Normal Load Levels	
	1.5	2	$\overline{2}$	
Low Voltage Violations	1.6	6		3
	1.7	13	Q	9
	1.8	30	21	22

Table 6.13. Voltage Violations in the IEEE 118-Bus System: Dispatch 1d

Table 6.14. Voltage Violations in the IEEE 118-Bus System: Dispatch 1e

	Load Level	# Violations	# CLT	# CLR
	0.2	7		3
High Voltage Violations	0.3	3	3	$\mathcal{D}_{\mathcal{L}}$
	0.4	\mathcal{D}	\mathcal{D}	
	0.5			
	$0.6 - 1.4$		Normal Load Levels	
	1.5			
Low Voltage Violations	1.6	2	\mathcal{D}	
	1.7	13		10
	1.8	23	13	15

Details of the voltage violations observed in this experiment are presented in [APPENDIX B.](#page-115-0)

6.7. ANN-Based State Estimation on the IEEE 118-Bus System

The IEEE 118-bus was successfully trained for one million epochs, using the BP neural network, for state estimation analysis. The inputs to the ANN are the 151 critical variables identified above using the threshold method. The outputs are the 236 voltage magnitudes and phase angles of the buses of the system. As described above, the loads were varied within the range of $\pm 40\%$ of the shoulder level of the power system. This permutation range covers the light load and peak load levels, which correspond to the normal operating range of power systems. A total of 1506 observations (patterns) were generated using PSS/E. Each observation comprises 151 data points for the inputs and 236 data points for the outputs, a total of 387 data points. From these, 502 patterns were used in training the ANN and 167 patterns from Dispatch 1 were used in the testing stage. An ANN with three processing layers was used: two hidden layers with 60 and 40 neurons, respectively and one output layer with 236 neurons. The BP network is designed using MATLAB Neural Network Toolbox.

The performance of the proposed ANN state estimator on the IEEE 118-bus system was measured by using the Mean Square Error (MSE) of the training and testing results, which is 2.27 \times 10⁻⁶ for the training data and is 2.04 \times 10⁻⁶ for the testing data. The plots of the ANN calculated voltage magnitudes and phase angles with respect to the actual voltage magnitudes and phase angles for buses 1, 10, 36, 67, 91, 118 of the IEEE 118-bus system are shown in [Figure 6.5](#page-101-0) and [Figure 6.6,](#page-102-0) respectively. These buses were chosen randomly. The calculated values are shown in dots whereas the actual (target) values are shown in connected lines (diagonal lines). Note that the actual values were obtained from PSS/E software.

In the phase angle plots it might be difficult to differentiate between the two sets of values because the calculated values are practically equal to their calculated counterparts. In the voltage magnitude plots, it might seem that the state estimator has a low accuracy, because the calculated values seem far from the actual values. This is due to the high resolution of the plot values. This resolution is necessary because the voltage magnitude values are in per unit (p.u.) and so do not have a wide range. Nevertheless, the error in predicting the values is very low. For instance, the farthest value from the diagonal line for the Vmag Bus 36 plot has the ANN (calculated) value of 0.9925 p.u., while the actual (target) value is 1 p.u. This translates to an error of 0.0075 p.u., which is about 0.75% error. Note that the R-squared calculation takes into consideration all the plotted data points for the variable. Therefore, the state estimator is very accurate.

The performance of the proposed state estimator on the IEEE 118-bus system was also measured using the coefficient of determination (R-squared values) of the voltage magnitudes and phase angles from the testing results. The R-squared values ranged from between 85.71% to 100%. 63 buses (2, 4, 5, 6, 7, 8, 10, 11, 13, 16, 23, 24, 25, 26, 27, 30, 31, 40, 42, 44, 46, 48, 49, 50, 51, 53, 54, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 72, 73, 80, 81, 87, 88, 89, 90, 91, 92, 96, 97, 98, 99, 100, 101, 102, 107, 111, 112, 113, 116 and 117) showed the highest R-squared value (100%) whereas bus 18 showed the lowest R-squared value (85.71%) for the voltage magnitudes. The values for the 63 buses with the highest R-squared values were fairly constant; therefore ANN could estimate them with high accuracy. 37 buses (4, 5, 8, 9, 15, 23, 28, 29, 33, 34, 39, 42, 43, 45, 49, 50, 53, 54, 57, 60, 62, 63, 69, 70, 72, 73, 80, 82, 86, 87, 91, 96, 99, 101, 106, 108 and 117) showed the highest R-squared value (100%) whereas bus 67 showed the lowest

