PHYSICS-BASED AND DATA-DRIVEN ANALYTICS FOR ENHANCED PLANNING AND OPERATIONS IN POWER SYSTEMS WITH DEEP RENEWABLE PENETRATION

A Dissertation

by

MENG WU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Le Xie
Committee Members,	Mladen Kezunovic
	P. R. Kumar
	Jianhua Huang
Head of Department,	Miroslav M. Begovic

December 2017

Major Subject: Electrical Engineering

ABSTRACT

This dissertation is motivated by the lack of combined physics-based and data-driven framework for solving power system challenges that are introduced by the integration of new devices and new system components. As increasing number of stochastic generation, responsive loads, and dynamic measurements are involved in the planning and operations of modern power systems, utilities and system operators are in great need of new analysis framework that could combine physical models and measuring data together for solving challenging planning and operational problems.

In view of the above challenges, the high-level objective of this dissertation is to develop a framework for integrating measurement data into large physical systems modeled by dynamical equations. To this end, the dissertation first identifies four critical tasks for the planning and operations of the modern power systems: the data collection and pre-processing, the system situational awareness, the decision making process, as well as the post-event analysis. The dissertation then takes one concrete application in each of these critical tasks as the example, and proposes the physics-based/data-driven approach for solving the challenging problems faced by this specific application.

To this end, this dissertation focuses on solving the following specific problems using physics-based/data-driven approaches. First, for the data collection and pre-processing platform, a purely data-driven approach is proposed to detect bad metering data in the phasor measurement unit (PMU) monitoring systems, and ensure the overall PMU data quality. Second, for the situational awareness platform, a physics-based voltage stability assessment method is presented to improve the situational awareness of system voltage instabilities. Third, for the decision making platform, a combined physics-based and data-driven framework is proposed to support the decision making process of PMU-based

power plant model validation. Forth, for the post-event analysis platform, a physics-based post-event analysis is presented to identify the root causes of the sub-synchronous oscillations induced by the wind farm integration.

The above problems and proposed solutions are discussed in detail in Section 2 through Section 5. The results of this work can be integrated to address practical problems in modern power system planning and operations.

DEDICATION

To my parents, Mr. Zhiwei Wu and Ms. Shulan Liu

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Le Xie for the valuable guidance and constant support of my Ph.D study and related research.

I would also like to thank my committee members: Dr. Mladen Kezunovic, Dr. P. R. Kumar and Dr. Jianhua Huang, for their insightful comments and questions which widened my research scope from various perspectives.

I would also like to express my gratitude to my internship supervisors and colleagues at ISO New England: Mr. Qiang (Frankie) Zhang, Dr. Xiaochuan Luo, Dr. Slava Maslennikov, Dr. Eugene Litvinov, Mr. Bilgehan Donmez, and Mr. Dean LaForest, as well as the vendor representative at the PowerTech Labs, Dr. Xi Lin, for their valuable supports, discussions, and suggestions, which improved the quality of my research work.

I would also like to thank Dr. Lin Cheng of Tsinghua University, China, Dr. Rongfu Sun of China Jibei Power Grid, and Dr. Thomas Overbye of Texas A&M University, for their insightful discussions and valuable supports, which stimulated my research interests and broadened my research horizon.

My sincere thanks also go to our research group members: Dae-hyun Choi, Yingzhong (Gary) Gu, Omar Urquidez, Yang Chen, Yun Zhang, Anupam Thatte, Sadegh Modarresi, Xinbo Geng and Hao Ming, Hung-Ming Chou, Tong Huang, Yuanyuan (Crystal) Li, Yuqi Zhou, Xiaowen Lai, Dian Chen, Shuang Feng, and Yang Bai.

Last but not least, I would like to thank my parents for their constant support and unconditional love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was conducted with the guidance from a dissertation committee consisting of Professor Le Xie of the Department of Electrical and Computer Engineering, Professor Mladen Kezunovic of the Department of Electrical and Computer Engineering, Professor P. R. Kumar of the Department of Electrical and Computer Engineering, and Professor Jianhua Huang of the Department of Statistics.

Part of the work presented in Section 4 was performed in collaboration with Mr. Qiang (Frankie) Zhang, Dr. Xiaochuan Luo, Dr. Slava Maslennikov, and Dr. Eugene Litvinov at ISO New England. The test case analyzed in Section 5 was provided by China Jibei Power Grid, Beijing, China. Part of the work in Section 5 was performed in collaboration with Dr. Lin Cheng from Tsinghua University, Beijing, China, and Dr. Rongfu Sun from China Jibei Power Grid, Beijing, China.

All the other work conducted for the dissertation was completed by the student independently.

Funding Sources

This research is supported by the following funding sources:

1. Power Systems Engineering Research Center.
2. National Science Foundation (NSF) ECCS-1150944, DGE-1303378, and IIS-1636772.
3. NSF-DGE 1303378, TAMU-NSF China Grant.
4. Chinese National Natural Science Fund (51177078).

NOMENCLATURE

$\Delta\omega_w, \Delta\omega_g$	Mismatch of wind turbine and doubly-fed induction generator (DFIG) rotating speed from their steady-state values.
$\Delta\omega_{wg}$	Difference between $\Delta\omega_w$ and $\Delta\omega_g$.
τ_w, τ_g	Wind turbine mechanical torque and DFIG electro-magnetic torque.
i_{qs}, i_{ds}	DFIG stator current on q and d axes.
i_{qr}, i_{dr}	DFIG rotor current on q and d axes.
v_{qs}, v_{ds}	DFIG stator voltage on q and d axes.
v_{qr}, v_{dr}	DFIG rotor voltage on q and d axes.
P_e	DFIG total real power output.
Q_s	DFIG stator reactive power output.
P_r, P_g	Real power of DFIG rotor-side converter and grid-side converter.
C_{lnk}, v_{lnk}	Capacitance and voltage of DFIG DC-link capacitor.
X_{ls}, X_{lr}	DFIG stator and rotor leakage reactance.
X_m	DFIG magnetizing reactance.
R_s, R_r	DFIG stator and rotor resistance.
D, K	Damping and stiffness of wind turbine mechanical shaft.
J_w, J_g	Inertia of wind turbine mechanical shaft 2-mass model.
R_{nt}, L_{nt}	Transmission line resistance and inductance.

X_{xf}	Equivalent transformer reactance of transmission system.
C_{se}, C_{sh}	Series and shunt capacitance of transmission system.
L_i, C_i	Inductance and capacitance of ith inductor/capacitor in the network.
$i^q_{L_i}, i^d_{L_i}$	Current of ith network inductor on q and d axes.
$v^q_{L_i}, v^d_{L_i}$	Voltage of ith network inductor on q and d axes.
$v^q_{C_i}, v^d_{C_i}$	Voltage of ith network capacitor on q and d axes.
$i^q_{C_i}, i^d_{C_i}$	Current of ith network capacitor on q and d axes.
v^i_{qinf}, v^i_{dinf}	Voltage of ith remote power system on q and d axes.
$\frac{dx}{dt}, \dot{x}$	Time derivative of x.
v_{wind}	Wind speed.
δ_i	Dynamic mechanical rotating angle of i^{th} mass of wind turbine mechanical shaft.
ω_i, ω_i^*	Dynamic and steady-state mechanical rotating speed of i^{th} mass of wind turbine mechanical shaft.
$\Delta\omega_i$	Mismatch between ω_i and ω_i^*.
ω_r, ω_r^*	Dynamic and steady-state DFIG electrical rotor speed.
ω_i^{rate}	Rated mechanical rotating speed of i^{th} mass of wind turbine mechanical shaft.
ω_g^{rate}	Rated mechanical rotating speed of the generator.
ν_i	Transformation ratio of the gearbox associated with i^{th} mass of wind turbine mechanical shaft.
ω	Rotating speed of synchronous reference frame.
ω_b	Rotating speed corresponding to system rated frequency.
$pole$	Number of poles in DFIG.

τ_i	External mechanical torque applied to i^{th} mass of wind turbine mechanical shaft.
J_i	Moment of inertia of i^{th} mass of wind turbine mechanical shaft.
$D_{i,j}$	Damping coefficient of the wind turbine mechanical shaft sections between i^{th} mass and j^{th} mass.
$K_{i,j}$	Stiffness coefficient of the wind turbine mechanical shaft sections between i^{th} mass and j^{th} mass.
P_{rsc}, P_{gsc}	Real power delivery through rotor-side converter and grid-side-converter.
$P_{loss}^{converter}$	Real power loss of DFIG back-to-back converter.
Q_g, Q_g^{ref}	Reactive power delivery through grid-side converter and its reference value.
Q_s	Reactive power delivery through DFIG stator.
P_e, Q_e	DFIG real and reactive power output.
v_s, v_r, v_g	terminal voltage of DFIG stator, rotor and grid-side-converter.
v_{dc}	DC-link voltage of DFIG back-to-back converter.
i_s, i_r, i_g	DFIG stator, rotor and grid-side converter current.
i_l	DFIG terminal transmission line current.
i_{ql}, i_{dl}	DFIG terminal transmission line current on q and d axes.
i_{qg}, i_{dg}	DFIG grid-side converter current on q and d axes.
v_{qg}, v_{dg}	Grid-side-converter terminal voltage on q and d axes.
$\vec{x}, \lvert\vec{x}\rvert$	Phasor of electrical variable x and its magnitude.
R_g, X_g	resistance and reactance of DFIG back-to-back converter.

R_l, X_l	resistance and reactance of DFIG terminal transmission line.
r_r, r_s	resistance of DFIG rotor and stator circuits.
X_{rr}, X_{ss}	reactance of DFIG rotor and stator circuits.
C_{dc}	DC-link capactance of DIFG back-to-back converter.
$\dot{x}$	time-derivative of variable x.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

1.1 Modern Power Grids in The Data-Rich Environment

In the past few decades, increasing number of new devices are being integrated into the electric power systems, bringing new device models as well as new measurement data into the power system planning and operations. Different from traditional power systems which interconnect conventional power plant and non-responsive loads through the transmission power grids monitored by the supervisory control and data acquisition (SCADA) system, modern power systems consist of stochastic power generations such as wind farms and solar panels, responsive loads with distributed generations and price-responsive demands, as well as modern metering devices, such as the phasor measurement units (PMUs), which provides massive amount of data and information on the dynamics of the transmission power systems. These new generations, demands, and metering devices offer great opportunities for the entire power system business to generate, deliver, and consume electricity in a greener, smarter, more secure and more economical way. In the meantime, the stochastic nature of the renewable generations, the market participation of the active loading centers, and the "curse of dimensionality" introduced by the massive amount of PMU metering data also bring new challenges to the planning and operations of the modern power grids in this data-rich environment.

1.2 Challenges for Planning and Operations of The Data-Rich Power Grids

Figure 1.1 identifies several critical tasks for the planning and operations of the modern power grids in the data-rich environment. With the presentation of the massive amount of metering data and modeling parameters for system-level applications, a reliable data collection and pre-processing platform is needed to ensure the availability and accuracy of all the metering data and modeling parameters. After the initial collection and pre-processing

of all the metering data and modeling parameters, a situational awareness platform is established to assess the system security and reliability using the metering data and modeling parameters, and provide necessary alarms when critical conditions are identified. Once the critical conditions are identified, a decision making process is then involved to diagnose the problem and make necessary corrections. For problems that fail to be corrected through the decision making process, system instability events may be induced. These events are then recorded in the form of metering data (such as PMU curves) and saved for offline post-event analysis, in order to prevent them from happening again.

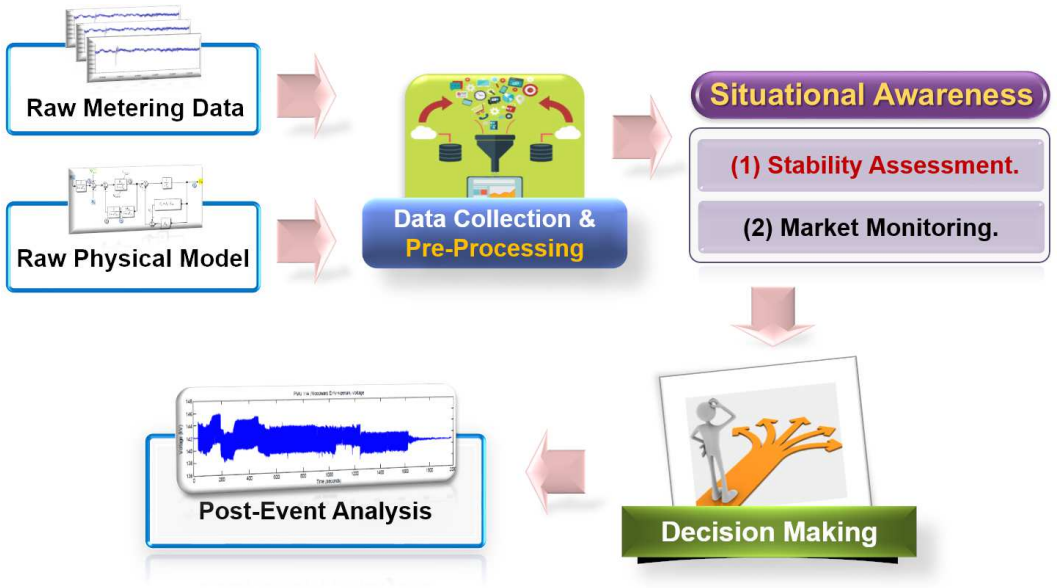

Figure 1.1: Critical tasks for the planning and operations of the modern power grids in the data-rich environment.

In the above process for power system planning and operations, the following critical challenges are identified for each of the major tasks:

1) A data collection and pre-processing platform: how to ensure the data quality of both metering data and modeling parameters, when the massive amount of data is being integrated into the modern power grids?

2) A situational awareness platform: how to take advantage of the recent developments in the control and optimization societies, in order to improve the situational awareness in modern power systems with stochastic generations and active loading centers?

3) A decision making platform: how to combine the physics-based and data-driven analytics to build an enhanced decision making platform and support the power system planning and operations in the data-rich environment?

4) A post-event analysis platform: how to better analyze the root cause of the system instability events induced by stochastic generations such as wind farms and solar panels?

Motivated by the above critical challenges faced by utilities and system operators for the planning and operations of the data-rich power grids, this dissertation takes one specific example in each of the four critical tasks, and provides a framework to apply combined physics-based and data-driven analytics for the modern power system planning and operations.

1.3 Combined Physics-Based and Data-Driven Analytics for Power System Planning and Operations

In this dissertation, a framework is proposed for combined physics-based and data-driven analytics for the modern power system planning and operations. The framework is introduced through a series of concrete examples for solving the above critical challenges using combined physics-based and data-driven techniques. To be specific, the following examples are presented to demonstrate the proposed framework:

1) The data collection and pre-processing platform: a purely data-driven approach is proposed to detect bad metering data in the PMU monitoring systems, and ensure the

overall PMU data quality.

2) The situational awareness platform: a physics-based voltage stability assessment approach is presented to improve the situational awareness of system voltage instabilities.

3) The decision making platform: a combined physics-based and data-driven framework is proposed to support the decision making process of PMU-based power plant model validation.

4) The post-event analysis platform: a physics-based post-event analysis is presented to identify the root causes of the sub-synchronous oscillations induced by the wind farm integration.

1.4 Suggested Contributions

The main contributions of this dissertation are suggested as follows:

1. A framework is proposed for integrating measurement data into large physical systems modeled using differential-algebraic equations.

2. A data-driven algorithm is developed for improving the data quality of PMU measurement systems.

3. A comprehensive approach is proposed for exploring the power flow solvability problem for the purpose of strengthened voltage stability monitoring of modern power grids.

4. A decision support framework is proposed for automating the entire process of PMU-based power plant model validation, with the capability of batch power plant model validation and automatic diagnosis of power plant modeling errors.

5. An enhanced analysis is presented for identifying the root causes of the wind-farm-induced sub-synchronous oscillations.

1.5 Dissertation Organization

The rest of the dissertation is organized as follows. Section 2 presents the purely data-driven approach to detect bad metering data in the PMU monitoring systems. Section 3 proposes the exploration of the power flow solvability problem for enhanced system voltage stability monitoring. Section 4 introduces the automation platform for the PMU-based power plant model validation as well as the diagnosis framework for power plant modeling errors. Section 5 presents the post-event analysis for identifying the root causes of the wind-farm-induced sub-synchronous oscillations. Section 6 provides concluding remarks to this dissertation.

2. DATA-DRIVEN PRE-PROCESSING FOR IMPROVED PMU DATA QUALITY*

2.1 Motivation

In recent years, there has been significant deployment of phasor measurement units (PMU) around the world. Compared with traditional metering units in supervisory control and data acquisition (SCADA) systems, PMUs provide measurements with much higher sampling rates. The high-resolution PMU measurements contain rich information on system dynamics, which stimulates the development of advanced analytics, such as dynamic state estimation [5], PMU-based model validation [6], and wide-area control and protection [7, 8]. However, as a large amount of data is streaming into the control center, the PMU data quality problem becomes one of the major challenges for system operators. Generally speaking, low-quality PMU data represents data that cannot accurately reflect the underlying system behavior. The inaccuracy can be caused by various problems such as sensing noises, data loss, and the global positioning system (GPS) time errors. As an example, the ratio of low-quality PMU data, reported by California Independent System Operator (ISO) in 2011, ranged from 10% to 17% [9]. In 2013, the ratio of low-quality PMU data in China was reported to range from 20% to 30% [10]. The online data quality monitoring of PMUs becomes a major barrier for any advanced PMU-based analytics.

In order to improve data quality of PMU systems, various methods have been proposed. In [11], a PMU-based state estimator is introduced to detect phasor angle bias and current magnitude scaling problems. In [12], the Kalman filtering technique is applied to detect low-quality PMU data. Both state estimator and Kalman filter-based approaches require prior knowledge on system topology and model parameters for detecting low-quality data.

*This section is in part a reprint with permission from Meng Wu and Le Xie of the material in the paper: "Online Detection of Low-Quality Synchrophasor Measurements: A Data Driven Approach", in *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2817-2827, July 2017 [1]. Copyright 2017 IEEE.

Therefore, the detection accuracy of the above approaches may be affected when gross errors are presented in system topology or parameters. Furthermore, these methods cannot operate successfully when state estimation diverges because of gross measurement errors, system physical disturbances, or stressful operating conditions. In [13, 14], serveral logic-based low-quality data detection schemes are presented. These approaches compare PMU data with certain thresholds, apply high-noise filters to raw PMU measurements, and perform cross-checking on PMU measurements obtained in nearby physical locations, in order to detect abnormal PMU measurements. However, these pre-defined logics may be rendered ineffective when large disturbances occur in the studied power grid. In [15], clustering algorithms are applied to extract information from power system time-varying data. These clustering techniques could potentially be applied to detect system anomalies such as low-quality PMU data or system physical disturbances. Reference [16, 17] pioneered a purely data-driven method to improve PMU data quality. This method applies low-rank matrix factorization techniques to detect and repair low-quality PMU data. It has satisfactory performance under both normal and fault-on operating conditions. However, since the matrix factorization techniques bear high computational burden such as nonlinear optimizations, it becomes a challenge when applied for real-time applications.

In view of the current efforts on PMU data quality improvement, this section presents a data-driven approach for online detection of low-quality PMU measurements. It leverages the spatio-temporal similarities among multi-time-instant PMU data, and applies density-based local outlier detection technique to detect low-quality PMU measurements. The major advantages of the proposed approach are summarized as follows. (1) This is a purely data-driven approach, without requiring any prior knowledge on system topology or model parameters, which eliminates the potential misdetections caused by inaccurate system information; (2) the proposed approach can operate without any converged state estimation results and is suitable for filtering out gross measurement errors for advanced

power system analytics; (3) the proposed approach has fast computational speed, which could be beneficial for real-time applications; and (4) the algorithm is able to perform detections under both normal and fault-on operating conditions. The proposed detection algorithm differentiates high-quality PMU data recorded during system physical disturbances (faults) from the low-quality data, which avoids potential false alarms caused by physical disturbances.

The rest of the section is organized as follows. Section 2.2 presents the problem formulation of the low-quality PMU data detection issue; Section 2.3 discusses the proposed data-driven approach for low-quality PMU data detection; Section 2.4 presents case study results to verify the proposed approach; Section 2.5 provides concluding remarks to this section.

2.2 Problem Formulation

This section presents the key features differentiating low-quality PMU measurements from the high-quality ones. Based on these features, low-quality PMU measurements are formulated as *spatio-temporal outliers* among high-quality measurements in the power grid. Accordingly, the low-quality PMU data detection problem is formulated to be a spatio-temporal outlier detection problem.

2.2.1 Key Features of High-Quality and Low-Quality PMU Data

Let $m \times n$ matrix M denote a set of PMU measurements collected from n PMU channels of the same type (i.e., all of them are voltage/current/power channels), within m time instants. This measurement matrix can be decomposed into the following two matrices:

$$M = L + D \tag{2.1}$$

where the k^{th} column of matrix L represents the accurate measurements corresponding to the k^{th} PMU channel in M, and D denotes the matrix containing inaccurate information caused by data quality problems. Each nonzero entry D_{ij} represents a measurement error of the j^{th} PMU channel at time instant i. Here, a PMU channel represents one of the following electrical quantities obtained by a PMU: voltage magnitude, voltage phasor angle, current magnitude, current phasor angle, real power, and reactive power. Therefore, M_{ij} is a real number instead of a complex number.