R-squared value (99.9987%) for phase angles. The values for the 37 buses with the highest R-squared values have a nearly linear relationship with the input variables (phase angle relationship with real power); therefore ANN could estimate them with high accuracy. The R-squared values for all the buses are shown in [Table D.1](#page-136-0) in [APPENDIX](#page-136-1) [D.](#page-136-1)

The proposed ANN state estimator program was coded in MATLAB and run on Intel i7 64-bit Dell Precision T1500 machine running a Windows 7 operating system.

Figure 6.5. Plots of the actual voltage magnitudes (vertical axes) vs. the calculated voltage magnitudes for IEEE 118-bus system (horizontal axes)

Figure 6.6. Plots of the actual voltage phase angles (vertical axes) vs. the calculated voltage phase angles for IEEE 118-bus system (horizontal axes)

CHAPTER 7

CONCLUSION

In this dissertation, two novel concepts are introduced. These include the identification of critical variables and locations of a power system, and an ANN-based technique for power systems state estimation. Critical locations are parts of the system whose measurements provide information that reflect the general state of the system. Two methods for the identification of critical locations in the system were presented and the results from these methods were analyzed and compared.

Identification of critical locations would enable utility companies to smartly utilize limited resources, for instance, putting the identified critical locations ahead on the priority list for maintenance and calibration of measurement units; installation of backup units at these locations in case the main units fail; prioritization of measurements from the units in these locations for steady state monitoring and control of the system; and prioritization of these locations for PMU deployment (for utilities that have not yet installed PMUs in their systems).

The effectiveness of monitoring the critical locations of the system was demonstrated on the IEEE 118-bus system using 48 scenarios. Bus voltage magnitudes of critical locations of the system were in each of the scenarios where voltage violations were detected; that is, if no critical location bus voltage magnitude violates the security margin, the system was observed to be secure. This illustrates the effectiveness of identifying and monitoring the critical locations of power systems in ensuring power system steady state security. Also, the concept proposed in this paper presents the additional benefit of having the minimum number of monitored nodes reduced; for the IEEE 14-bus and 118-bus systems the number of monitored nodes is reduced to about 64% of the total available nodes in the respective systems. And this will potentially help reduce the financial costs of running power systems.

Additionally, the proposed ANN-based state estimator uses the identified critical variables of the system, and hence, employs fewer measurements than conventional approaches. The main advantage of this approach is that it is very accurate. Additionally, it is robust and eliminates the need for running observability analysis prior to executing state estimation; the ANN does this in one pass. The reduced number of measurements would make the system data more manageable, and also allow a more efficient monitoring and control of the system by the system operator. Therefore, the proposed technique provides a great alternative to the conventional methods and is ideal for smart grid applications.