Definition 1. *M_{ij} is defined to be low-quality PMU data if its corresponding $|D_{ij}| > \tau$, where τ is a positive threshold to determine low-quality data.*

It has been shown in [16,17], when low-quality PMU data is presented in certain power system, the rank of matrix M would be higher than the rank of matrix L, due to the nonzero entries in matrix D. This phenomenon indicates the linear dependency (similarity) among PMU measurements would be weakened by data quality problems.

In order to demonstrate the above property of low-quality PMU measurements, Figure 2.1 shows voltage magnitude curves measured by two PMUs with nearby physical locations. Both curves were recorded at the same time period, when a line-tripping fault was presented in the system (from 3s to 5s). The upper curve contains low-quality data at around 1s. By observing only the upper curve, it is difficult to confirm whether the data spikes are caused by physical disturbance or data-quality problem, since all the data spikes have outlier behavior compared with their temporal neighbors. However, by comparing multiple PMU curves obtained in different locations of the system, it would be possible to differentiate spikes caused by data-quality problems and those caused by disturbances, since spikes caused by data-quality problems are outliers compared with their spatial neighbors, while spikes caused by disturbances appear in curves recorded by multiple PMUs and therefore cannot be considered as outliers compared with their spatial

neighbors.

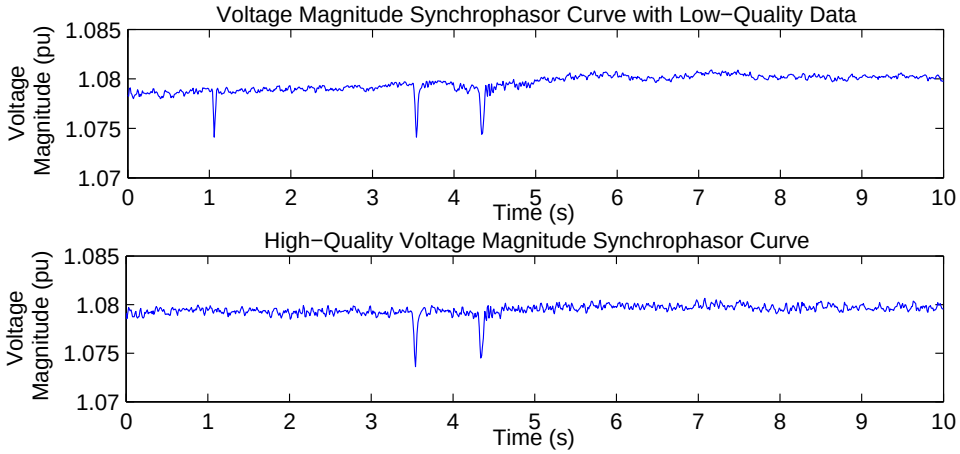

Figure 2.1: Comparison between PMU curves with and without low-quality data [1].

The above observations can be summarized as the following key features of low-quality and high-quality PMU data under normal/fault-on operating conditions:

Feature 1. *Both low-quality PMU measurements and fault-on PMU measurements exhibit weak temporal similarities with the measurements obtained at the neighboring time periods, while high-quality PMU measurements obtained during normal operating conditions exhibit strong temporal similarities with the measurements obtained at the neighboring time periods.*

Feature 2. *Low-quality PMU measurements exhibit weak spatial similarities with the measurements obtained by the neighboring PMUs at the same time period, while fault-on PMU measurements exhibit strong spatial similarities with the measurements obtained by the neighboring PMUs at the same time period.*

It should be noted that strong electrical connections among neighboring PMUs are required in order for the above features to be valid. Therefore, higher PMU measurement

redundancy would lead to better accuracy in low-quality data detection, and lack of measurement redundancy could cause miss detections for the proposed algorithm. As more and more PMUs are being installed in power grids around the world, the measurement redundancy would be enhanced, and therefore the detection accuracy of the proposed algorithm would be improved.

2.2.2 Formulation of Low-Quality PMU Data as Spatio-Temporal Outliers

According to the discussions in the previous section, low-quality PMU measurements have weaker spatio-temporal similarities with their high-quality neighbors, under both normal and fault-on operating conditions. Therefore, these low-quality measurements can be formulated as spatio-temporal outliers among all the PMU measurements in the system. With a proper definition of similarity metrics for PMU curves, the degree of similarity between two PMU curves can be quantified, and data-mining techniques can be applied to detect the spatio-temporal outliers whose degrees of similarity are significantly different from other PMU curves.

For a measurement matrix M obtained within a certain period of time, general steps to formulate the detection problem are described as follows:

***Step 1*:** Define a proper similarity metric (distance function) $f(M_i, M_j)$, which quantifies the degree of similarity between the i^{th} and j^{th} column of M.

***Step 2*:** Map each column of M (a data curve obtained from certain PMU channel) to the space S where the distance function $f(M_i, M_j)$ is defined. Each column of M can be represented as a point in S.

***Step 3*:** Examine the outlier behavior of the points in S, according to distance function $f(M_i, M_j)$. Points lying far from the majority are more likely to be outliers with low-quality data.

Figure 2.2 demonstrates the above formulation through a simple example. Three 2×8 measurement matrices $M(1)$, $M(2)$, $M(3)$ are sampled from the same set of PMU channels at three different time periods. Each matrix contains 8 PMU curves within 2 consecutive time instants. $M(1)$ contains 6 high-quality PMU curves and 2 low-quality PMU curves obtained under normal operating condition. $M(2)$ and $M(3)$ contain 8 high-quality PMU curves obtained under fault-on and normal operating conditions, respectively. The Euclidean distance is used as the similarity metric (distance function), and each PMU curve in the three matrices is projected to the 2D Euclidean space shown in Figure 2.2. The x and y coordinates of each point are the data values at the first and second time instant of the corresponding PMU curve, respectively.

The following observations can be drawn from Figure 2.2: (1) The cluster of fault-on PMU data (fault-on cluster) lies far from the clusters of high-quality PMU data under normal operating condition (normal-condition cluster), indicating weak temporal similarity between the two clusters; (2) all the points within the fault-on cluster lie close to each other, indicating strong spatial similarities among points within the fault-on cluster; and (3) the two points representing low-quality PMU curves lie far from the normal-condition cluster, as well as the majority of points in the low-quality cluster, indicating weak spatial and temporal similarities with their neighboring points. Therefore, the low-quality data points can be defined as spatio-temporal outliers under this formulation.

2.3 Online Detection of Low-Quality PMU Data

Based on the previous discussion, we propose a density-based local outlier factor (LOF) analysis to detect low-quality PMU data. In [18], similar LOF-based techniques are introduced for the detection of high sensing noises and false data injections in PMU data. This section improves the similarity metrics for PMU curves, which lead to more robust performance on detecting various types of data quality problems, including not only

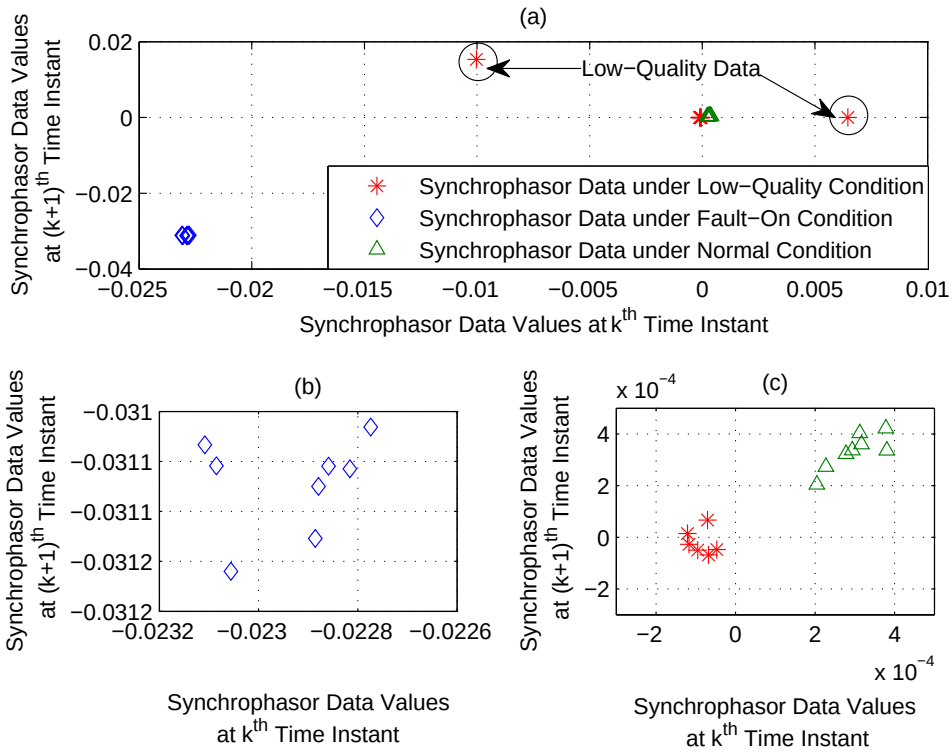

Figure 2.2: 2D points representing PMU curves under normal/fault-on/low-quality conditions. (a) Overall figure with all the 2D points under normal/fault-on/low-quality conditions. (b) Zoomed-in figure with all the 2D points under fault-on condition. (c) Zoomed-in figure with all the 2D points under normal condition and high-quality 2D points under low-quality condition [1].

sensing noises and false data injections, but also data spikes and un-updated data problems.

2.3.1 Similarity Metrics Between Synchrophasor Curves

In this subsection, two similarity metrics are proposed for detecting low-quality PMU data whose variance is significantly higher or lower than its spatio-temporal neighborhoods.

Definition 2. *Let $M(k)$ denote the PMU measurement matrix obtained at the k^{th} time period. The length of each time period equals to the length of the moving data window of the proposed algorithm. Let $M_i(k)$ and $M_j(k)$ denote the i^{th} and j^{th} columns of $M(k)$.*

Let $\sigma_i(k)$ denote the standard deviation of $M_i(k)$, Let C denote the data set of all the PMU measurements identified to be clean (without data quality problems) by the proposed algorithm. The normalized standard deviation for PMU data obtained from the i^{th} channel at the k^{th} time period is defined as follows:

$$\sigma_i^{Norm}(k) = \frac{\sigma_i(k)}{\frac{\sum_{t=1}^{t=k-1} \sigma_i(t)\chi_C(M_i(t))}{\sum_{t=1}^{t=k-1} \chi_C(M_i(t))}} \tag{2.2}$$

where

$$\chi_C(M_i(t)) = \begin{cases} 1 & (M_i(t) \in C) \\ 0 & (M_i(t) \notin C) \end{cases} \tag{2.3}$$

The normalized deviation $\sigma_i^{Norm}(k)$ represents the standard deviation of data curve obtained from the i^{th} PMU channel at the k^{th} time period, normalized by the average standard deviation of the historical clean measurements obtained from the same PMU channel. Considering $\sigma_i(k)$ as a indicator of the strength of system dynamic response recorded by i^{th} PMU channel at the k^{th} time period, $\sigma_i^{Norm}(k)$ is a normalized indicator which compares the current strength of system dynamic response with the average historical strength recorded by the same sensing channel. This normalization process removes the influence of PMU physical locations on the dynamic strength of the PMU curves.

2.3.1.1 Similarity Metric for Low-Quality PMU Data with High Variance

The similarity metric (distance function) $f_H(i,j)$ between $M_i(k)$ and $M_j(k)$ is defined as follows:

$$f_H(i,j) = \left|\sigma_i^{Norm} - \sigma_j^{Norm}\right| \tag{2.4}$$

2.3.1.2 *Similarity Metric for Low-Quality PMU Data with Low Variance*

The similarity metric (distance function) $f_L(i,j)$ between $M_i(k)$ and $M_j(k)$ is defined as follows:

$$f_L(i,j) = max\left(\left|\frac{\sigma_i^{Norm}}{\sigma_j^{Norm}}\right|, \left|\frac{\sigma_j^{Norm}}{\sigma_i^{Norm}}\right|\right) \tag{2.5}$$

The above two similarity metrics measure the difference between dynamic strength of data curves $M_i(k)$ and $M_j(k)$. Since during the same time period k, clean PMU curves across the system tend to have similar dynamic strength (similarly low/high strength under normal/fault-on operating condition), $f_H(i,j)$ and $f_L(i,j)$ values tend to be small for clean measurements. However, the dynamic strength of low-quality PMU curves tend to be different from that of the clean curves, since dynamics of low-quality PMU curves are mainly driven by the dynamics of the data quality problems, rather than the true system dynamics. Therefore, $f_H(i,j)$ and $f_L(i,j)$ values tend to be large for low-quality PMU measurements.

Although both similarity metrics could reflect the outlier behavior of both low-quality data with high variance (such as sensing noises, data spikes, etc.) and low variance (such as un-updated data), $f_H(i,j)$ tends to be more sensitive to high-variance data problems and $f_L(i,j)$ tends to be more sensitive to low-variance data problems. Under normal operating conditions, the performance of $f_H(i,j)$ in detecting low-variance data problems (such as un-updated data) could be unsatisfactory. This is because under normal operating conditions, the normalized standard deviations for clean measurements tend to be close to one, while the normalized standard deviations for low-variance data (such as un-updated data) tend to be close to zero. Therefore, under normal operating conditions, $f_H(i,j)$ between clean data and un-updated data would remain close to one, while $f_L(i,j)$ between clean data and un-updated data would be a very large number. However, under normal oper-

ating conditions, $f_H(i,j)$ between two clean data sets would be a small positive number (close to zero), and $f_L(i,j)$ between two clean data sets would lie around one. Therefore, the $f_L(i,j)$ value between un-updated data and clean data tends to be much larger than $f_L(i,j)$ value between two clean data sets, leading to a better detection performance. This performance difference is further demonstrated through case studies.

2.3.2 Density-Based Outlier Detections for PMU Data

Built upon the above similarity metrics, LOF analysis, which is a density-based outlier detection technique, is applied to solve the low-quality data detection problem. In this subsection, procedures for calculating LOFs are briefly discussed. The mathematical definition of "density" is presented below. Details of LOF analysis can be found in [19].

2.3.2.1 Calculation of K-Distance(P)

Let the measurement matrix M be a database consisting of synchrophasor measurements. Let p, q, o be some objects in M, each object represents a column in M. Let k be a positive integer. The distance between p and q, denoted by $d(p,q)$, is defined by $f_H(p,q)$ or $f_L(p,q)$.

For any positive integer k, the $k-distance$ of object p, denoted by $k-distance(p)$, is defined as the distance $d(p,o)$ between p and an object $o \in M$ such that:

a) for at least k objects $o' \in M \backslash \{p\}$ it holds that $d(p,o') \leq d(p,0)$, and

b) for at most $k-1$ objects $o' \in M \backslash \{p\}$ it holds that $d(p,o') < d(p,0)$.

In the above definition, $o' \in M \backslash \{p\}$ denotes $\{o' : o' \in M, o' \notin \{p\}\}$

Intuitively, $k-distance(p)$ represents the distance between object p and the k^{th} nearest neighbor of p. The value of $k-distance(p)$ provides a measure on the density around the object p. For the same number of k, smaller $k-distance(p)$ indicates higher density around p.

2.3.2.2 *Identification of K-Distance Neighborhood of P*

Given $k-distance(p)$, the ***k-distance neighborhood of p*** contains every object whose distance from p is not greater than the $k-distance$. This concept is defined in (2.6).

$$N_{k-distance(p)}(p) = \{q \in M \backslash \{p\} | d(p,q) \leq k-distance(p)\} \tag{2.6}$$

These objects q are called the ***k-nearest neighbors of p***.

2.3.2.3 *Calculation of Reachability Distance of Object P from Object O*

The ***reachability*** [†] ***distance of object p with respect to object o*** is defined in (2.7).

$$reach-dist_k(p,o) = max\{k-distance(o), d(p,o)\} \tag{2.7}$$

Intuitively, if object p is far away from object o, then the reachability distance between p and o is simply their actual distance $d(p,o)$. However, if they are "sufficiently" close to each other, the actual distance $d(p,o)$ is replaced by the $k-distance(o)$. The reason is that in doing so, the statistical fluctuations of $d(p,o)$ for all the p's close to o can be significantly reduced. The strength of this smoothing effect can be controlled by the parameter k. The higher the value of k, the more similar the reachability distances for objects within the same neighborhood.

Figure 2.3 illustrates the relationship among true distance $d(p_3,o)$, $k-distance(o)$, $reach-dist_k(p_1,o)$, and $reach-dist_k(p_5,o)$. In this example, $k = 3$, and true distance $d(\cdot)$ is the Euclidean distance[‡]. According to the above definitions, $k-distance(o)$ represents the distance between object o and the k^{th} nearest neighbor of o. Therefore,

[†]It should be noted that the notion of reachability in this section does not refer to reachability concept in hybrid system literature.

[‡]It should be noted that Euclidean distance is used here only for the illustration of the concepts of $k-distance(\cdot)$ and $reach-dist_k(\cdot)$.In the proposed low-quality data detection algorithm, the true distance $d(\cdot)$ is defined by similarity metrics $f_H(\cdot)$ and $f_L(\cdot)$

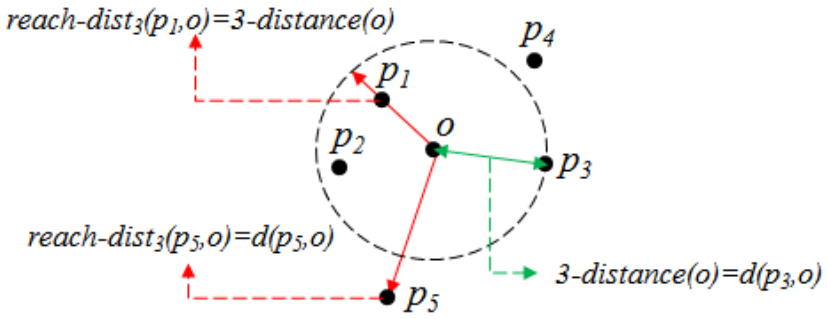

Figure 2.3: $k-distance(o)$, $reach-dist_k(p_1, o)$, and $reach-dist_k(p_5, o)$ when $k = 3$ [1].

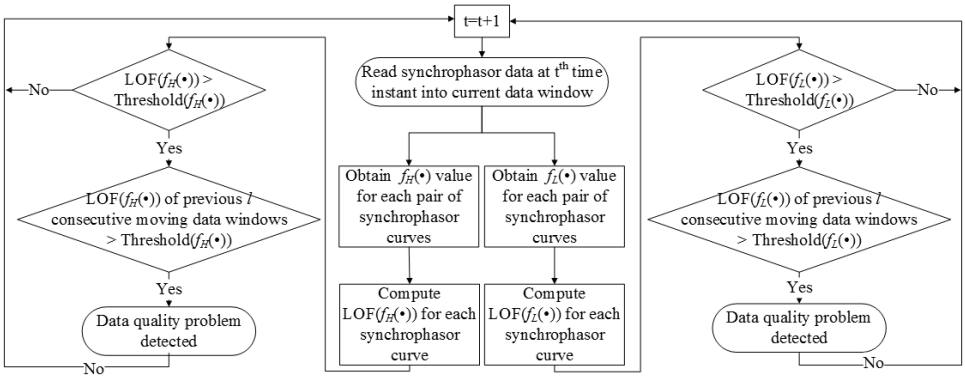

Figure 2.4: Overall flowchart of the proposed approach [1].

when $k = 3$, $k-distance(o) = d(p_3, o)$, where p_3 is the third nearest neighbor of o. The radius of the circle in Figure 2.3 represents $k-distance(o)$. Since true distance $d(p_1, o) < k-distance(o)$, and true distance $d(p_5, o) > k-distance(o)$, the reachability distance between p_1 and o is $reach-dist_k(p_1, o) = k-distance(o)$, while the reachability distance between p_5 and o is $reach-dist_k(p_1, o) = d(p_5, o)$. These reachability distances $reach-dist_k(\cdot)$, developed through the comparison between the true distances $d(\cdot)$ and the $k-distance(o)$, will then be used to formulate the local outlier factor.

2.3.2.4 Calculation of Local Reachability Density of P

The *local reachability density of p* is defined as

$$lrd_{MinPts}(p) = 1/(\frac{\sum_{o \in N_{MinPts}(p)} reach{-}dist_{MinPts}(p,o)}{|N_{MinPts}(p)|}) \tag{2.8}$$

where $N_{MinPts}(p) = N_{MinPts-distance(p)}(p)$, and $MinPts$ is a positive integer.

Intuitively, the local reachability density of an object p is the inverse of the average reachability distance based on the $MinPts$-nerest neighbors of p.

2.3.2.5 Calculation of LOF of P

The *local outlier factor of p* is defined as

$$LOF_{MinPts}(p) = \frac{\sum_{o \in N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{|N_{MinPts}(p)|} \tag{2.9}$$

The local outlier factor of object p captures the degree to which p is an local outlier. It is the average of the ratio of the local reachability density of p and those of p's $MinPts$-nearest neighbors. It is easy to see that the lower p's local reachability density is, and the higher the local reachability densities of p's $MinPts$-nearest neighbors are, the higher the LOF value of p is.