REFERENCES

- [1] P. Kundur, Power System Stability and Control, McGraw-Hill, Inc, 1994.
- [2] A. Abur and A. G. Exposito, Power System State Estimation: Theory and Implementation, New York: CRC Press, 2004.
- [3] "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," April 2004.
- [4] A. G. Phadke and J. S. Thorp, Synchronized phasor measurements and their applications, New York: Springer Science+Business Media, LLC, 2008.
- [5] V. Madani, M. Parashar, J. Giri, S. Durbha, F. Rahmatian, D. Day, M. Adamiak and G. Sheble, "PMU placement considerations — A roadmap for optimal PMU placement," in *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*, March 2011.
- [6] Y. Liao, "Power transmission line parameter estimation and optimal meter placement," in *IEEE SoutheastCon 2010 (SoutheastCon), Proceedings of the*, March 2010.
- [7] M. Priyadharshini and R. Meenakumari, "Probabilistic approach based optimal placement of phasor measurement units via the estimation of dynamic vulnerability assessment," in *Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on*, March 2014.
- [8] Y. Song, S. Ma, L. Wu, Q. Wang and H. He, "PMU Placement Based on Power System Characteristics," in *Sustainable Power Generation and Supply, 2009. SUPERGEN '09. International Conference on*, April 2009.
- [9] B. Xu and A. Abur, "Optimal Placement of Phasor Measurement Units for State Estimation: Final Project Report," Power Systems Engineering Research Center, Ithaca, NY, October 2005.
- [10] A. Z. Gamm, I. N. Kolosok, A. M. Glazunova and E. S. Korkina, "PMU placement criteria for EPS state estimation," in *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, April 2008.
- [11] M. Rihan, M. Ahmad, M. S. Beg and M. A. Anees, "Robust and economical placement of phasor measurement units in Indian Smart Grid," in *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2013 IEEE*, November 2013.
- [12] A. S. Deese, T. Nugent and S. Coppi, "A comparative study of optimal PMU placement algorithms for cost minimization," in *PES General Meeting | Conference & Exposition, 2014 IEEE*, July 2014.
- [13] J. Wu, J. Xiong, P. Shil and Y. Shi, "Optimal PMU placement for identification of multiple power line outages in smart grids," in *Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on*, Aug. 2014.
- [14] J. I. Fadiran, S. Chowdhury and S. P. Chowdhury, "A multi-criteria optimal phasor measurement unit placement for multiple applications," in *Power and Energy Society General Meeting (PES), 2013 IEEE* , July 2013.
- [15] B. Gou and R. G. Kavasseri, "Unified PMU Placement for Observability and Bad Data Detection in State Estimation," in *Power Systems, IEEE Transactions on*, November 2014.
- [16] F. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model," *IEEE Trans. on Power Apparatus and Systems, vol. PAS-89, No. 1,* pp. 120- 125, January 1970.
- [17] E. Ghahremani and I. Kamwa, "Dynamic State Estimation in Power System by Applying the Extended Kalman Filter With unknown Inputs to Phasor Measurements," *IEEE Trans. on Power Systems, vol. 26, No.4,* pp. 2556-2566, November 2011.
- [18] S. Wang, W. Gao and A. S. Meliopoulos, "An Alternative Method for Power System Dynamic State Estimation Based on Unscented Transform," *IEEE Trans. on Power Systems, vol. 27, No.2,* pp. 942-950, May 2012.
- [19] A. Jain and N. Shivakumar, "Power system tracking and dynamic state estimation," in *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, 15-18 March 2009.
- [20] G. R. Krumpholz, K. A. Clements and P. W. Davis, "Power system observability: a practical algorithm using network topology," *IEEE Trans. on Power Apparatus and Systems, vol. PAS-99,* pp. 1534-1542, July/Aug. 1980.
- [21] A. Monticelli and F. F. Wu, "Network observability: Theory," *IEEE Trans. on Power Apparatus and Systems, Vol. PAS-104, No. 5,* pp. 1042-1048, May 1985.
- [22] O. Kosut, L. Jia, R. Thomas and L. Tong, "On malicious data attacks on power system state estimation," in *Universities Power Engineering Conference (UPEC), 2010 45th International* , Cardiff, Wales, Sept. 2010.
- [23] S. Haykin, Neural Networks: A Comprehensive Foundation, New Jersey: Prentice Hall, 1999.
- [24] D. V. Kumar, S. Srivastava, S. Shah and S. Mathur, "Topology processing and static state estimation using artificial neural networks," *IEE Proc. -Gener. Transm. Distrib., Vol. 143. No. 1,* pp. 99-105, January 1996.
- [25] D. Singh, J. P. Pandey and D. S. Chauhan, "Radial basis neural network state estimation of electric power networks," in *Electric Utility Deregulation, Restructuring and Power Technologies, 2004. (DRPT 2004). Proceedings of the 2004 IEEE International Conference on*, April 2004.