2.3.3 Robust Detection Criterion and Parameter Selections

In order to improve the robustness of the proposed approach, the following detection criterion and parameter selection procedure are applied to the algorithm.

2.3.3.1 Robust Detection Criterion

Due to the propagation delay of electro-magnetic waves, PMUs installed at different locations of a large-scale power system may respond to physical disturbances at the time instants slightly asynchronous with each other. If a short moving data window is chosen

for the algorithm, this slight time shift may cause false alarms under fault-on operating conditions. In order to avoid the false alarms without introducing too much computational burden, PMU measurements within the current moving data window are identified to contain low-quality data only if there are already l consecutive moving data windows prior to this current window, whose LOF values exceed the threshold value. l is a integer slightly less than the length of the moving data window. This criterion would introduce a small detection delay to the proposed algorithm. However, since the length of the moving data window is set to be short for the purpose of online application, the delay would be a insignificant value.

2.3.3.2 Parameter Selections

Three parameters need to be determined for the proposed algorithm: number of nearest neighbors ($MinPts$) of each object, length of the moving data window, and LOF thresholds for various similarity metrics. These parameters can be determined through off-line training using historical data. In order to reduce the detection delay, the length of moving data window should remain short. The $MinPts$ value can be selected to be around half of the total number of PMU channels, by assuming the total number of low-quality curves at each time window should be less than the total number of high-quality PMU curves.

According to the previous discussions, the overall flowchart of the proposed algorithm is shown in Figure 2.4. Key steps for implementing this low-quality data detection approach are as follows.

***Step 1*:** Create the current moving data window by reading in PMU measurements at the latest time instant.

***Step 2*:** Compute $f_H(\cdot)$ and $f_L(\cdot)$ values for each pair of PMU curves.

***Step 3*:** Compute LOF value of each PMU curve, based on $f_H(\cdot)$ and $f_L(\cdot)$. For each

PMU curve, the LOF value can be calculated following the equations in the previous subsection.

***Step 4*:** If the LOF value corresponding to $f_H(\cdot)$ or $f_L(\cdot)$ of the i^{th} PMU curve exceeds the threshold, go to Step 5; otherwise, go to Step 7.

***Step 5*:** If the previous l consecutive LOF values corresponding to $f_H(\cdot)$ or $f_H(\cdot)$ of the i^{th} PMU curve exceed the threshold, go to Step 6; otherwise, go to Step 7.

***Step 6*:** The i^{th} PMU curve is detected to contain low-quality data at current time window.

***Step 7*:** Move the data window to the next time instant, and go back to Step 1.

Although the above calculation procedure involves looping process for the LOF calculation of each PMU curve, there is no time-consuming computation (such as matrix inversion, decomposition, etc.) involved in the above procedure. All the operations within the looping process request light computational efforts. The computational burden of the entire process is not significant. The computational performance of the proposed algorithm is demonstrated through the case studies.

2.4 Case Studies

The proposed approach is tested using both synthetic and real-world PMU data. Low-quality measurements caused by various reasons are used to verify the effectiveness of the approach. In all the following test cases, a unique set of algorithm parameters are used: moving data window length = 20 data points; LOF threshold corresponding to $f_H(\cdot)$ = 10; LOF threshold corresponding to $f_L(\cdot)$ = 100; Number of neighboring data for LOF algorithm = 0.5 $\times$ number of PMU curves. In order to demonstrate the proposed method is capable to detect low-quality data under fault-on operating conditions, a system physical disturbance (fault) is recorded by the PMU data in each test case.

2.4.1 Case Study with Synthetic Data

The synthetic PMU measurements are sampled from the simulation results of a standard IEEE-14 test system, with a sampling rate of 50Hz. A three-phase line-to-ground fault is presented while running the simulation. In each test case, one type of low-quality data is randomly inserted into a subset of the test data.

2.4.1.1 Synthetic Data with High Sensing Noise

This test data set contains 14 synthetic voltage magnitude measurement curves, where 3 of them (No. 1, 5, 14) contain Gaussian noises lasting from 6s to 6.4s, with a signal-to-noise ratio (SNR) of 40dB. Figure 2.5 shows the 3 curves with data quality problems.

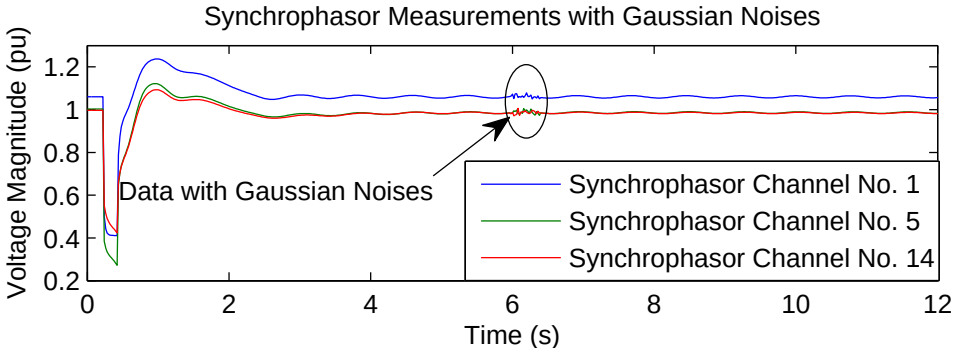

Figure 2.5: Synthetic PMU measurements with high sensing noise [1].

Table I presents the detection results. It shows that all the 3 noisy data segments are successfully detected, without introducing any false alarm by the physical disturbance. A small detection delay (less than 0.38s) is introduced, due to the length of the moving data window. The average computing time for each moving data window is *0.0161s*. Figure 2.6 presents the LOF values of all the PMU curves, when data quality problem or physical disturbance is presented. This comparison shows that the LOF values exceed the thresh-

old when low-quality data is presented, while remain below the threshold when physical disturbance is presented. The results indicate the proposed method is able to detect low-quality PMU data while avoiding false alarms caused by system physical disturbances.

Table I: Detection Results for Synthetic PMU Data with High Sensing Noise [1]

Index of PMU with High Noise	Starting Time of Noisy Segment	Ending Time of Noisy Segment
1	6.22s (LOF = 620.5)	6.78s (LOF = 31.9)
5	6.34s (LOF = 429.1)	6.78s (LOF = 73.3)
14	6.34s (LOF = 418.6)	6.76s (LOF = 48.2)

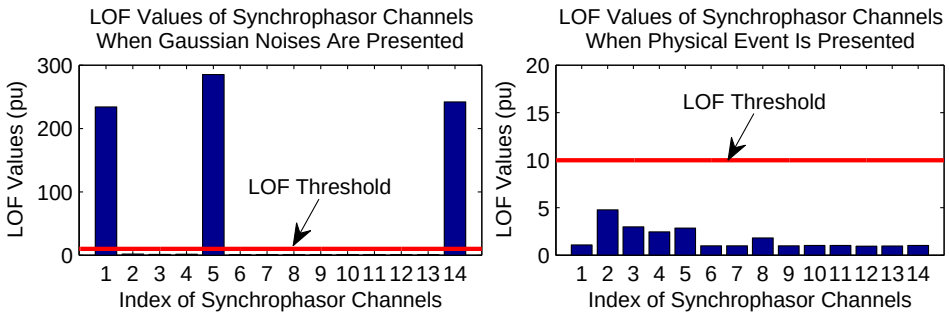

Figure 2.6: LOF values of synthetic PMU channels when physical disturbance (right) or high sensing noise (left) is presented [1].

2.4.1.2 Synthetic Data with Spikes

This test data set contains 47 synthetic real power measurement curves, where 4 of them (No. 3, 6, 30, 45) contain data spikes lasting from 6.3s to 6.4s. These spikes can be caused by problems such as data loss or time skew of GPS clock [14]. Figure 2.7 shows the 4 curves with data quality problems.

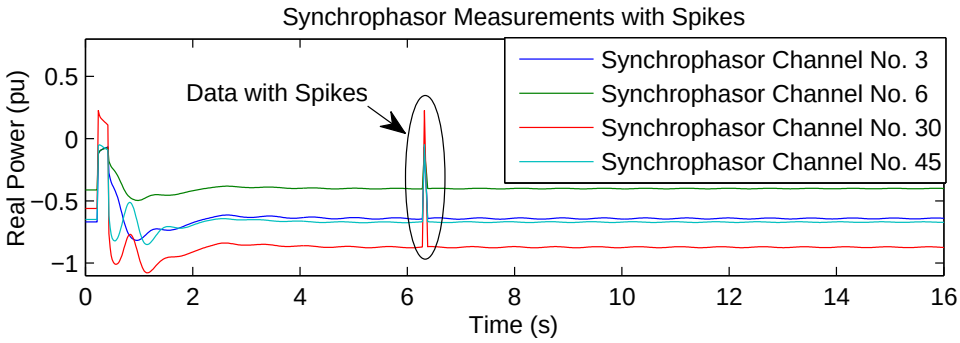

Figure 2.7: Synthetic PMU measurements with data spikes [1].

The detection results are shown in Table II. All the 4 spikes are detected and no false alarm is introduced by physical disturbance. The detection delay introduced by the length of the moving data window is less than 0.36s. The average computing time for each moving data window is *0.0627s*. Figure 2.8 presents the LOF values of all the PMU curves, when data quality problem or physical disturbance is presented. It is clear that low-quality data would cause the LOF values to exceed the threshold, while system physical disturbances would not cause a significant increment in LOF values.

Table II: Detection Results for Synthetic PMU Data with Spikes [1]

Index of PMU with Data Spike	Starting Time of Spike Segment	Ending Time of Spike Segment
3	6.46s (LOF = 107.7)	6.76s (LOF = 85.2)
6	6.46s (LOF = 113.9)	6.76s (LOF = 90.4)
30	6.48s (LOF = 102.3)	6.76s (LOF = 71.5)
45	6.44s (LOF = 270.3)	6.76s (LOF = 58.7)

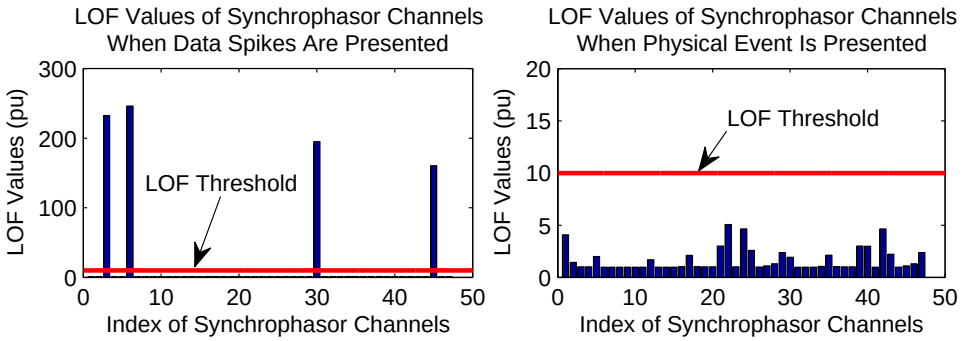

Figure 2.8: LOF values of synthetic PMU channels when physical disturbance (right) or data spike (left) is presented [1].

2.4.1.3 Synthetic Data with Un-Updated Data

This test data set contains 14 synthetic voltage magnitude measurement curves, where 3 of them (No. 6, 12, 13) contain un-updated data lasting from 6s to 6.4s. Figure 2.9 shows the 3 curves with data quality problems.

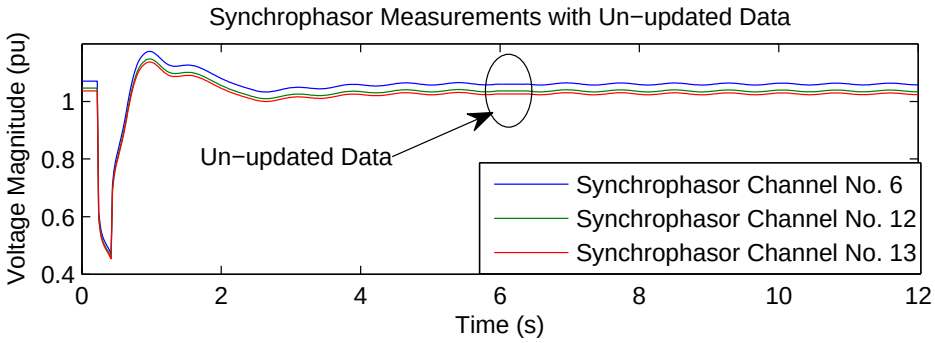

Figure 2.9: Synthetic PMU measurements with un-updated data [1].

Table III presents the detection results. The 3 un-updated data segments are detected, while the presence of physical disturbance does not cause any false alarm. The detection delay introduced by the length of the moving data window is less than 0.36s, and the

average computation time for each moving time window is *0.0128s*.

Table III: Detection Results for Synthetic PMU Data with Un-updated Data [1]

Index of PMU with Un-updated Data	Starting Time of Un-updated Segment	Ending Time of Un-updated Segment
6	6.36s (LOF = 3423.6)	6.40s (LOF = 3519.5)
12	6.36s (LOF = 3423.6)	6.40s (LOF = 3519.5)
13	6.36s (LOF = 3423.6)	6.40s (LOF = 3519.5)

2.4.1.4 Synthetic Data with False Data Injection

This test data set contains 47 synthetic real power measurement curves, where 4 of them (No. 15, 21, 29, 42) contain false data injections lasting from 6s to 6.4s. Figure 2.10 shows the 4 curves with data quality problems.

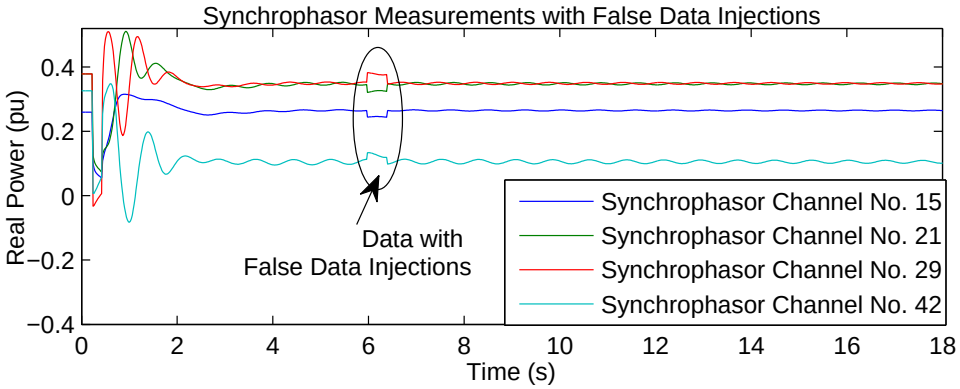

Figure 2.10: Synthetic PMU measurements with false data injection [1].

The detection results are shown in Table IV. Although physical disturbance is presented, all the 4 false data injections are correctly detected and no false alarm is intro-

duced. The detection dely caused by the length of the moving data window is less than 0.38s. The average computing time for each moving data window is *0.0627s*.

In all the above case studies using synthetic PMU measurements, the maximum detection delay is less than 0.4s, and the maximum computing time for each moving data window is less than 0.1s. It is summarized in [20] that the data latency requirements for online quasi-steady-state applications (state estimation, small signal stability analysis, oscillation analysis, voltage stability analysis, etc.) range from 1s to 5s. It is clear that both the detection delay and the computing time of the proposed method satisfy the latency requirements for PMU-based online quasi-steady-state applications. Therefore, the proposed method is suitable for online dtection of low-quality PMU measurements, in order to improve the accuracy of these PMU-based applications.

Table IV: Detection Results for Synthetic PMU Data with False Data Injections [1]

Index of PMU with False Data Injection	Starting Time of Injected Data Segment	Ending Time of Injected Data Segment
15	6.32s (LOF = 39.7)	6.78s (LOF = 30.9)
21	6.32s (LOF = 25.7)	6.78s (LOF = 19.9)
29	6.32s (LOF = 14.1)	6.78s (LOF = 10.7)
42	6.34s (LOF = 10.8)	6.72s (LOF = 10.9)

2.4.2 Case Study with Real-World Data

High-quality PMU measurements obtained from a real-world power grid are used to test the proposed approach. The sampling rate of the data is 100Hz. A line-tripping fault is recorded by the data. In each test case, one type of low-quality data is manually inserted to a randomly-chosen subset of the test data, so that the ground truth of the existence of

low-quality data is known for sure.

2.4.2.1 Real-World Data with High Sensing Noise

This test data set contains 39 real-world voltage magnitude measurement curves, where 4 of them (No. 10, 15, 23, 29) contain Gaussian noises lasting from 1s to 1.2s, with a SNR of 40dB. The SNR of the original clean data set is tested to be well below 40dB. Figure 2.11 shows the 4 curves with data quality problems.

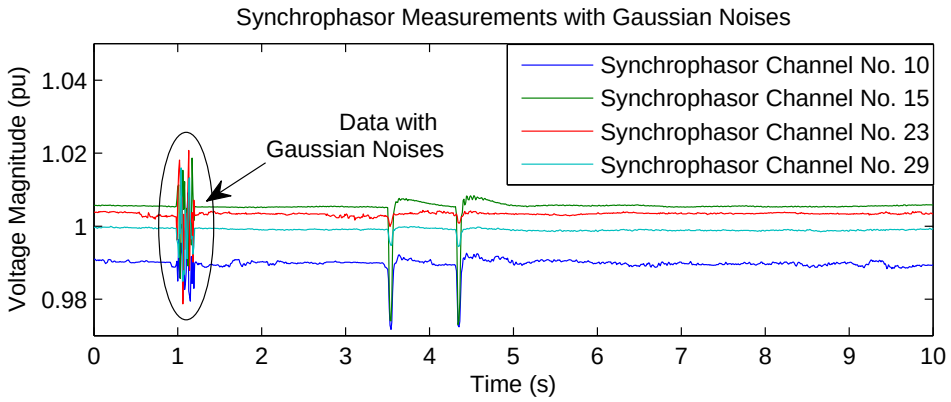

Figure 2.11: Real-world PMU measurements with high sensing noise [1].

Table V presents the detection results. It shows that all the 4 noisy data segments are successfully detected, without introducing any false alarm by the physical disturbance. A small detection delay (less than 0.19s) is introduced, due to the length of the moving data window. The average computing time for each moving data window is *0.0376s*. Figure 2.12 presents the LOF values of all the PMU curves, when data quality problem or physical disturbance is presented. This comparison shows that the LOF valus exceed the threshold when low-quality data is presented, while remain below the threshold when physical disturbance is presented. The results indicate the proposed method is able to detect low-quality PMU data while avoiding false alarms caused by physical disturbances.

Table V: Detection Results for Real-World PMU Data with High Sensing Noise [1]

Index of PMU with High Noise	Starting Time of Noisy Segment	Ending Time of Noisy Segment
10	1.17s (LOF = 286.9)	1.39s (LOF = 30.0)
15	1.16s (LOF = 577.8)	1.39s (LOF = 43.3)
23	1.16s (LOF = 206.3)	1.39s (LOF = 12.3)
29	1.16s (LOF = 328.2)	1.39s (LOF = 35.1)

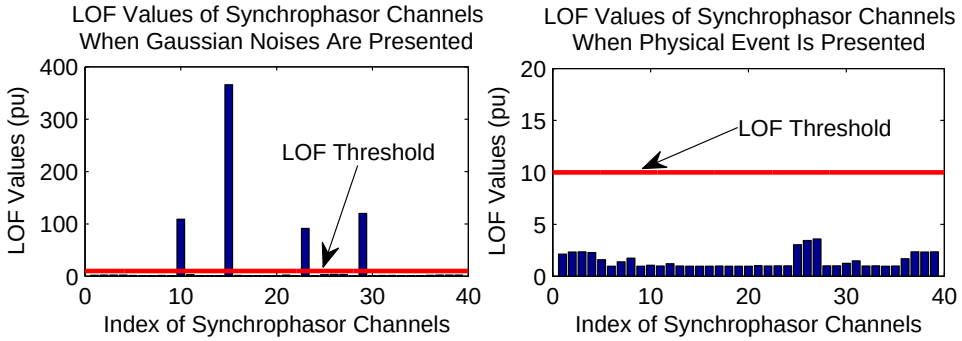

Figure 2.12: LOF values of real-world PMU channels when physical disturbance (right) or high sensing noise (left) is presented [1].

2.4.2.2 Real-World Data with Spikes

This test data set contains 22 real-world real power measurement curves, where 4 of them (No. 3, 6, 20, 21) contain data spikes at the time instant of 1.06s. In this test case, the length of each data spike is one sample. This test scenario is created in order to test the performance of the algorithm in detecting single data dropout. Figure 2.13 shows the 4 curves with data quality problems.

The detection results are shown in Table VI. All the 4 spikes are detected and no false alarm is introduced by physical disturbance. The detection delay introduced by the length of the moving data window is less than 0.19s. The average computing time for

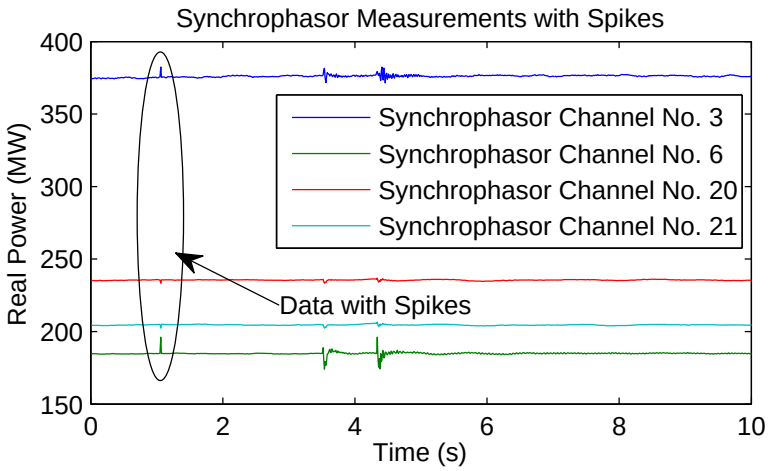

Figure 2.13: Real-world PMU measurements with data spikes [1].

each moving data window is *0.0150s*. Figure 2.14 presents the LOF values of all the PMU curves, when data quality problem or physical disturbance is presented. It is clear that low-quality data would cause the LOF values to exceed the threshold, while system physical disturbances would not cause a significant increment in LOF.

Table VI: Detection Results for Real-World PMU Data with Spikes [1]

Index of PMU with Data Spike	Starting Time of Spike Segment	Ending Time of Spike Segment
3	1.22s (LOF = 52.0)	1.25s (LOF = 28.2)
6	1.22s (LOF = 124.8)	1.25s (LOF = 69.2)
20	1.22s (LOF = 50.5)	1.25s (LOF = 27.2)
21	1.22s (LOF = 71.7)	1.25s (LOF = 39.5)

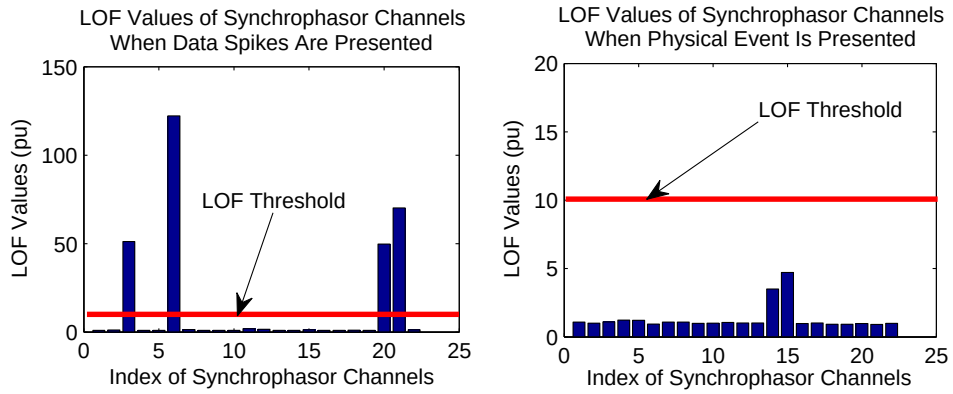

Figure 2.14: LOF values of real-world PMU channels when physical disturbance (right) or data spike (left) is presented [1].

2.4.2.3 Real-World Data with Un-Updated Data

This test data set contains 13 real-world current magnitude measurement curves, where 4 of them (No. 1, 5, 7, 13) contain un-updated data lasting from 1s to 1.2s. Figure 2.15 shows the 4 curves with data quality problems.

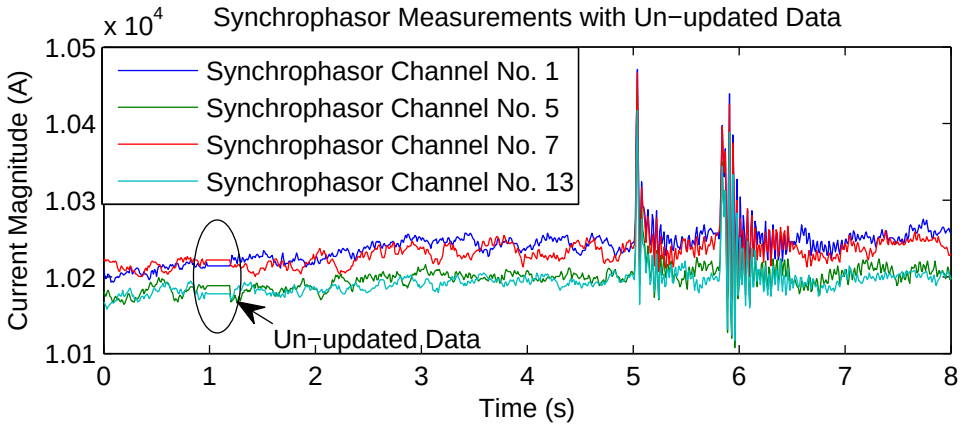

Figure 2.15: Real-world PMU measurements with un-updated data [1].

Table VII presents the detection results. The 4 un-updated data segments are detected,

while the presence of physical disturbance does not cause any false alarm. The detection delay introduced by the length of the moving data window is less than 0.18s, and the average computation time for each moving data window is *0.0115s*.

Table VII: Detection Results for Real-World PMU Data with Un-updated Data [1]

Index of PMU with Un-updated Data	Starting Time of Un-updated Segment	Ending Time of Un-updated Segment
1	1.18s (LOF = 4637.2)	1.20s (LOF = 4537.2)
5	1.18s (LOF = 4637.2)	1.20s (LOF = 4537.2)
7	1.17s (LOF = 3317.8)	1.20s (LOF = 4537.2)
13	1.18s (LOF = 4637.2)	1.20s (LOF = 4537.2)

Figure 2.16 presents the current magnitude data obtained from PMU channels No. 1 and No. 2, where PMU channel No. 1 contains un-updated data from 1s to 1.2s, and PMU channel No. 2 contains clean data only. Figure 2.17 presents the normalized deviations of the two PMU channels, as the computation data window moves with time. It is clear that: 1) under normal operating conditions, the normalized deviations of clean data segments lie close to one; 2) under fault-on operating conditions, the normalized deviations of clean data segments increase significantly; 3) the normalized deviations of un-updated data segments decrease towards zero.

Figure 2.18 presents the $f_H(i,j)$ and $f_L(i,j)$ values of PMU channels No. 1 and No. 2, as the computation data window moves with time. Figure 2.19 presents the LOF values of PMU channels No. 1 and No. 2, as the computation data window moves with time. It is clear from Figure 2.18 and Figure 2.19 that $f_L(i,j)$ is more sensitive to the un-updated data than $f_H(i,j)$, and therefore leads to a better detection performance for un-updated data.

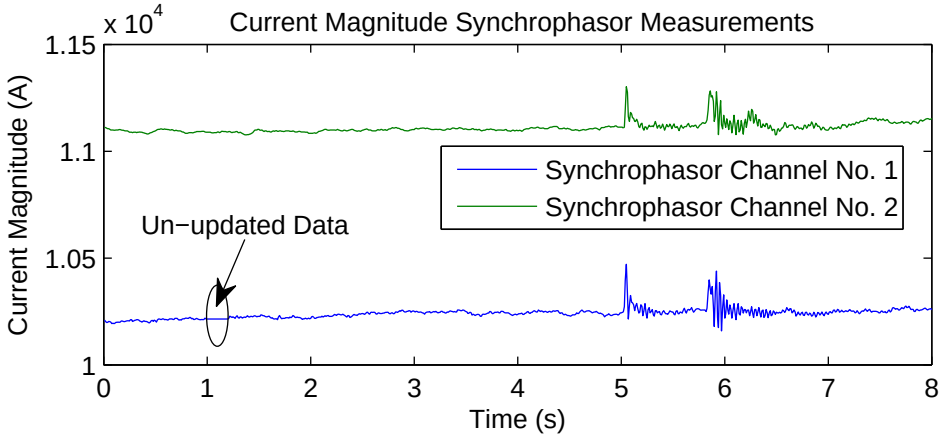

Figure 2.16: Real-world current magnitude PMU measurements [1].

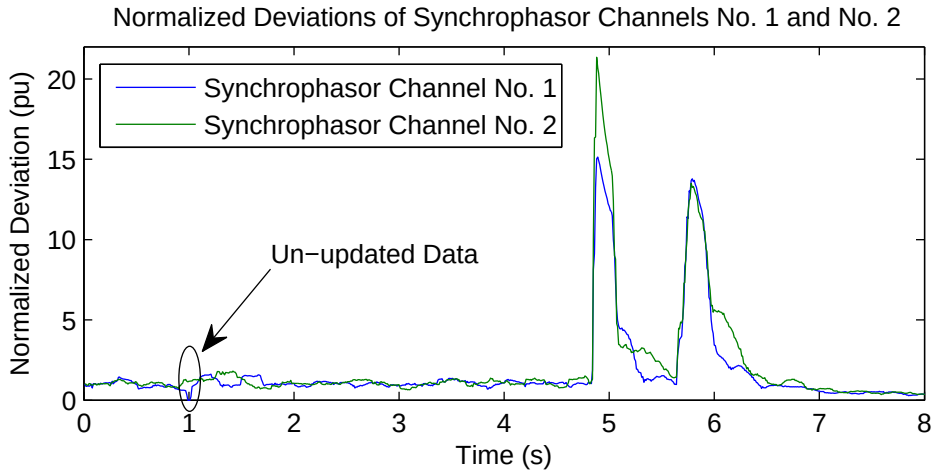

Figure 2.17: Normalized deviation of PMU channel No. 1 and No. 2 [1].

2.4.2.4 Real-World Data with False Data Injection

This test data set contains 39 real-world voltage magnitude measurement curves, where 4 of them (No. 2, 20, 27, 37) contain false data injections lasting from 1s to 1.2s. Figure 2.20 shows the 4 curves with data quality problems.

The detection results are shown in Table VIII. Although physical disturbance is pre-

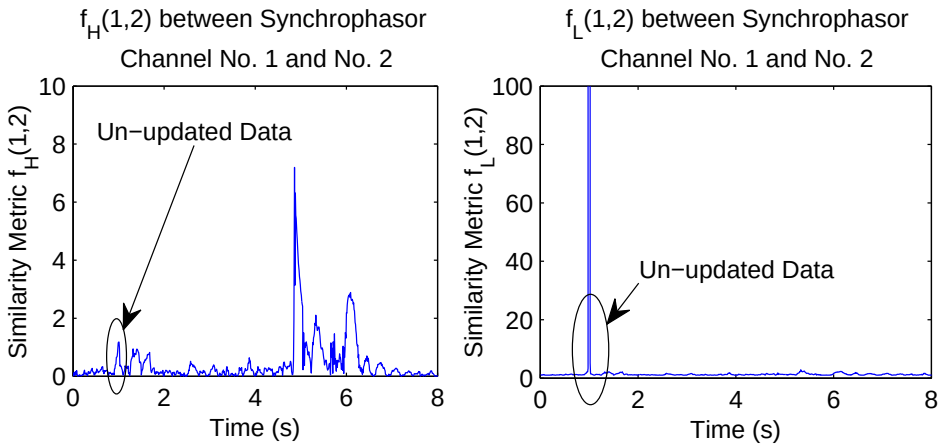

Figure 2.18: Similarity metric $f_H(i,j)$ (left) or $f_L(i,j)$ (right) between PMU channels No. 1 and No. 2 [1].

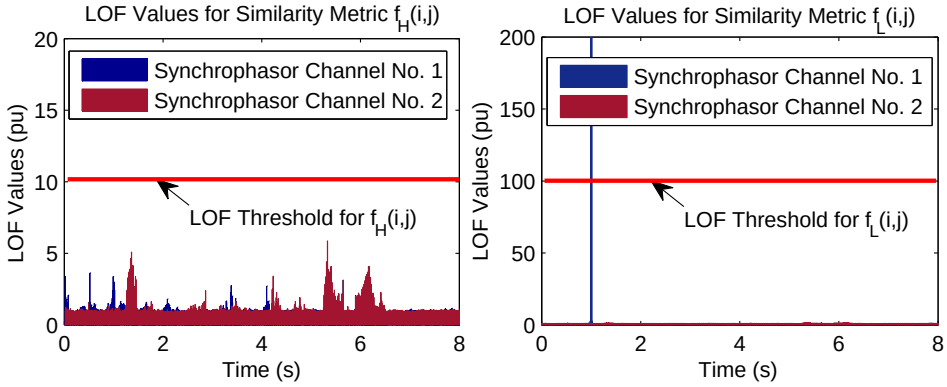

Figure 2.19: LOF values when similarity metric $f_H(i,j)$ (left) or $f_L(i,j)$ (right) is applied [1].

sented, all the 4 false data injections are correctly detected and no false alarm is introduced. The detection delay caused by the length of the moving data window is less than 0.19s. The average computing time for each moving data window is *0.0475s*.

In all the above case studies using real-world PMU measurements, the maximum detection delay is less than 0.2s, and the maximum computing time for each moving data window is less than 0.05s. It is summarized in [20] that the data latency requirements for

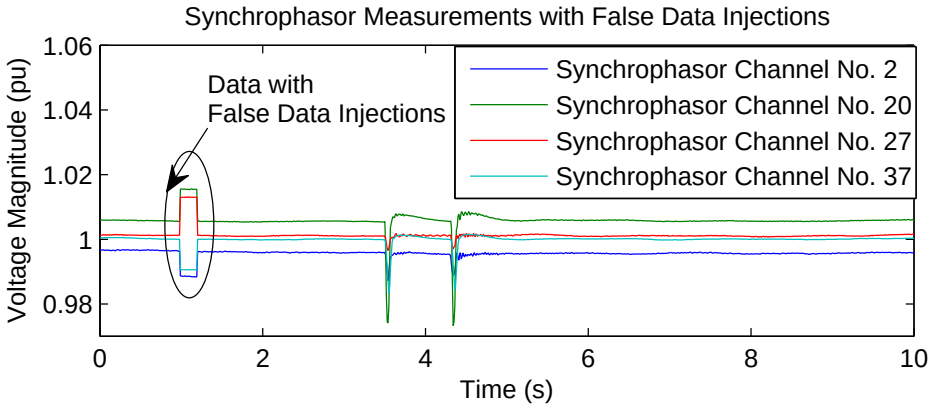

Figure 2.20: Real-world PMU measurements with false data injection [1].

online quasi-steady-state applications (such as state estimation, small signal stability analysis, oscillation analysis, voltage stability analysis, etc.) range from 1s to 5s. It is clear that both the detection delay and the computing time of the proposed method satisfy the latency requirements for PMU-based online quasi-steady-state applications. Therefore, the proposed method is suitable for online dtection of low-quality PMU measurements, in order to improve the accuracy of these PMU-based applications.

Since the detection delay of the proposed algorithm is mainly caused by the length of the moving data window, the delay could be estimated and removed when the occurrence time of the low-quality data is reported. By doing this, the reported occurrence time of the low-quality data could be very close to its actual occurrence time.

For power grids with a large number of PMUs, the computation speed of the proposed algorithm could be further improved by applying the detection algorithm in a decentralized framework. In large systems, multiple detection engines could be applied to process PMU measurements obtained from different physical locations or control areas (such as different states or different local control centers). PMUs lying far from each other could be grouped into different subgroups, and be processed in parallel by different detection engines. This

decentralized framework could help reduce the number of PMU channels that need to be processed by each detection engine, and therefore improve the computation speed of each detection engine. Since this method does not require any system-wide information (such as system topology), it can be easily decentralized without spending extra effort on creating the reduced or equivalent system model.

Meanwhile, parallel processing could also help improve the online computation performance of the proposed algorithm. Multiple processors could be applied at each detection engine, so that several consecutive moving data windows could be processed by different processors at the same time. This parallel technique could improve the overall computation speed when the proposed algorithm is applied to power systems with a significant number of PMUs.

Table VIII: Detection Results for Real-World PMU Data with False Data Injections [1]

Index of PMU with False Data Injection	Starting Time of Injected Data Segment	Ending Time of Injected Data Segment
2	1.16s (LOF = 74.3)	1.39s (LOF = 71.2)
20	1.16s (LOF = 117.7)	1.39s (LOF = 111.1)
27	1.16s (LOF = 95.5)	1.39s (LOF = 91.6)
37	1.16s (LOF = 383.4)	1.39s (LOF = 365.7)

2.5 Section Conclusion

This section presents a framework that is possible for online detection and improvement of PMU data quality issues. The proposed approach formulates the low-quality PMU data as spatio-temporal outliers among all the PMU measurements, and performs detection through a density-based local outlier detection algorithm. Similarity metrics are proposed

to quantify the spatio-temporal similarities among multi-time-instant PMU measurements. The proposed approach has satisfactory performance under both normal and fault-on operating conditions. It requires no prior information on system modeling and topology. The computation speed of the proposed algorithm is suitable for online applications. Synthetic and real-world PMU measurements are used to verify the effectiveness of the proposed approach. This framework, if successful, could potentially boost up system operators' confidence of PMU-based analytics in modern power systems.

Built upon this work, future research could focus on developing similarity metrics with more sensitive and robust performance, identifying root causes of the low-quality problems, and correcting the low-quality PMU data.

3. PHYSICS-BASED SITUATIONAL AWARENESS FOR VOLTAGE STATBILITY ASSESSMENT

3.1 Motivation

Power systems with deep renewable penetration are more vulnerable to voltage instabilities, due to the increased generation uncertainties and decreased reactive power support of the renewable power plants [21]. Fundamentally speaking, the voltage instability problem is induced by the underlying saddle-node bifurcation happening in the power system [22–24]. The degree of voltage instability of a certain power system can be measured through its voltage stability margin. Specifically, the power system voltage stability margin measures the distance between the closest bifurcation point and the current operating condition. All the possible bifurcation points in the system form the hyper-surface of the power flow solution boundary.

There is a large literature studying the voltage stability margins and the power flow solution boundaries. In [25, 26], continuation methods are proposed to trace the closest bifurcation point along a pre-specified direction of load increment. In [25], the pre-specified loading direction is allowed to have one degree of freedom, while in [26], the pre-specified loading direction may have two degrees of freedom, making it possible to visualize the power flow solution boundary in a two-dimensional space. In [27], an upper bound is provided for the voltage stability margin along a pre-specified loading direction, through the second-order cone programming. In [28, 29], the bifurcation point that is locally closest to the current operating condition is obtained through iterative method, direct method, and optimization-based method.

In order to further explore the power flow solution boundary and the voltage stability margin in high-dimensional loading space, we propose two approaches to visualize

the power flow solution boundary in high-dimensional loading space, and to calculate the global lower bound of the voltage stability margin. Compared with the existing techniques for obtaining the voltage stability margin, the proposed approaches merit the following advantages: 1) the proposed approaches do not assume any specified direction of load increment, which reduces the impact of load uncertainty to the voltage stability assessment; 2) the proposed approaches study the global properties of the power flow solution boundary and the voltage stability margin around a certain operating condition, instead of searching for the locally closest bifurcation points.

The rest of this section is organized as follows: Section 3.2 describes the proposed visualization tool for the power flow solution boundary in high-dimensional space; Section 3.3 presents the proposed method for calculating the lower bound of the voltage stability margin; Section 3.4 provides concluding remarks for this section.

3.2 Visualization of The Power Flow Solution Boundary

3.2.1 Introduction to Power Flow Solution Boundary

The nonlinear power flow problem (in polar coordinating system) can be expressed using the following set of equations:

$$f(x, \lambda) = 0 \tag{3.1}$$

where $x \in R^n$ represents the bus voltage magnitudes and phase angles, $\lambda \in R^p$ represents the bus real/reactive power generations/loads.

For a point lying on the power flow solution boundary, (3.1) must be satisfied, i.e., it must be a power flow solution. However, since all the points lying on the power flow solution boundary should be bifurcation points of the corresponding dynamic system, it must also satisfy the following equation:

$$detf_x(x,\lambda) = 0 \tag{3.2}$$

where $f_x(x,\lambda)$ represents the Jacobian matrix of the power flow equations, evaluated at the current operating condition (x,λ), and det represents the matrix determinant.

3.2.2 Formulation of The Visualization Problem

The objective of the visualization problem can be stated as follows:

For a given operating point (x^0,λ^0), find a collection of operating points (x^*,λ^*), such that:

1) Each point (x^*,λ^*) lies on the power flow solution boundary (i.e., satisfies (3.1) and (3.2)).

2) The collection of points (x^*,λ^*) describes the geology of the power flow solution boundary around (x^0,λ^0).

In order to achieve the above objective, the visualization problem is formulated as the following optimization problem, and a dynamic-programming-based approach is proposed to solve the optimization problem:

$$\begin{aligned} &min||\lambda^0-\lambda^*||_2 \\ &s.t. f(x^0,\lambda^0)=0 \\ &\quad f(x^*,\lambda^*)=0 \\ &\quad detf_x(x^*,\lambda^*)=0 \end{aligned} \tag{3.3}$$

The above optimization framework searches for the power flow solution boundary point (x^*,λ^*), which has the shortest Euclidean distance towards the given operating point (x^0,λ^0). A dynamic-programming-based approach is proposed to solve the above optimization problem. After running the dynamic programming algorithm, a collection of points (x^*,λ^*) on the power flow solution boundary can be obtained, and the geology of

the power flow solution boundary around (x^0, λ^0) can be characterized.