- [26] A. Jain, R. Balasubramanian and S. C. Tripathy, "Topological Observability: Artificial Neural Network Application Based Solution for a Practical Power System," in *Power Symposium, 2008. NAPS '08. 40th North American*, 28-30 Sept. 2008.
- [27] [Online]. Available: www.ferc.gov/oversight. [Accessed 10 October 2014].
- [28] J. Shlens, "A Tutorial on Principal Component Analysis," 7 April 2014. [Online]. Available: http://arxiv.org/pdf/1404.1100v1.pdf. [Accessed 12 May 2014].
- [29] Y. Guo, K. Li, Z. Yang, J. Deng and D. M. Laverty, "A novel radial basis function neural network principal component analysis scheme for PMU-based wide-area power system monitoring," *Electric Power Systems Research,* pp. 197-205, 2015.
- [30] M. Afshari, A. Tavasoli and J. Ghaisari, "A PCA-based Kalman estimation approach for system with colored measurement noise," in *Iranian Conference on Electrical Engineering*, Tehran, Iran, 2012.
- [31] S. Wang, T. He and X. Ma, "PCA-based preprocessing method of electronic data in power grid," in *International Conference on Computer, Mechatronics, Control and Electronic Engineering*, Changchun, 2010.
- [32] D. J. Burke and M. J. O'Malley, "A Study of Principal Component Analysis Applied to Spatially Distributed Wind Power," *IEEE Transactions on Power Systems,* vol. 26, no. 4, pp. 2084-2092, Nov. 2011.
- [33] E. Barocio, B. C. Pal, D. Fabozzi and N. F. Thornhill, "Detection and visualization of power system disturbances using principal component analysis," in *Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*, Rethymnon, Greece, 2013.
- [34] K. K. Anaparthi, B. Chaudhuri, N. F. Thornhill and B. C. Pal, "Coherency Identification in Power Systems Through Principal Component Analysis," *IEEE Transactions on Power Systems,* vol. 20, no. 3, pp. 1658-1660, Aug. 2005.
- [35] Y. Guo, K. Li, D. M. Laverty and Y. Xue, "Synchrophasor-Based Islanding Detection for Distributed Generation Systems Using Systematic Principal Component Analysis Approaches," *IEEE Transactions on Power Delivery,* vol. 30, no. 6, pp. 2544-2552, Dec. 2015.
- [36] J. E. Jackson, A User's Guide To Principal Components, New York: John Wiley and Sons, Inc., 1991.
- [37] A. C. Rencher, Methods of Multivariate Analysis, 2nd ed., New York: John Wiley and Sons, Inc., 2002.
- [38] I. T. Jolliffe, "Discarding Variables in a Principal Component Analysis. I: Artificial Data," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 21, no. 2, pp. 160-173, 1972.
- [39] I. T. Jolliffe, "Discarding Variables in a Principal Component Analysis. II: Real Data," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 22, no. 1, pp. 21-31, 1973.
- [40] "Power Systems Test Case Archive," [Online]. Available: https://www.ee.washington.edu/research/pstca/.
- [41] "Load Modeling Guide for ISO New England Network Model," 12 December 2012. [Online]. Available: http://www.isone.com/rules_proceds/isone_plan/othr_docs/load_modeling_guide.pdf. [Accessed 8 June 2016].
- [42] "MATLAB," The MathWorks, Inc., [Online]. Available: http://www.mathworks.com.
- [43] W. L. Martinez and M. Cho, Statistics in MATLAB: A Primer, Boca Raton: CRC Press, 2014.
- [44] [Online]. Available: http://www.al-roomi.org/power-flow/118-bus-system. [Accessed 8 August 2016].
- [45] J. J. Grainger and W. D. Stevenson, Power System Analysis, New York: McGraw-Hill, Inc., 1994.
- [46] H. Demuth and M. Beale, Neural Networks Toolbox User's Guide, Natick, MA: The MathWorks, Inc., 1996.
- [47] X. Chen and R. S. Womersley, "Existence of solutions to systems of underdetermined equations and spherical designs," *SIAM J. Numer. Anal.,* vol. 44, no. 6, pp. 2326-2341, November 2006.
- [48] "Smart Grid | Department of Energy," U.S. Department of Energy, [Online]. Available: http://energy.gov/oe/technology-development/smart-grid. [Accessed 26 February 2014].
- [49] G. Alefeld, A. Gienger and F. Potra, "Efficient numerical validation of solutions of nonlinear systems," *SIAM J. Numer. Anal,* vol. 31, no. 1, pp. 252-260, February 1994.
- [50] J. Hao and W. Xu, "Extended transmission line loadability curve by including voltage stability constrains," in *Electric Power Conference, 2008. EPEC 2008. IEEE Canada* , Vancouver, BC, Oct. 2008.
- [51] GE Energy, Mechanics of running PSLF dynamics / Dynamic simulation application using PSLF, Schenectady, NY, August 2011.
- [52] Y.-c. Li and K. Gao, "A KPCA and SVR Based Dynamic State Estimation Method for Power System," in *Computer Modeling and Simulation, 2010. ICCMS '10. Second International Conference on*, 2010.
- [53] Z. Qu, J. Meng and W. Yuan, "Preprocessing of online analysis in power grid based on state estimation optimized by PCA," in *Computer Science and Software Engineering, 2008 International Conference on*, 2008.
- [54] E. Manitsas, R. Singh, B. C. Pal and G. Strbac, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurement

Modeling," *IEEE Transactions on Power Systems,* vol. 27, no. 4, pp. 1888-1896, November 2012.

- [55] M. Biserica, Y. Besanger, R. Caire, O. Chilard and P. Deschamps, "Neural Networks to Improve Distribution State Estimation—Volt Var Control Performances," *IEEE Transactions on Smart Grid,* vol. 3, no. 3, pp. 1137-1144, September 2012.
- [56] A. Kumar and S. Chakrabarti, "ANN-based Hybrid State Estimation and Enhanced Visualization of Power Systems," in *Innovative Smart Grid Technologies - India (ISGT India)*, Dec. 2011.
- [57] A. Gómez-Expósito, A. de la Villa Jaén and C. Gómez-Quiles, "A taxonomy of multi-area state estimation methods," *Electric Power Systems Research,* vol. 81, pp. 1060-1069, April 2011.
- [58] D. Junce and C. Zexiang, "Mixed Measurements State Estimation Based on Wide-Area Measurement System and Analysis," in *Transmission and Distribution Conference and Exhibition: Asia and Pacific*, 2005.
- [59] L. Xie, D.-H. Choi, S. Kar and H. V. Poor, "Fully Distributed State Estimation for Wide-Area Monitoring Systems," *Smart Grid, IEEE Transactions on,* vol. 3, no. 3, pp. 1154,1169, Sept. 2012.
- [60] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic and A. Phadke, "Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks," in *Proceedings of the IEEE*, Jan. 2011.
- [61] M. Shahraeini and M. H. Javidi, "Wide Area Measurement Systems," in *Advanced Topics in Measurements*, InTech, 2012, pp. 303-322.
- [62] S. Azizi, A. S. Dobakhshari, A. M. Ranjbar and S. A. Nezam Sarmadi, "Optimal PMU Placement by an Equivalent Linear Formulation for Exhaustive Search," *Smart Grid, IEEE Transactions on,* vol. 3, no. 1, pp. 174 - 182, March 2012.
- [63] E. Caro and A. J. Conejo, "State estimation via mathematical programming: a comparison of different estimation algorithms," *IET Generation, Transmission & Distribution,* vol. 6, no. 6, pp. 545-553, 2012.
- [64] H. Livani and C. Y. Evrenosoglu, "State Forecasting of Power Systems with Intermittent Renewable Sources Using Viterbi Algorithm," in *IEEE Power and Energy Society General Meeting*, San Diego, CA, 2011.
- [65] J.-M. Lin, S.-J. Huang and K.-R. Shih, "Application of Sliding Surface-Enhanced Fuzzy Control for Dynamic State Estimation of A Power System," *IEEE Transactions on Power Systems,* vol. 18, no. 2, pp. 570-577, 2003.
- [66] M. Nejati, N. Amjady and H. Zareipour, "A New Stochastic Search Technique Combined With Scenario Approach for Dynamic State Estimation of Power Systems," *IEEE Transactions on Power Systems,* vol. 27, no. 4, pp. 2093-2105, 2012.