3.2.3 Proposed Solution for The Visualization Problem

The dynamic-programming-based algorithm for solving the visualization problem is described using the following pseudo codes:

Algorithm 1: Visualization of Power Flow Solution Boundary

---

1 Initialization: determine a small step size τ for the change of parameters λ;

2 Obtain initial power flow solution (x_0, λ_0) by solving $f(x_0, \lambda_0) = 0$;

3 Starting from (x_0, λ_0), obtain all the possible next moving steps by the following iterations:

4 **for** *k=1:1:length(λ_0)* **do**

5 $\quad \lambda_1^{(k)} = \lambda_0 + [0, \cdots, \tau, \cdots, 0]^T$ (τ is the k^{th} element);

6 $\quad f(x_1^{(k)}, \lambda_1^{(k)}) = 0$ (f is the power flow equations) ;

7 **end**

8 Solve the following recursion problem using dynamic programming - memorization technique:

9 $G(x_0, \lambda_0) =$

$1 + MIN(G(x_1^{(1)}, \lambda_1^{(1)}), \cdots, G(x_1^{(k)}, \lambda_1^{(k)}), \cdots, G(x_1^{(length(\lambda_0))}, \lambda_1^{(length(\lambda_0))}))$;

---

where $G(\cdot)$ represents the optimization problem formulated in (3.3).

After running the above dynamic-programming-based algorithm, a collection of points lying on the power flow solution boundary around (x_0, λ_0) can be obtained, and the geology of the power flow solution boundary around (x_0, λ_0) can be characterized accordingly.

3.3 Calculation of Lower Bound of The Voltage Stability Margin

One of the key disadvantages of the above visualization algorithm is that, the dynamic-programming-based algorithm performs a greedy search for the power flow solution bound-

ary around a certain operating condition, and therefore is computationally expensive and difficult to be applied to large-scale practical power systems.

In order to overcome the above disadvantage, another algorithm is proposed to calculate the global lower bound of the voltage stability margin. The formulation of this algorithm is presented in the following section.

3.3.1 Problem Formulation for Calculating The Lower Bound of The Voltage Stability Margin

The following optimization problem is formulated for the calculation of the lower bound of the voltage stability margin:

$$\begin{aligned} & min||\lambda^0 - \lambda^*||_2^2 \\ & s.t. f^r(x^0, \lambda^0) = 0 \\ & \quad f^r(x^*, \lambda^*) = 0 \\ & \quad v^T f_x^r(x^*, \lambda^*) = 0 \\ & \quad v^T v = 1 \\ & \quad x^{min} \leq x^* \leq x^{max} \\ & \quad \lambda^{min} \leq \lambda^* \leq \lambda^{max} \end{aligned} \tag{3.4}$$

where $f^r(\cdot)$ denotes the power flow equations in rectangular coordinating system; $f_x^r(\cdot)$ denotes the Jacobian matrix of the power flow equations in rectangular coordinating system; v denotes the left eigenvector of the power flow Jacobian matrix $f_x(x^*, \lambda^*)$; x^{max} and x^{min} denote the upper and lower bounds for x^*, respectively; and λ^{max} and λ^{min} denote the upper and lower bounds for λ^*, respectively.

The key advantage of (3.4) is that, by applying power flow equations in rectangular coordinating system, and introducing the left eigenvector of the power flow Jacobian matrix, the above optimization problem can be formulated as a polynomial-constrained polyno-

mial optimization problem. The lower bound of the minimal value of the above problem can be obtained through a series of sum-of-squares (SOS) relaxations. The proposed algorithm for obtaining the SOS-based lower bound of the original problem is presented in the following section.

3.3.2 Proposed Solution for Calculating The Lower Bound of The Voltage Stability Margin

The following algorithm is proposed for obtaining the lower bound of the polynomial-constrained polynomial optimization problem described in (3.4):

Algorithm 2: Calculating Lower Bound of The Voltage Stability Margin

1 Initialization: Formulate (3.4) into the following polynomial-constrained polynomial optimization form:
$\zeta^* = min. \;\; f_0(x) \;\; s.t. \;\; f_k(x) \geq 0 \;\; (k = 1, \cdots, m)$;

2 Obtain the following generalized Lagrangian function:
$L(x, \mu_1, \cdots, \mu_m) = f_0(x) - \Sigma_{k=1}^{m} \mu_k(x) f_k(x), \quad for \;\; \forall x \in R^n, \forall \mu_k \in SOS_*$;

3 Obtain the following generalized Lagrangian dual:
$max \quad \eta \quad s.t. \quad L(x, \mu_1, \cdots, \mu_m) - \eta \geq 0 \quad (\forall x \in R^n), \quad \mu_1 \in SOS_*, \cdots, \mu_m \in SOS_*$;

4 Obtain sum-of-squares (SOS) relaxation of order r:
$\eta^r = max \quad \eta \quad s.t. \quad L(x, \mu_1, \cdots, \mu_m) - \eta \in SOS_r \quad (\forall x \in R^n), \quad \mu_1 \in SOS_{r1}, \cdots, \mu_m \in SOS_{rm}$;

5 Solve the above SOS relaxation using SOS optimization engine (via semi-definite programming), and obtain the lower bound η^r ;

6 Check the tightness of η^r. If needed, update r using $r + 1$ and re-run the algorithm ;

where SOS_{rk} denotes the set of sum of square polynomials with degree $\leq r_k$; r denotes the relaxation order of the SOS relaxation; $r_k = r - \lceil degree(f_k)/2 \rceil \;\; (k = 1, \cdots, m)$ is

chosen to balance the degrees of all the terms $\mu_k(x)f_k(x) \ \ (k = 1, \cdots, m)$.

The details of the above sum-of-squares relaxation can be found in [30–32]. It is proved in [30, 31], that under a moderate assumption which requires the feasible region of the original polynomial optimization problem to be compact, the global minimum of the original polynomial optimization problem can be approximated as closely as desired by solving a finite sequence of SOS optimization problems through semi-definite programming (i.e., $\eta^r \leq \eta^{r+1}$, $\eta^r \to \zeta^*$ as $r \to \infty$).

3.4 Case Studies

The proposed visualization algorithm is tested using a three-bus test system. Figure 3.1 shows the one-line diagram of the three-bus system. In the test system, Bus 1 is modeled as the slack bus, Bus 2 is modeled as the PQ bus, and Bus 3 is modeled as the PV bus. When visualizing the power flow solution boundary in high dimensional space, the following parameters remain unchanged: the terminal voltage magnitude of Bus 3 (the PV bus), as well as the terminal voltage magnitude and voltage phase angle of Bus 1 (the slack bus). The following three variables can be changed independently in order to obtain the power flow solution boundary in the high-dimensional space: the real power load variation of Bus 2 (PQ bus), the real power generation variation of Bus 3 (PV bus), and the reactive power load variation of Bus 2 (PQ bus).

Figure 3.2 shows the visualization of the power flow solution boundary (around the current operating condition) in the three-dimensional space, where the three axes denote the real power load variation of Bus 2 (PQ bus), the real power generation variation of Bus 3 (PV bus), and the reactive power load variation of Bus 2 (PQ bus), respectively. It can be seen from Fig. Y, that there are multiple local minimal points existed in the non-convex three-dimensional surface.

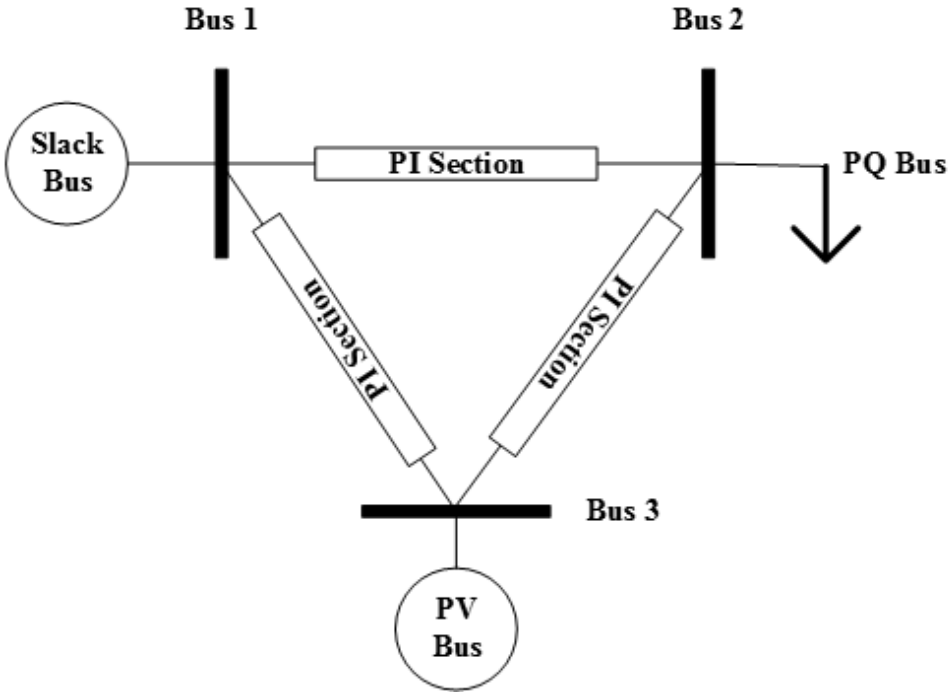

Figure 3.1: One-line diagram of the three-bus test system.

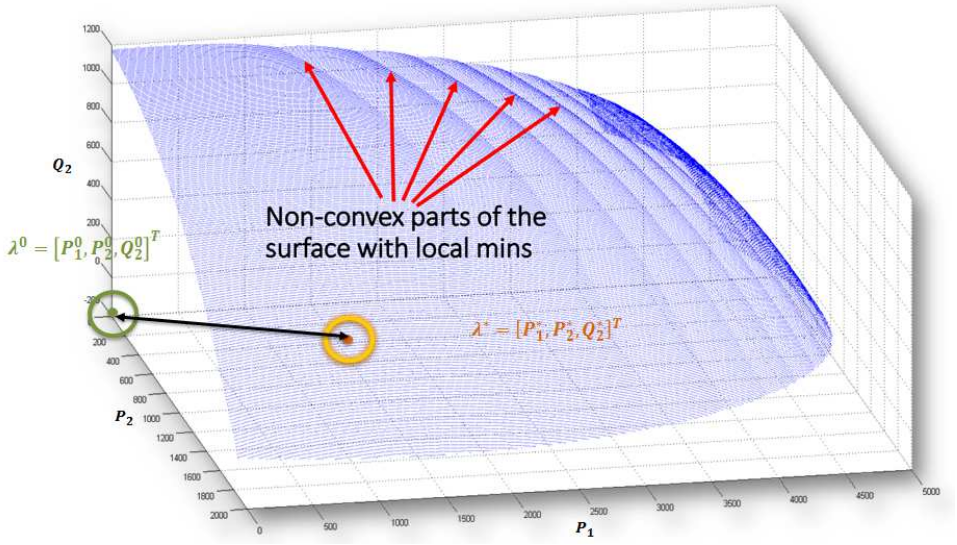

Figure 3.2: Three-dimensional power flow solution boundary around the current operating condition.

3.5 Section Conclusion

In this section, we propose two approaches to visualize the power flow solution boundary in high-dimensional loading space, and to calculate the global lower bound of the voltage stability margin. Compared with the existing techniques for obtaining the voltage stability margin, the proposed approaches merit the following advantages: 1) the proposed approaches do not assume any specified direction of load increment, which reduces the impact of load uncertainty to the voltage stability assessment; 2) the proposed approaches study the global properties of the power flow solution boundary and the voltage stability margin around a certain operating condition, instead of searching for the locally closest bifurcation points.

In the future work, we would explore the power flow solution boundaries of various test systems, and further improve the proposed approaches according to the verification results.

4. COMBINED PHYSICS AND DATA DECISION SUPPORT FOR POWER PLANT MODEL VALIDATION*

4.1 Motivation

The phasor measurement unit (PMU) based power plant model validation (PPMV) has been widely adopted in North America for identifying and correcting inaccurate power plant dynamic models. In the North America Electric Reliability Corporation (NERC) guidelines and standards [33–36], the generator owners (GOs) are suggested to perform PMU-based PPMV to determine interim power plant model parameters in supplement of their long-term plans for the offline tests, and the planning coordinators (PCs) are required to implement a documented process to perform model validations using PMU disturbance recordings, which includes not only the determination of an unacceptable model, but also the resolution to the underlying model problems. These guidelines and standards urge the PCs and GOs to establish a systematic and standardized process for applying PMU-based PPMV to all the eligible power plants within their systems. Given the huge number of power plants available in North America power grids, it would be a challenging and time-consuming task for PCs and GOs to perform the PMU-based PPMV on a regular basis, if convenient software tools are not available.

4.1.1 The Concept of PMU-Based PPMV

In order to use PMU-recorded disturbance data to validate the dynamic model of a specific power plant, a PMU must be installed at the point of interconnection (POI) between the power plant and the rest of the system. The dynamics of the rest of the system is then

*This section is in part a reprint with permission from Meng Wu, Weihong Huang, Frankie Qiang Zhang, Xiaochuan Luo, Slava Maslennikov, and Eugene Litvinov of the material in the paper: "Power Plant Model Verification at ISO New England", at *2017 IEEE PES General Meeting*, Chicago, IL [2]. Copyright 2017 IEEE.

represented using the PMU measurements obtained at the POI. Detailed dynamic models of the system outside the studied power plant are ignored, and the entire system (except for the studied power plant) is equivalenced as an infinite bus (the POI bus).

After obtaining the PMU voltage magnitude data $|V|^{PMU}$ and voltage phase angle data θ^{PMU} at the POI, these measurements are injected into the POI bus, in order to represent the system dynamics outside the studied power plant. A transient simulation study is then performed to the reduced system with the dynamic models of the studied power plant and the PMU injection curves. The simulated real and reactive power curves at the POI, P^{Simu} and Q^{Simu}, are then compared with the corresponding PMU curves at the POI, P^{PMU} and Q^{PMU}. If significant mismatch is found between the PMU curves and the corresponding simulation curves, the studied power plant model is considered inaccurate, and further model diagnosis and calibration need to be performed.

In this section, a PPMV case refers to the set of data, $\{P^{Simu}, Q^{Simu}, P^{PMU}, Q^{PMU}\}$, generated by running the above simulation procedure once for the studied power plant model. A mismatched PPMV case refers to the PPMV case with significant mismatch between the PMU curves and the corresponding simulation curves. A mismatched PPMV case indicates the underlying model problems for the studied power plant model. When multiple sets of $\{P^{PMU}, Q^{PMU}\}$ curves are obtained different transient events, various PPMV cases can be generated accordingly for the studied power plant model. If the underlying model problem of a mismatched PPMV case is already known, this case could serve as a labeled data point for the model diagnosis framework described in the following section. These labeled data points could be generated through simulation studies with incorrect power plant models, as well as practical PPMV studies with detailed engineering judgment for the model problems.

4.1.2 The Current Procedure for PMU-Based PPMV

The current procedure for PMU-Based PPMV is shown in Figure 4.1. Three types of inputs are needed for the PMU-based PPMV: the PMU data during suitable physical events, the power flow case data, and the dynamic modeling data for transient simulations. These inputs are sent through a transient simulation engine with the capability of running the PMU-based PPMV. A series of simulation curves can then be obtained. Based on the simulation results, the comparison study is performed to identify critical mismatches between the simulation curves and the corresponding PMU curves. Once critical mismatches are identified, the model calibration process is performed for the adjustment of the model parameters or structures of the power plant.

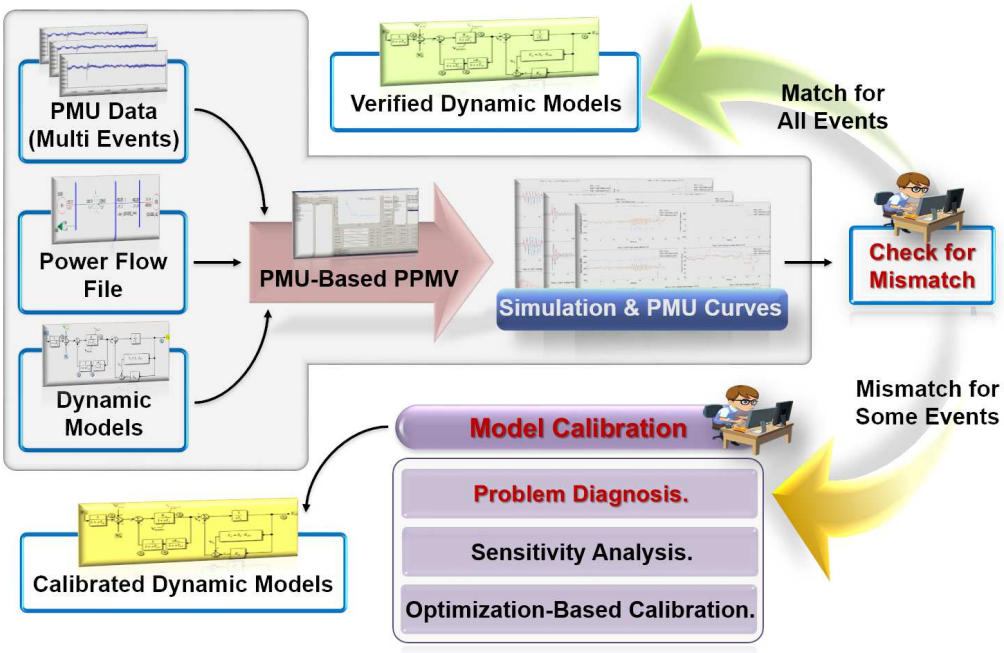

Figure 4.1: The current procedure of PMU-based PPMV.

Within the process of model calibration, three key functions need to be performed: 1) the model problem diagnosis which identifies the specific devices with modeling errors using engineering judgement; 2) the sensitivity analysis [37] which identifies sensitive parameters for the adjustment; 3) the optimization-based calibration which applies optimization techniques such as the ensemble Kalman filter [37], the variable projection method [38], and the particle swarm optimization [39] to perform model calibration automatically.

Currently, most utilities and independent system operators (ISOs) perform the PMU-based PPMV in a manual way. Although there is huge potential for automating the entire process shown in Figure 4.1, production-grade software tools for PMU-based PPMV have not been widely applied to the industrial practices. In this section, the following efforts towards an automated PMU-based PPMV are introduced: 1) a batch power plant model validation tool is designed and developed to automatically perform the model validation (without calibration) for multiple power plants using a single disturbance event; 2) a feature-based diagnosis framework is proposed as an automatic screening tool to mimic the engineering judgement process and diagnose the underlying modeling problems before detailed calibrations.

The rest of this section is organized as follows. Section 4.2 presents the main functions of the batch power plant model validation tool; Section 4.3 introduces the automatic power flow initialization process developed for the batch power plant model validation tool. Section 4.4 proposes the feature-based diagnosis framework for the PMU-based PPMV. Section 4.5 presents the case study results for the proposed PPMV diagnosis framework. Section 4.6 provides the concluding remarks for this section.

4.2 Batch Power Plant Model Validation Tool

The batch power plant model validation (BPPMV) tool is designed to automatically perform the model validation for multiple power plants using a single disturbance event. The PSAT and TSAT programs (developed by Powertech Labs) [40, 41] are executed to perform core functions of power flow computation and time-domain simulation, respectively. After reading the user inputs of selected power plants and disturbance event period (start/end time), the BPPMV tool automatically queries the commercial PMU and SCADA databases for necessary measurements, calls PSAT for power flow initialization module, and creates TSAT cases for playback simulations. Upon completing simulations, the model validation results are saved automatically into Matlab data files, Excel data sheets, Matlab figure files, and PNG figure files. The above functions can be executed either sequentially (in automatic operation mode), or separately (in manual operation mode).

4.2.1 Main User Interface of BPPMV Tool

The BPPMV tool is equipped with a graphical user interface (see Figure 4.2). All the functions implemented in the BPPMV tool can be launched from this main window. The operational status of each function can be viewed from the status bar at the bottom of the main window. When the BPPMV main window is launched for the first time, the following settings can be initialized: 1) locations of the TSAT and PSAT programs, offline model validation case files, BPPMV mapping file, and path for validation outputs; 2) PMU and SCADA data sampling rates, and PMU and SCADA database credentials; 3) TSAT simulation length, time step, and integration method; 4) Model validation input signals and output data format. Once initialized, the above settings can be saved into a default configuration file, so that users can load them the next time they open it. Figure 4.3 shows the user interface for parameter settings.

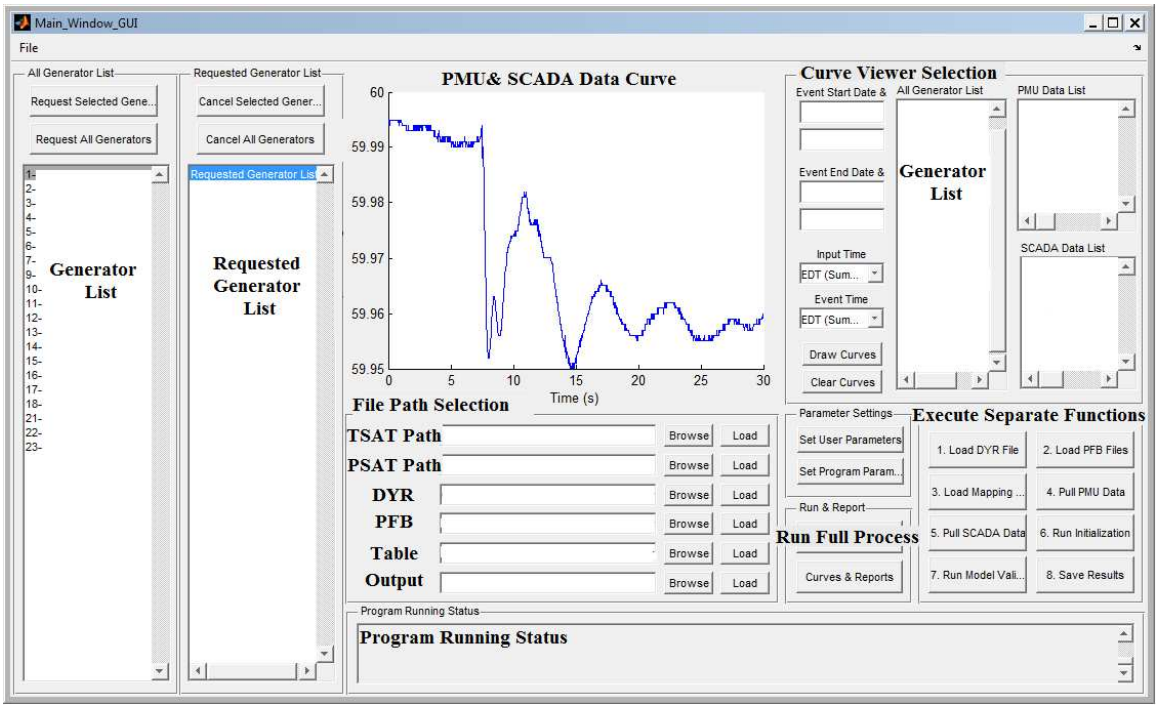

Figure 4.2: Main user interface of the BPPMV tool [2].

4.2.2 Raw PMU/SCADA Data Display

After loading the initial configurations, all the power plants eligible for model validation will be listed on the left of the main user interface (in Figure 4.2). Users can select the desired power plants and one disturbance period (start/end time) to set up a BPPMV case. The main user interface is equipped with the function of displaying raw PMU/SCADA data for user-selected power plants, so that users can visually inspect the selected disturbance and check the data quality. When raw signals with different units are selected for visualization, a separate window will pop up with all the selected signals grouped into different subplots, based on their units (see Figure 4.4).

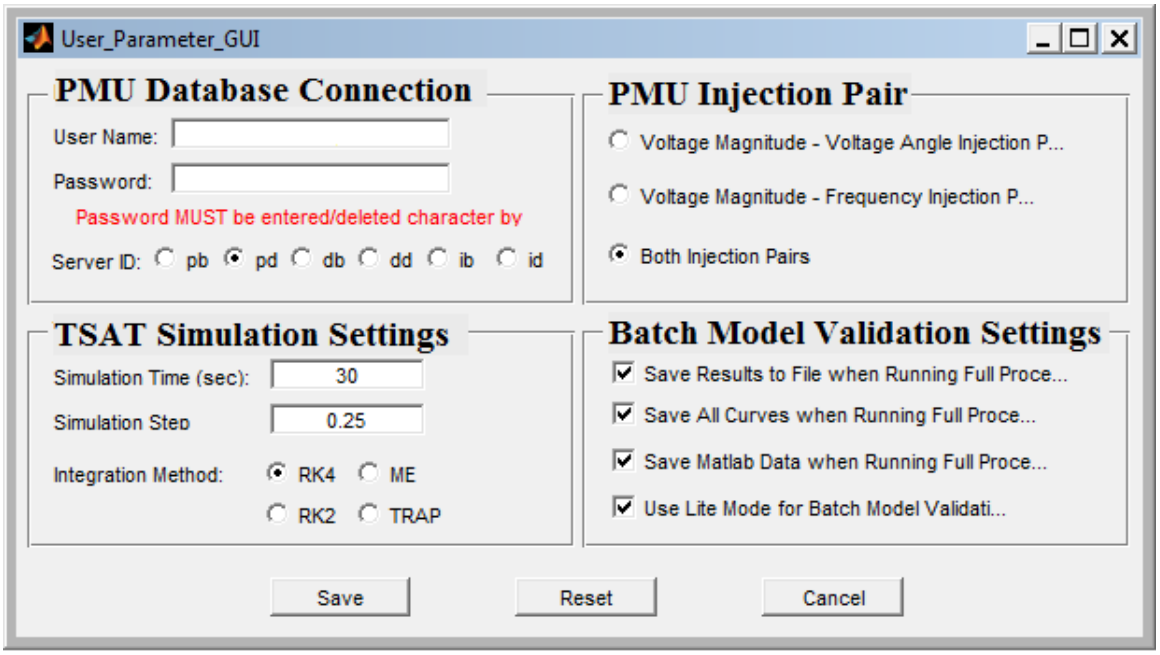

Figure 4.3: Separate user interface for parameter settings [2].

4.2.3 Model Validation Results Display

A separate user interface (see Figure 4.5) is designed for the results display, which is accessible from the main user interface. Comparisons between actual PMU curves and TSAT simulation curves can be displayed for each selected power plant. The comparison results of the following quantities can be visualized: current magnitude, current angle, real power, reactive power, voltage magnitude, voltage angle and frequency.

4.3 Power Flow Initialization Process for The BPPMV Tool

This section presents a convenient power flow initialization process, which is a necessary step to start a time-domain simulation. The challenge here is to estimate the Generator Step-up Transformer (GSU) losses, derive the actual generator output, and match the point-of-interconnection (POI) initial conditions of the model validation cases with actual PMU data. The key advantages of the proposed initialization process are: 1) it involves

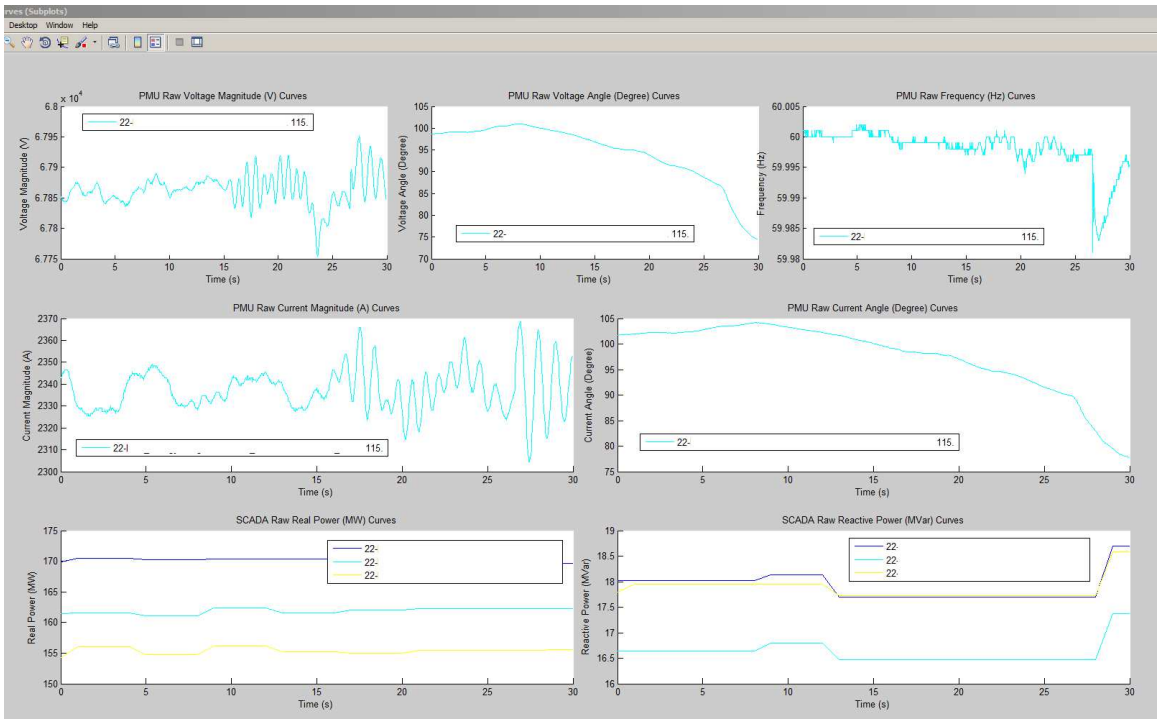

Figure 4.4: Pop-up window for raw PMU/SCADA data display [2].

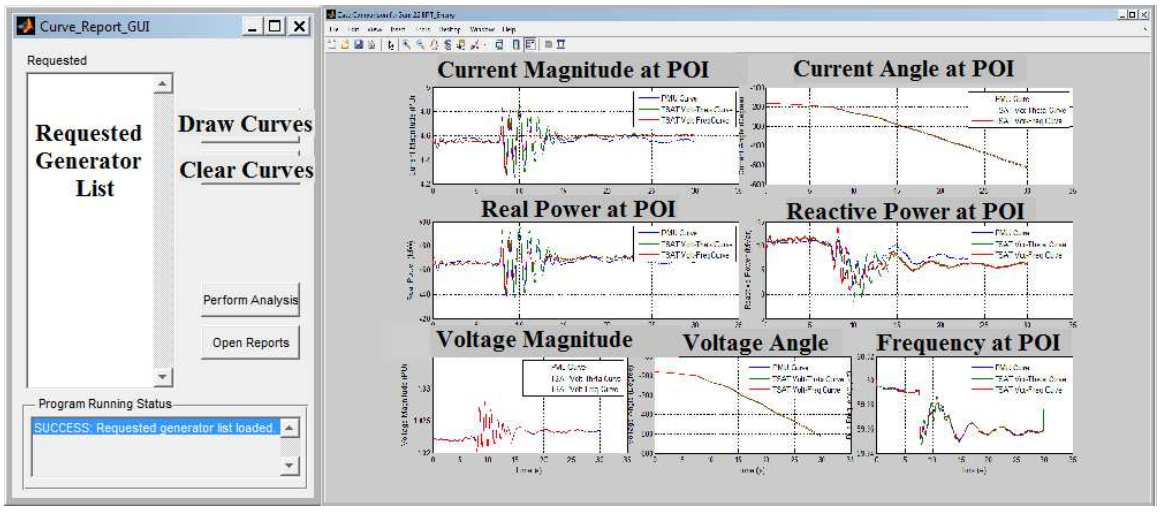

Figure 4.5: User interface for validation results display [2].

only power flow calculations provided by software vendors; 2) it does not require extra programming efforts on power flow or state estimation calculations; 3) it involves no iterative searches or heuristic trial-and-errors. Detailed initialization approaches for single generator case and multi-generator-single-POI case are described as follows.

4.3.1 Single Generator Case

The proposed initialization method for single-generator cases involves two power flow calculations and the following PMU measurements at the POI bus are needed: voltage magnitude V^{PMU}, voltage angle θ^{PMU}, real power injection P^{PMU}, and reactive power injection Q^{PMU}. The detailed steps are highlighted below:

***Step 1*:** Replace the rest of the system with an equivalent generator (PMU-Gen) model attached to the POI bus.

***Step 2*:** Model the POI bus as a PV bus, and set the following power flow parameters for the PMU-Gen: set the desired real power injection as $P^{PMU-Gen} = P^{PMU}$, and the reactive power injection as $Q^{PMU-Gen} = Q^{PMU}$. Fix its reactive output by setting its upper/lower reactive power limits to Q^{PMU}.

***Step 3*:** Model the subject generator bus as a slack bus, and set the following power flow parameters: let it regulate the voltage at the POI bus and set the desired voltage as V^{PMU}; set the desired voltage angle of the generator bus as $\theta^{Gen} = \theta^{PMU}$.

***Step 4*:** Model other buses as PQ buses.

***Step 5*:** Run the power flow calculation using the above settings and save the results.

***Step 6*:** Based on the power flow case saved in Step 5, remodel the POI bus as a slack bus and the generator bus as a PV bus; set the desired voltage angle of POI bus as $\theta^{POI} = \theta^{PMU}$.

***Step 7*:** Run the second power flow calculation using these settings and save the final results.

The main purpose of the first power flow (Step 1 to Step 5) is to match the initial power flow condition of the POI bus with the PMU measured values: V^{PMU}, P^{PMU}, and Q^{PMU}. The key idea is to model the POI bus as PV bus only without voltage regulation. Since the POI bus voltage is regulated by the subject generator, the actual generator reactive power output is solved by the first power flow calculation. The subject generator's real power output is solved because it is modeled as a slack bus. The second power flow (Step 6 and Step 7) is to match the voltage angle of the POI bus with the actual PMU value θ^{PMU}.

Compared with the popular iterative state estimation approach, the proposed method only applies two power flow calculations and is easy to implement using any commercial software.

4.3.2 Multi-generator-single-POI Case

The proposed initialization method for multi-generator-single-POI cases are described as follows. Since PMU data only has the total output (at POI), SCADA measurements of each generator's output are also needed: real power P^{SCADA}, and reactive power Q^{SCADA}. Power flow initialization is conducted with the following steps:

***Step 1*:** Replace the equivalent system with an equivalent generator (PMU-Gen) model attached to the POI bus.

***Step 2*:** Model the POI bus as a slack bus, and set the following power flow parameters for the POI bus: set the desired active power injections as $P^{PMU-Gen} = P^{PMU}$, and the reactive power as $Q^{PMU-Gen} = Q^{PMU}$; fix its reactive power output by setting its upper/lower reactive power limits to Q^{PMU}.

***Step 3*:** Model the subject generators' buses as PV buses, and set the following power flow parameters: let them regulate the voltage at the POI bus and set the desired voltage as the PMU value V^{PMU}; set their Q^{Gen} values to be their Q^{SCADA} values; set their P^{Gen} values to be their P^{SCADA} values; set the desired voltage angle of the generator buses as $\theta^{Gen} = \theta^{PMU}$.

***Step 4*:** Enable governor response in the solution parameters dialog, run the power flow calculation using these settings and save the results.

In the power flow results, all the initial power flow conditions at the POI bus are matched by the measured values: V^{PMU}, P^{PMU}, Q^{PMU}, and θ^{PMU}. Besides, the active and reactive power outputs for each non-POI generator are matched by the measured values: P^{SCADA} and Q^{SCADA}; offline units are turned off accordingly. The key idea here is to use the governor response to resolve the small active power mismatches caused by GSU losses and match the POI boundary conditions.

It is worth mentioning that even though the proposed initialization process is designed for PSAT and narrated with PSAT terminologies, the principals are generic and could be applicable to other similar commercial software as well.

4.4 Feature-Based Diagnose Framework for Power Plant Model Validation

The importance of phasor measurement unit (PMU) based power plant model validation (PPMV) is being recognized by researchers and practitioners. In North America, the PMU-based PPMV is typically performed manually by utilities and independent system operators (ISOs) for diagnosing and calibrating power plant model problems. In order to guarantee the accuracy of the calibration results, engineers need to provide manual judgment to the mismatched PPMV cases, so that the type of the modeling problem (such as wrong machine parameters, missing governor models, etc.) can be determined before the model is sent for a detailed calibration. This manual judgment process has become

the major bottleneck for automating the PMV-based PPMV. To overcome this difficulty, this section proposes a feature-based diagnosis framework to determining the types of the power plant model problems. It mimics the human engineering judgment process via a supervised learning engine. Instead of applying purely curve fitting or sensitivity analysis, this approach uses the engineering experience extracted from the labeled historical PPMV cases, establish the critical feature space, and then perform artificial learning to determine the type of a power plant model problem. The proposed framework could serve as a screening tool for the PPMV engineering judgment process, which could potentially help automate the entire PMU-based PPMV applications.

4.4.1 Formulation of The Model Diagnosis Problem

In this section, the power plant model diagnosis problem is formulated as a statistical classification problem. Figure 4.6 shows the overall flowchart of the model diagnosis problem. The model diagnosis framework contains the following three sub-problems.

4.4.1.1 Feature Extraction

To diagnose the power plant model problems automatically, critical features need to be extracted from the labeled data points (obtained from the practical or simulated PPMV cases). For the PPMV applications, features are defined to be individual measurable characteristics of a critical mismatch pattern between the simulated curve and the corresponding PMU curve. These characteristics / mismatch patterns could strongly suggest the underlying model problem of the power plant. Figure 4.7 shows several examples of critical features that can be extracted from mismatched PPMV cases. According to the engineering experience [37] and the underlying physics of the power plant model, the post-event steady state mismatch is a strong indicator of the turbine governor model problem; the inverse swing pattern during the transient period suggests the model problem with the power system stabilizer (PSS); and the time shift between the Q^{PMU} and the Q^{Simu} curves indi-

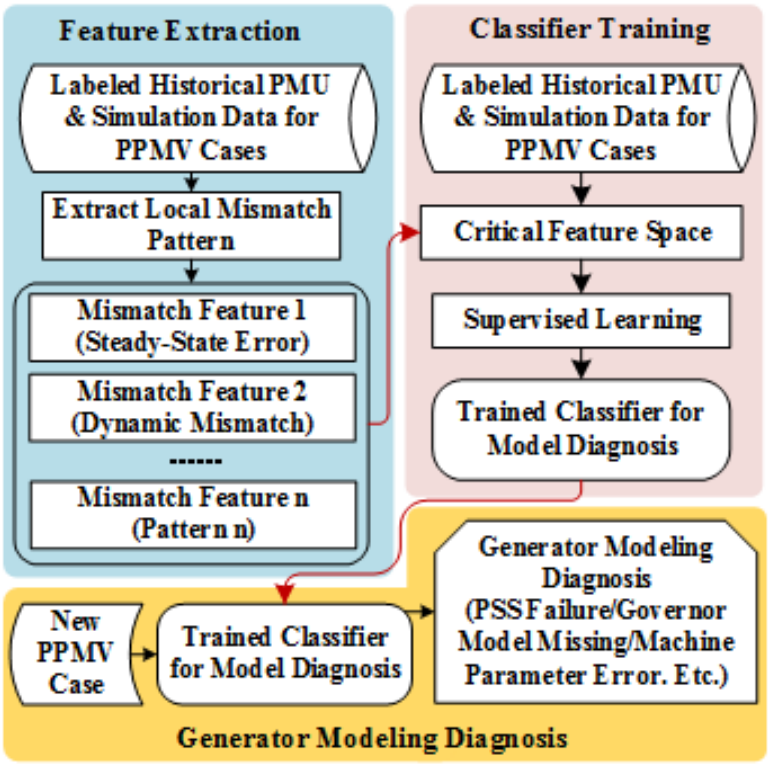

Figure 4.6: Flowchart of the model diagnosis problem.

cates model problem with the controller time constant.

In this section, we extract critical features using engineering judgment of critical mismatch patterns that would suggest unacceptable power plant models. For each critical feature extracted, a quantification metric is proposed in order to quantify the degree of mismatch between the simulated curve and the PMU curve. A high-dimensional feature space is then established, in which each of the feature has one quantification metric (distance function).

4.4.1.2 Classifier Training

Once the critical feature space is established, all the historical/simulated PPMV cases are projected onto this feature space. The degree of mismatch between the PMU curve and

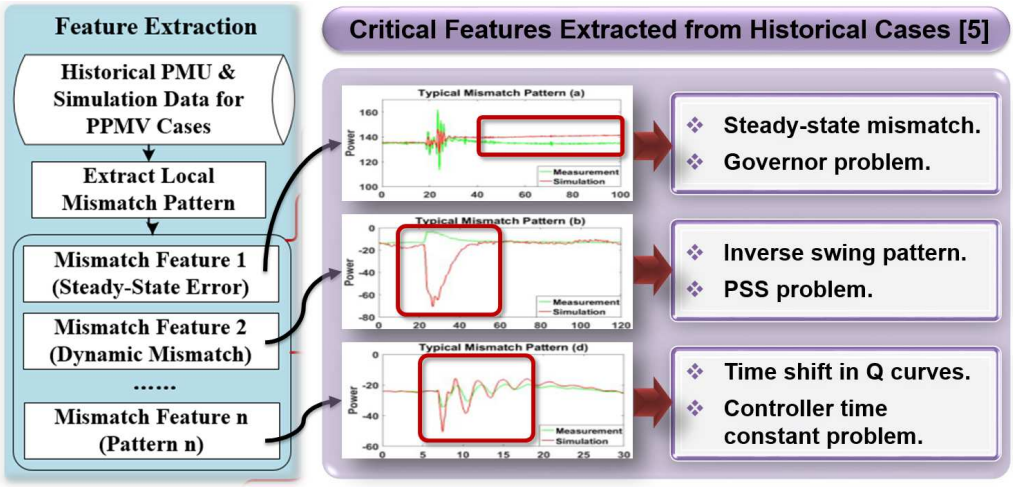

Figure 4.7: Examples of critical feature extraction from mismatched PPMV cases.

the simulated curve in each PPMV case are then quantified through the high-dimensional quantification metric (distance function). In this section, the supervised learning approach is applied to train the classifier for the model diagnosis. The practical/simulated PPMV cases with identified model problems serve as the labeled data points for the supervised learning engine. Figure 4.8 presents the flowchart for the classifier training sub-problem.