APPENDIX A: DIAGRAM OF THE IEEE 118-BUS SYSTEM

Figure A.1. Diagram of the IEEE 118-bus system

This diagram was obtained from reference [44].

APPENDIX B: ADDITIONAL RESULTS FOR BUS VOLTAGE VIOLATIONS FOR THE IEEE 118-BUS SYSTEM

Table B.1. Base Case Bus Voltage Magnitudes of the IEEE 118-Bus System and the

Minimum (95%) and Maximum (105%) Allowable Voltage Levels

Table B.2. Dispatch 1 High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

Table B.3. Dispatch 1a High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

Table B.4. Dispatch 1b High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

Table B.4. continued

Table B.5. Dispatch 1c High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

Table B.6. Dispatch 1d High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

Table B.6. continued

Table B.7. Dispatch 1e High and Low Voltage Violations

Highlighted numbers indicate violation. Load levels indicate percentage of the base case

APPENDIX C: INDEPENDENT SYSTEM OPERATOR OF NEW ENGLAND

(ISO-NE) REPORT SUBMITTED TO FEDERAL ENERGY

REGULATORY COMMISSION (FERC)

Figure C.1. ISO-NE daily report submitted to FERC

This report was obtained from reference [27].

APPENDIX D: R-SQUARED VALUES OF THE IEEE 118-BUS SYSTEM

Table D.1. R-Squared Values of the IEEE 118-Bus System

	R-Squared	R-Squared		R-Squared	R-Squared
Bus	Values for	Values for	Bus	Values for	Values for
Number	Voltage	Phase	Number	Voltage	Phase
	Magnitudes	Angles		Magnitudes	Angles
75	0.998894		97		
76	0.997662		98		
77	0.993934		99		
78	0.998248		100	1	
79	0.999264		101	1	
80			102		1
81			103	0.969565	1
82	0.999874		104	0.980354	1
83	0.99996		105	0.991686	1
84	0.999609		106	0.998564	
85	0.999168		107		
86	0.999726		108	0.993753	$\mathbf{1}$
87		1	109	0.990215	1
88			110	0.968182	1
89			111		
90			112		
91			113		1
92			114	0.999156	1
93	0.999845		115	0.99646	
94	0.999784		116		1
95	0.999924		117		1
96		1	118	0.998503	1

Table D.1. continued

BIOGRAPHY OF THE AUTHOR

Amamihe Onwuachumba was born in Orlu, Nigeria. He graduated from Federal Government Academy (Centre for the Gifted and Talented), Suleja, Nigeria in 2000. He received his Bachelor of Science and Master of Science degrees in Electrical and Electronics Engineering in 2008 and 2010, respectively, from Moscow Power Engineering Institute (National Research University), Moscow, Russia.

Amamihe started his graduate study in Electrical and Computer Engineering at the University of Maine in January 2011 and was appointed a Graduate Assistant. He interned with RLC Engineering, Hallowell, ME in the Power Systems Studies Group in the summers of 2013, 2014 and 2015. In August 2013, he received a Master of Science degree in Electrical and Computer Engineering from the University of Maine. His research interests are in power system state estimation, power system observability analysis, power system stability analysis, application of neural networks and fuzzy systems in the smart grid, design and applications of power electronic devices. He has several scientific papers published by the Institute of Electrical and Electronics Engineers (IEEE), Journal of Energy and Power Engineering, Moscow Power Engineering Institute Publishing House, Moscow, Russia and Herald of the Electrical Engineering Department for Students' Scientific Work, Donetsk, Ukraine.

Amamihe was awarded the 2014-2015 Dean's Fellowship by the University of Maine Graduate School. He is a member of IEEE, Golden Key International Honour Society and the National Society of Black Engineers (NSBE). He is a candidate for the Doctor of Philosophy degree in Electrical and Computer Engineering from the University of Maine in August 2016.