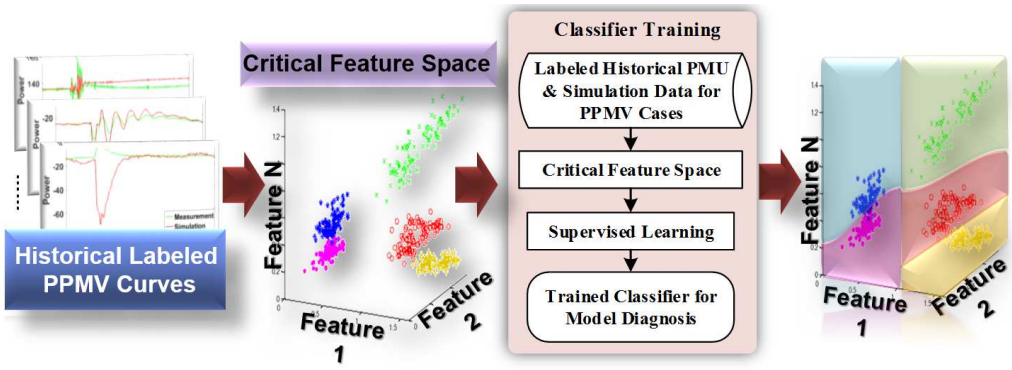

Figure 4.8: Flowchart for the classifier training sub-problem.

4.4.1.3 Power Plant Model Diagnosis

After the classifier is obtained through the supervised learning approach, it is then applied to classify the model problems of new PPMV cases into one of the known categories. The classification results suggest the underlying model problems of the new PPMV cases. Figure 4.9 presents the flowchart for the power plant model diagnosis sub-problem.

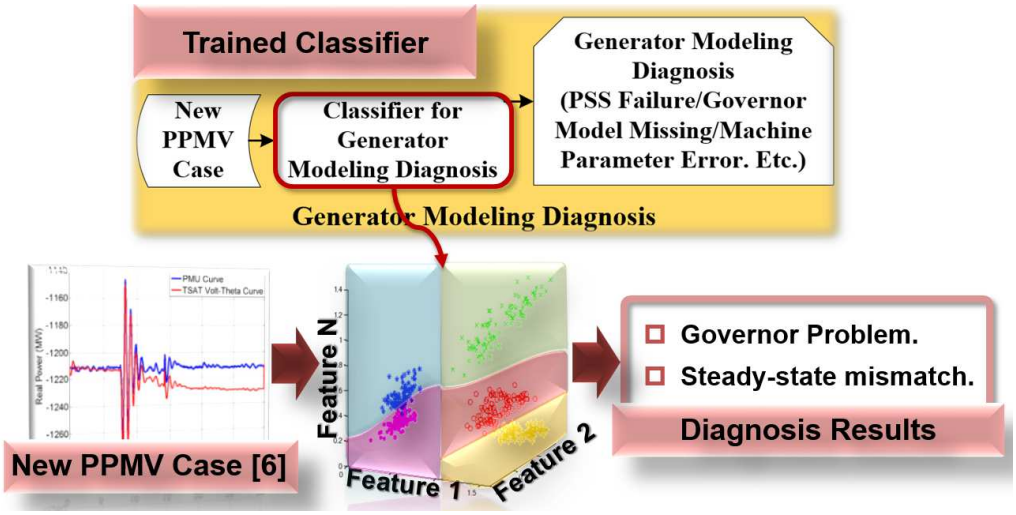

Figure 4.9: Flowchart for the power plant model diagnosis sub-problem.

4.4.2 Proposed Approach

In order to diagnose the power plant model problem using the proposed framework, this section presents a supervised learning approach to differentiate two kinds of power plant model problems: 1) the missing turbine governor model problem; and 2) the wrong machine damping parameter problem. The supervised learning engine is built upon the support vector machine (SVM) method. Two quantification metrics (distance functions) are developed based on the key mismatch features of the above power plant model problems.

4.4.2.1 Key Feature for The Turbine Governor Problem

The following feature is used to identify if the turbine governor model is missing from the entire power plant model:

Feature 1: If the turbine governor model is missing from the entire power plant model, the post-event steady-state real power output of the power plant remains the same as its pre-event steady-state value. If the turbine governor model is included in the entire power plant model, the post-event steady-state real power output of the power plant is different from its pre-event steady-state value. These facts lead to a significant post-event steady-state mismatch between the real power curves obtained using the power plant models with and without the turbine governor model.

The above feature is induced by the underlying physics of the power plant turbine governors. Designed to adjust the mechanical power input (and therefore affecting the electrical power output) of the power plant after a transient event, the turbine governor maintains the real power balance of the power grid by changing the post-event steady-state real power generation to a new set point. Details of various turbine governor models can be found in [42].

The quantification metric (distance function) in (4.1) is proposed to quantify Feature 1.

$$D_1(P^{PMU}, P^{Simu}) = \left| \frac{1}{n} \sum_{k=1}^{n} (p_k^{PMU-Sdy} - p_k^{Simu-Sdy}) \right| \tag{4.1}$$

where P^{PMU} and P^{Simu} denote the vectors containing the real power PMU and simulation outputs, respectively; $P^{PMU-Sdy}$ and $P^{Simu-Sdy}$ denote the $n \times 1$ vectors containing the post-event steady-state real power PMU and simulation outputs, respectively; $p_k^{PMU-Sdy}$ and $p_k^{Simu-Sdy}$ denote the k^{th} data points of $P^{PMU-Sdy}$ and $P^{Simu-Sdy}$, respectively; $D_1(\cdot)$ denotes the distance function between the PMU curve and the simulated curve, when

Feature 1 is applied. The post-event steady state period can be determined using the rate of change of the PMU/simulated curves. When the average rate of change of the curve segment is less than a per-defined threshold, this segment of the curve can be identified as the post-event steady state segment.

4.4.2.2 Key Feature for The Machine Damping Problem

The following feature is used to identify the wrong machine damping parameter problem:

Feature 2: Due to the difference in the settings of the machine damping parameters, the real power PMU and simulation outputs tend to have different dynamic patterns. Significant mismatches can be observed between the PMU real power curve and the simulated real power curve, before the transient event settles down to a new steady state.

The above feature is caused by the fact that the machine damping parameter would have significant impact on the oscillating behavior of the power plant real power output. The quantification metric (distance function) in (4.2) is proposed to quantify Feature 2.

$$D_2(P^{PMU}, P^{Simu}) = ln(DTW(P^{PMU-Detrend}, P^{Simu-Detrend})) \tag{4.2}$$

where P^{PMU} and P^{Simu} denote the vectors containing the real power PMU and simulation outputs, respectively; $P^{PMU-Detrend}$ and $P^{Simu-Detrend}$ denote the vectors containing the real power PMU and simulation outputs, whose mean values are removed from the corresponding P^{PMU} and P^{Simu}, respectively; $DTW(\cdot)$ denotes the dynamic time warping value between two time series, $ln(\cdot)$ denotes the natural logarithm of a variable; $D_2(\cdot)$ denotes the distance function between the PMU curve and the simulated curve, when Feature 2 is applied.

In (4.2), the dynamic time warping is applied to quantify the similarity between the simulated curve and the PMU curve. The dynamic time warping is a technique for effi-

ciently quantifying the similarity between two time series with approximately the same overall component shapes [43]. The formulation of this technique is briefly presented as follows. More details of the dynamic time warping can be found in [43–46].

Let $Q = q_1, Q_2, ..., q_i, ...q_n$ and $C = c_1, c_2, ..., c_j, ..., c_m$ denote the two time series Q and C, of length n and m, respectively. Let D denotes the $n \times m$ matrix whose (i^{th}, j^{th}) element denotes the Euclidean distance between the two points q_i and c_j. A warping path W is a set of matrix elements that defines a mapping between Q and C. The k^{th} element of W is defined as $w_k = (i, j)_k$. Therefore, $W = w_1, w_2, ..., w_k, ..., w_K$, where $max(m, n) \leq K \leq m + n - 1$.

The dynamic time warping technique finds the warping path W which minimizes the warping cost shown in (4.3):

$$DTW(Q, C) = min(\sqrt{\sum_{k=1}^{K} w_k}/K) \tag{4.3}$$

In the meantime, the warping path W needs to satisfy the following constraints:

1. Boundary Conditions: $w_1 = (1, 1)$ and $w_k = (m, n)$. This means the warping path needs to start and finish in diagonally opposite corner cells of the matrix.

2. Continuity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a - a' \leq 1$ and $b - b' \leq 1$. This restricts the allowable steps in the warping path to adjacent cells.

3. Monotonicity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a' - a \leq 0$ and $b' - b \leq 0$. This forces the points in W to be monotonically spaced in time.

According to the above discussion, the dynamic time warping technique can be formulated as a constrained optimization problem. The optimal path W can be found via dynamic programming. To solve this problem, the following recurrence is evaluated in the dynamic programming algorithm:

$$r(i,j) = d(q_i, c_j) + min\{r(i-1, j-1), r(i-1, j), r(i, j-1)\} \tag{4.4}$$

where $r(i,j)$ denotes the cumulative distance at the current recursion.

4.4.2.3 Support Vector Machine Classifier for Power Plant Model Diagnosis

Based on the two key features extracted above, a two-dimensional feature space is established for the classifier training. The proposed distance functions $D_1(\cdot)$ and $D_2(\cdot)$ serve as the quantification metrics for the two dimensions of the feature space. A support vector machine classifier is trained to classify the two different power plant model problems. The classifier training process is briefly described as follows. Details of the support vector machine technique can be found in [47].

Let $(\vec{x_1}, y_1), ..., (\vec{x_n}, y_n)$ denotes the training data set for the classifier. The i^{th} training data point, $(\vec{x_i}, y_i)$, is the labeled data point obtained from the i^{th} PPMV training case. $y_i = 1$ if the turbine governor model is missing from the original model of the i^{th} PPMV case; $y_i = -1$ if the machine damping parameter is incorrect for the i^{th} PPMV case. The 2×1 vector $\vec{x_i} = [D_1(P_i^{PMU}, P_i^{Simu}), D_2(P_i^{PMU}, P_i^{Simu})]^T$, where P_i^{PMU} and P_i^{Simu} denote the PMU and simulated real power curves for the i^{th} PPMV case, respectively. The above definition projects all the training PPMV cases to the two-dimensional feature space.

In this two-dimensional feature space, the support vector machine classifier is trained through solving the following optimization problem:

$$\begin{aligned} & min \ ||\vec{\omega}||_2 \\ & s.t. \ \ y_i(\vec{\omega} \cdot \vec{x_i} - b) \geq 1, \ \ for \ i = 1, ..., n \end{aligned} \tag{4.5}$$

where $\vec{\omega}$ is a 1×2 vector; b is a real value; $||\vec{\omega}||_2$ denotes the 2-norm of $\vec{\omega}$; the hyperplane determined by $\vec{\omega} \cdot \vec{x} - b = 0$ is the separating hyperplane for the two types of modeling problems. The $\vec{\omega}$ and b obtained by solving (4.5) determine the following support vector machine classifier: $\vec{x} \mapsto sgn(\vec{\omega} \cdot \vec{x} - b)$, where $sgn(\cdot)$ denotes the sign function.

Following the above process, a trained linear classifier can be obtained to diagnosis two types of power plant model problems: 1) the missing turbine governor model problem; and 2) the wrong machine damping parameter problem. The trained classifier can then be applied to determine the underlying model problems for the unlabeled PPMV cases.

4.4.3 Case Study

The proposed solution approach is tested using simulated PPMV cases generated from the Western System Coordinating Council (WSCC) 9-bus-system. Figure 4.10 shows the one-line diagram of the test system.

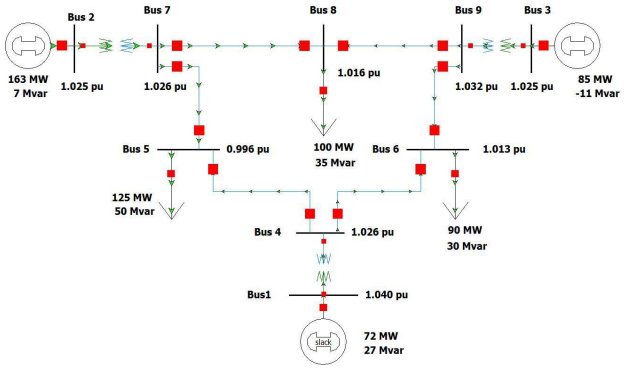

Figure 4.10: One-line diagram of the test system.

In the original system, each power plant model contains a synchronous machine model, an exciter model, and a turbine governor model. In order to simulate the scenarios induced by incorrect power plant models, the power plant model at Bus 2 is modified to simulate

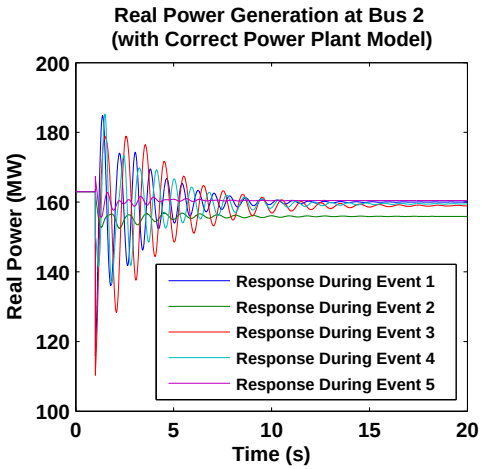

Figure 4.11: Real power generation at bus 2 (with correct power plant model).

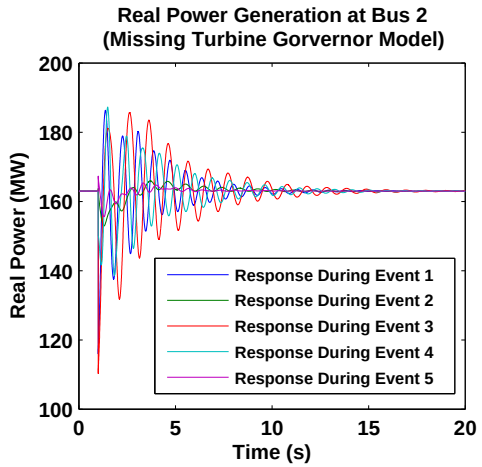

Figure 4.12: Real power generation at bus 2 (without turbine governor model).

two different model parameter/structure errors: 1) missing turbine governor model, where the turbine governor model is removed from the original power plant model; 2) wrong machine damping parameter, where the machine damping parameter is tuned to an incorrect value. Figure 4.11, Figure 4.12, and Figure 4.13 show the real power generation curves at bus 2 during various system events, when the correct power plant model, the incorrect power plant model without the turbine governor model, and the incorrect power plant

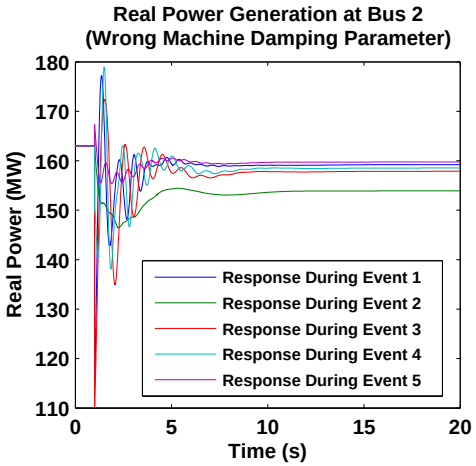

Figure 4.13: Real power generation at bus 2 (with wrong machine damping parameter).

model with wrong machine damping parameter are applied, respectively. By comparing Figure 4.11 and Figure 4.12, it is clear that, when the turbine governor is missing from the original power plant model, the post-event steady-state values of responses during different events return to their pre-event steady-state values, indicating the underlying problem with the turbine governor model. By comparing Figure 4.11 and Figure 4.13, it is clear that there are significant dynamic pattern mismatches between the curves in Figure 4.11 and the curves in Figure 4.13.

In this case study, the following two critical features are extracted for this study: 1) post-event steady-state error; 2) the dynamic pattern mismatch. These features are extracted using methods described in the previous section.

Figure 4.14 shows the steady-state mismatch values for the real power generations during different events, when different model errors are applied. The green bars indicate the steady-state mismatch values of the power plant model without the turbine governor, and the blue bars indicate the steady-state mismatch values of power plant model with the wrong machine damping parameter.

Figure 4.15 shows the dynamic time warping values for the real power generations

during different events, when different model errors are applied. The green bars indicate the dynamic time warping values of the power plant model without the turbine governor, and the blue bars indicate the dynamic time warping values of power plant model with the wrong machine damping parameter.

It is clear from Figure 4.14 and Figure 4.15 that, when the turbine governor model is missing from the power plant model, the steady state mismatches of the real power outputs tend to be more significant; when the wrong machine damping parameter is set for the power plant model, the dynamic pattern mismatches (quantified by the dynamic time warping values) of the real power outputs tend to be more significant.

The trained classifier and the diagnosis result for the test case is shown in Figure 4.16. The similarity metrics are calculated according to the approaches in the previous section. Within the feature space established using the above two critical features, the two different modeling errors can be classified effectively. The correct model diagnosis information for

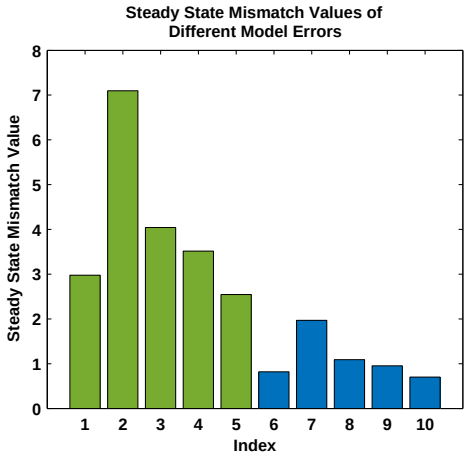

Figure 4.14: Steady state mismatch values for real power generations during different events, when different model errors are applied (the first 5 bars (in green) are steady state mismatch values of the power plant model without turbine governor, the last 5 bars (in blue) are steady state mismatch values of power plant model with wrong machine damping parameter).

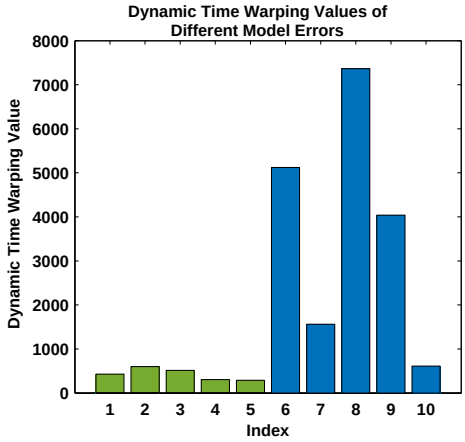

Figure 4.15: Dynamic time warping values for real power generations during different events, when different model errors are applied (the first 5 bars (in green) are dynamic time warping values of the power plant model without turbine governor, the last 5 bars (in blue) are dynamic time warping values of power plant model with wrong machine damping parameter).

test PPMV case can be generated after the classification.

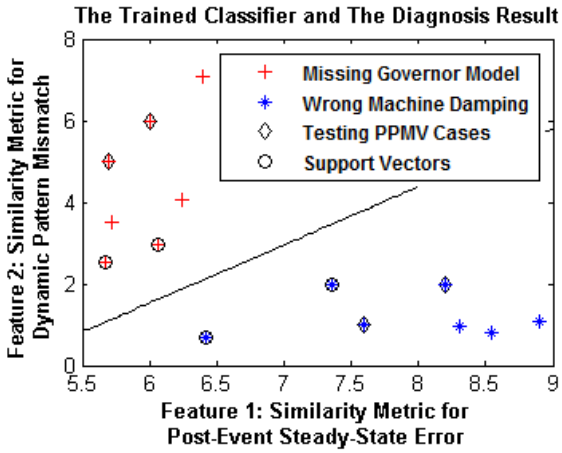

Figure 4.16: The trained classifier and the diagnosis result for the test PPMV case.

4.5 Section Conclusion

In this section, a batch power plant model validation software tool is presented. This tool aims at automatically performing the model validation (without calibration) for multiple power plants using a single disturbance event. Built upon this software tool, a feature-based diagnosis framework is proposed for automating the entire process of PMU-based PPMV. It mimics the human engineering judgement process via a supervised learning engine. This proposed approach could serve as a screening tool for the PPMV engineering judgement process, which could help automate the entire PPMV application.

Future work will include extending the proposed power plant model diagnosis framework to other types of power plant model problems. More mismatch features will be extracted; comprehensive similarity metrics (distance functions) will be proposed for different types of model problems; various classification techniques will be applied in order to obtain better classifiers.

5. PHYSICS-BASED POST EVENT ANALYSIS FOR WIND FARM SUB SYNCHRONOUS OSCILLATIONS*

5.1 Motivation

In recent decades, increasing amounts of wind farms have been integrated into power systems around the world, helping relieve both energy and environmental concerns, while bringing challenging operational issues to system operators at the same time [48]. One of the recently-discovered system instability phenomena is the sub-synchronous oscillation (SSO) induced by wind farm integration. In October 2009, the Electric Reliability Council of Texas (ERCOT) reported a SSO event in their wind-integrated system [49]. The oscillation was triggered by a single line-to-ground fault, and eventually caused unstable power and voltage oscillation in the nearby area. During the past few years, similar wind-induced SSO events have also been observed repeatedly in other wind-integrated systems including Oklahoma Gas and Electric Company (OG&E) in 2011 [50] and China Jibei Power Grid in 2013 and 2014 [51]. Different from SSO incidents in ERCOT system, the events discovered in OG&E and China Jibei Power Grid were not triggered by large disturbances in nearby transmission network. Motivated by this phenomenon, this section tend to analyze wind farm SSO from small-signal-stability point of view, instead of transient stability issues. A recent publication [52] provides detailed modal analysis as well as electro-magnetic simulations to SSO events in China Jibei Power Grid. This section further studies the impact of wind farm spatial distribution on SSO events through applying quantitative sensitivity analysis and parameter adjustment verification to a multi-

*This section is in part a reprint with permission from Meng Wu, Le Xie, Lin Cheng, and Rongfu Sun of the material in the paper: "A Study on The Impact of Wind Farm Spatial Distribution on Power System Sub-Synchronous Oscillations", in *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2154-2162, May 2016 [3]. Copyright 2016 IEEE, and in part a reprint with permission from Meng Wu and Le Xie of the material in the paper: "Calculating Steady-State Operating Conditions for DFIG-Based Wind Turbines", in *IEEE Transactions on Sustainable Energy* (accepted, to appear) [4]. Copyright 2017 IEEE.

wind-farm model derived from a practical wind-integrated system.

Significant progress has been made in the analysis and mitigation of wind-induced SSO. Several key issues on wind-induced SSO have been investigated. Fundamentally, the nature of these SSO incidents observed in wind-integrated systems are similar to the induction generator effect discussed in SSOs related to conventional power plants. These SSOs are caused by the resonance coupling between electrical and magnetic parts of the generator circuits and series-compensated transmission lines. Anderson et. al [53] summarizes theoretical framework and systematical analysis for this kind of electro-magnetic resonance in the context of *synchronous generators*. However, recent studies reveal that the integration of wind farms have introduced new challenges to the analysis and control of this problem: a) among various types of wind turbines, doubly-fed-induction-generator (DFIG)-based wind turbines are the most vulnerable to SSO, and the converter controllers of DFIG are highly participated in the SSO mode [54]; b) the frequency of wind-induced SSO depends closely on DFIG controller parameters, and therefore has a wide range, which increases the difficulty of controller design and parameter tuning [55]; c) in a given transmission network, wind farm consists of many spatially-distributed wind turbines, which requires multi-machine modeling and analysis in the view of SSO. Possible control strategies to mitigate SSO using existing DFIG and FACTS controllers in the system have been proposed, and control signal selections have been discussed for wind-induced SSO mitigation [56–58].

Given the increasing presence of distributed wind power in the grid, there is an increasing concern to fundamentally investigate the cause and counter-measure of such SSO. While a large body of literature exists for the issue of conventional synchronous generator-induced SSO, the phenomena described in this section differ in the following sense: (1) there is a cluster of spatially distributed asynchronous wind generators, and (2) the operating condition (e.g. wind speed) is inherently stochastic. In recent study of wind-induced

SSO, most of the research has been focusing on single-machine-infinite-bus or purely simulation-based study. This section aims at filling the gap of providing rigorous analytical study on the wind farm spatial distribution's impact on SSO.

Motivated by the above gap between current research progress and industrial expectations, this section studies the impact of wind farm spatial distribution on wind-induced SSO using a multi-machine wind farm model derived from a practical wind-integrated power system. Eigenvalue sensitivity with respect to system parameters as well as operating conditions is performed to explore SSO in a multi-wind-farm environment, the coupling of wind turbines with identical and different parameter settings on the studied SSO modes is investigated.

The rest of the section is organized as follows. Section II describes the system modeling techniques adopted in this section; Section III proposes an initialization technique for obtaining the steady-state operating condition of DFIG-based wind turbines; Section IV presents the parameter sensitivity analysis for studying the impact of wind farm spatial distribution on system SSO; Section V shows the case study results and corresponding analysis regarding the relationship between wind turbine parameters and system SSO behavior in a multi-wind-farm environment; Section VI presents the concluding remarks.

5.2 Modeling of Multi-Wind-Farm-Integrated Systems for SSO Studies

Studies discussed in this section are based on the system structure shown in Figure 5.1.

In the proposed multi-machine system structure, wind turbines with the same parameters and nearby physical locations are grouped together and represented by an integrated model of DFIG-based wind farm. Multiple wind farm models with different parameters are connected to the bulk power system through certain network dynamic model. Since studies presented in this section focus on the SSO problem caused by wind farm integration, which has been shown to be related to series compensation of transmission lines

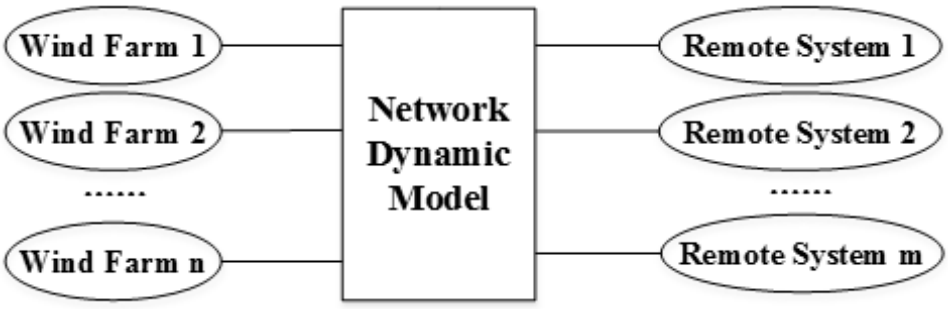

Figure 5.1: Overall structure of system model [3].

close to the wind farms [54], remote bulk power systems are modeled by infinite buses, and therefore no dynamics are considered. The above system structure is suitable for studying the impact of wind farm spatial distribution on SSO problem, without introducing too much computational complexity.

It should be noted that model of wind turbines could be of different level of details. However, in this section, components that are essential for the purpose of SSO analysis and control are selected to be modeled in detail. Discussions regarding essential models for wind farm SSO analysis can be found in [56,59–61]. In order to check the model fidelity of wind turbines, online damping estimation methods proposed in recent publications [62–67] could be used for the purpose of comparing damping ratios between practical wind turbines and their models, once the maturity of online damping estimation methods is suitable for practical applications.

5.2.1 Modeling of DFIG-Based Wind Farm

For each integrated DFIG-based wind turbine model, wind turbine model described in [59] is adopted. Figure 5.2 shows the overall structure of a DFIG-based wind turbine with grid integration. Four dynamical subsystems are modeled:

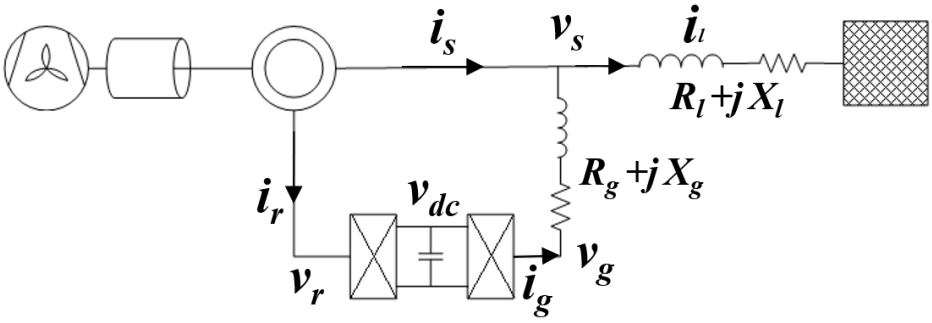

Figure 5.2: Structure of DFIG-based wind turbine with grid integration [3].

5.2.1.1 Wind Turbine Mechanical Shaft Model

The multi-mass wind turbine mechanical shaft model is adopted in this section. Its dynamics for i^{th} mass can be represented by the ODEs shown in (5.1) [68]:

$$\begin{cases} \dot{\delta}_i = \Delta\omega_i = \omega_i - \omega_i^* \\ J_i\dot{\Delta\omega}_i = \tau_i + K_{i,i+1}(\delta_{i+1} - \delta_i) \\ \qquad\qquad - K_{i,i-1}(\delta_i - \delta_{i-1}) \\ \qquad\qquad + D_{i,i+1}(\Delta\omega_{i+1} - \Delta\omega_i) \\ \qquad\qquad - D_{i,i-1}(\Delta\omega_i - \Delta\omega_{i-1}) - D_i\Delta\omega_i \end{cases} \tag{5.1}$$

where $\omega_i^* = \frac{\omega_r^*}{\nu_i(pole/2)}$, $\nu_i = \frac{\omega_g^{rate}}{\omega_i^{rate}}$.

5.2.1.2 DFIG Electro-magnetic Model

The DFIG electro-magnetic dynamics can be represented by the 4th order dynamic model shown in (5.2) [69]:

$$\begin{bmatrix} v_{qs} \\ v_{ds} \\ v_{qr} \\ v_{dr} \end{bmatrix} = M_{gen}(\omega_r) \begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{qr} \\ i_{dr} \end{bmatrix} + N_{gen} \begin{bmatrix} \dot{i}_{qs} \\ \dot{i}_{ds} \\ \dot{i}_{qr} \\ \dot{i}_{dr} \end{bmatrix} \tag{5.2}$$

where

$$M_{gen}(\omega_r) = \begin{bmatrix} -r_s & -\frac{\omega}{\omega_b}X_{ss} & 0 & -\frac{\omega}{\omega_b}X_m \\ \frac{\omega}{\omega_b}X_{ss} & -r_s & \frac{\omega}{\omega_b}X_m & 0 \\ 0 & -\frac{\omega-\omega_r}{\omega_b}X_m & -r_r & -\frac{\omega-\omega_r}{\omega_b}X_{rr} \\ \frac{\omega-\omega_r}{\omega_b}X_m & 0 & \frac{\omega-\omega_r}{\omega_b}X_{rr} & -r_r \end{bmatrix} \tag{5.3}$$

$$N_{gen} = \begin{bmatrix} -\frac{X_{ss}}{\omega_b} & 0 & -\frac{X_m}{\omega_b} & 0 \\ 0 & -\frac{X_{ss}}{\omega_b} & 0 & -\frac{X_m}{\omega_b} \\ -\frac{X_m}{\omega_b} & 0 & -\frac{X_{rr}}{\omega_b} & 0 \\ 0 & -\frac{X_m}{\omega_b} & 0 & -\frac{X_{rr}}{\omega_b} \end{bmatrix} \tag{5.4}$$

In (5.2)-(5.4), full-order dynamics for DFIG stator and rotor are modeled under synchronous reference frame. The DFIG is assumed to operate in three-phase balanced condition.

5.2.1.3 DFIG Converter Controller Model

In general, DFIG converter controllers can be represented using (5.5):

$$\begin{cases} \dot{x_{ctrl}} = A_{ctrl}x_{ctrl} + B_{ctrl}u_{ctrl} \\ y_{ctrl} = C_{ctrl}x_{ctrl} + D_{ctrl}u_{ctrl} \end{cases} \tag{5.5}$$

The converter controller model shown in (5.5) usually applies d-q decoupled control to regulate DFIG power/torque and currents [68]. u_{ctrl}, y_{ctrl} are typically electrical variables of the DFIG-based wind turbine, and A_{ctrl}, B_{ctrl}, C_{ctrl}, D_{ctrl} are determined by the structure of the PI control loops.

5.2.2 DFIG DC-Link Model

DFIG DC-link dynamics can be represented by the ODE shown in (5.6) [56]:

$$C_{dc}v_{dc}\dot{v_{dc}} = P_{rsc} - P_{gsc} - P_{loss}^{converter} \tag{5.6}$$

where

$$\begin{cases} P_{rsc} = v_{qr}i_{qr} + v_{dr}i_{dr} \\ P_{gsc} = v_{qs}i_{qg} + v_{ds}i_{dg} \\ P_{loss}^{converter} = i_{qg}^2R_g + i_{dg}^2R_g \end{cases} \tag{5.7}$$

Equations (5.6) and (5.7) describe the dynamics of real power delivery through DFIG rotor and back-to-back converters. It should be noted that DFIG reactive power delivery through rotor and converters, Q_g, is usually regulated by DFIG converter controllers. In most situations, Q_g^{ref} is set to be close to zero. The algebraic equation shown in (5.8) represents steady-state relationship for DFIG reactive power delivery through rotor and converters:

$$Q_g^{ref} = Q_g \tag{5.8}$$

where $Q_g = v_{qs}i_{dg} - v_{ds}i_{qg}$.

5.2.3 DFIG-based Wind Turbine Grid Integration Model

The algebraic equations shown in (5.9) represent the steady-state bus current injection relationship when a DFIG-based wind turbine is connected with certain transmission system:

$$\begin{cases} i_{ql} = i_{qs} + i_{qg} \\ i_{dl} = i_{ds} + i_{dg} \end{cases} \tag{5.9}$$

Based on (5.9), the overall dynamic model for DFIG-based wind turbine with grid integration can be represented by (5.1)-(5.9). This model is adopted in the rest of the section to calculate the steady-state operating conditions of DFIG-based wind turbines.

5.2.4 Modeling of Dynamic Network System

In this section, dynamic network system consisting resistors, inductors and capacitors is modeled under synchronous reference frame, inductor currents and capacitor voltages are treated as state variables. It should be emphasized that there are two types of networks which need to be considered separately: the non-degenerate network and degenerate network. A network is said to be degenerate if [53] (1) it contains a circuit (loop) composed only of capacitors and/or independent or dependent voltage sources; (2) it contains a cut set composed only of inductors and/or independent or dependent current sources. For degenerate network, not all the capacitor voltages and inductor currents are independent, and therefore, those dependent ones cannot be considered to be state variables. In this section, the following procedures are taken to build general dynamic network model:

Step B.1: classify all system components into four types: voltage source (wind farms and remote systems), resistors (transmission line resistance), capacitors (series and shunt capacitors, etc.), and inductors (transmission line inductance, shunt inductors, etc.).

Step B.2: merge all parallel-connected components of the same type into one integrated

component.

Step B.3: create super bus: if a bus is connected with more than three integrated component, identify it as a super bus.

Step B.4: create super component: merge all series-connected integrated components of the same type between two super buses into one super component.

Step B.5: identify dependent capacitors: check each loop in the system, if there exists a loop containing only super capacitors and super voltage sources, then identify one of the grounded super capacitors as dependent capacitor, if there is no grounded super capacitors presented, identify the capacitor with smallest susceptance as dependent capacitor.

Step B.6: identify dependent inductors: check each cut set in the system, if there exists a cut set containing only super inductors, then identify one of the grounded super inductors as dependent inductor, if there is no grounded super inductors presented, identify the inductor with smallest susceptance as dependent inductor.

Step B.7: build network model: terminal voltages of super voltage sources are considered as inputs of the network. For each independent inductor/capacitor, the inductor current/capacitor voltage under synchronous reference frame are considered to be state variables. The dynamics of independent inductor/capacitor is represented by the following equations:

$$\begin{cases} \omega L_i i^d_{L_i} + L_i \dot{i^q_{L_i}} = v^q_{L_i} \\ -\omega L_i i^q_{L_i} + L_i \dot{i^d_{L_i}} = v^d_{L_i} \end{cases} \tag{5.10}$$

$$\begin{cases} \omega C_j v^d_{C_j} + C_j \dot{v^q_{C_j}} = i^q_{C_j} \\ -\omega C_j v^q_{C_j} + C_j \dot{v^d_{C_j}} = i^d_{C_j} \end{cases} \tag{5.11}$$

The currents/voltages of dependent inductors/capacitors as well as resistors can be represented by network inputs and state variables through Kirchhoff's law.

Through the above procedures, the following model for general network dynamic sys-

tem can be obtained:

$$\dot{x_{nt}} = A_{nt}x_{nt} + B_{nt}u_{nt} \tag{5.12}$$

where $x_{nt} = [i^q_{L_1}, i^d_{L_1}, \cdots, i^q_{L_s}, i^d_{L_s}, v^q_{C_1}, v^d_{C_1}, \cdots, v^q_{C_t}, v^d_{C_t}]^T$ denotes the state variables for the network system, $u_{nt} = [v^1_{qs}, v^1_{ds}, \cdots, v^n_{qs}, v^n_{ds}, v^1_{qinf}, v^1_{dinf}, \cdots, v^m_{qinf}, v^m_{dinf}]^T$ denotes the inputs for the network system, s, t, m, n denotes the total number of inductors, capacitors, remote power systems, and wind farms modeled in the study system, respectively; A_{nt} and B_{nt} denotes the state matrix and input matrix of network system, respectively.

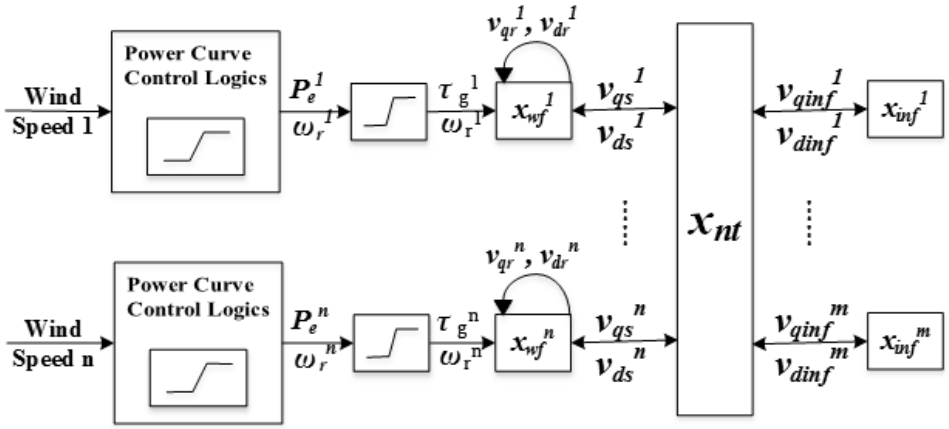

Figure 5.3: System input/output relationship [3].

5.2.5 System Integration and Initialization

According to the previous discussions, the input/output relationship of the entire system can be represented by Figure 5.3. Wind farm models and network model are interconnected through DFIG stator voltages. The entire system model can be described by the following nonlinear ODEs:

$$\dot{x_{all}} = f_{all}(x_{all}, u_{all}, p_{all}) \tag{5.13}$$

where $x_{all} = [x_{wf}^1, \cdots, x_{wf}^n, x_{nt}]^T$ denotes system state variables; $p_{all} = [p_{wf}^1, \cdots, p_{wf}^n, p_{nt}]^T$ denotes system parameters; $u_{all} = [P_e^1, \omega_r^1, \cdots, P_e^n, \omega_r^n, v_{qinf}^1, v_{dinf}^1, \cdots, v_{qinf}^m, v_{dinf}^m]^T$ denotes system inputs.

In order to perform small signal analysis, steady-state operating point of the system under certain operating condition should be obtained through model initialization. In this section, the wind farm model and multi-machine network model are initialized separately using system power flow solutions.

5.2.5.1 Initialization of Network Model

Dynamic network model is initialized through the following steps:

Step C.1: build power flow model for the multi-machine network according to its dynamic model. Each wind farm is modeled as PQ bus with positive real and reactive power injection to the network. One infinite bus is modeled as slack bus, the other infinite buses are modeled as PV bus with negative real and reactive power injection to the network.

Step C.2: solve power flow problem for the network and obtain steady-state voltages for all wind farms and infinite buses. Calculate the corresponding voltages on q and d axes.

Step C.3: obtain steady-state operating points of the network system by solving (5.12) with $\dot{x_{nt}} = 0$.

Through the above procedures, initial conditions for network state variables can be obtained directly from power flow solution of the corresponding steady-state network model.

5.3 Initialization of The DFIG-Based Wind Turbine Model

Initialization of DFIG-based wind farm model has been discussed in recent literature [70–72], while agreement on the most appropriate approach has not been reached. This section proposes an analytical approach to calculate the initial steady-state operating conditions for DFIG-based wind turbines, after valid power flow solutions are obtained.

The key idea stems from formulating the original problem using a set of linear-quadratic equations, which makes the problem analytically solvable. Advantages of the proposed method over previous direct initialization methods can be summarized as follows: (1) analytical formulation and closed-form solution to the initialization problem are provided, which can be easily implemented in commercial software for power system transient simulation; (2) a full-order dynamic model under d-q reference frame is adopted as representation for DFIG system; (3) loss of DFIG back-to-back converter is considered; (4) DFIG dq-reference frame can be of arbitrary direction, i.e. $v_{ds} = |\vec{v_s}|$ and $v_{qs} = 0$ are not assumed; (5) nonzero reactive power delivery through DFIG grid-side converter can be supported; (6) model re-initialization can be performed freely at any time, and initialization for wind turbine operations below rated output can be well supported. The proposed method takes advantage of the relationship among wind speed, electrical power output, and DFIG rotor speed, which can be obtained through either field tests or designed control logics.

5.3.1 Formulation of the Initialization Problem for DFIG-based Wind Turbines

In order to obtain steady-state operating conditions for DFIG-based wind turbines, all the time-derivative terms in the dynamic models shown in (5.1)-(5.5) need to be set to zero, that is, the DFIG-based wind turbine needs to operate at steady state. The algebraic equations shown in (5.14)-(5.24) represent the steady-state relationship for the model initialization process.

$$\begin{aligned}\tau_i = &- K_{i,i+1}(\delta_{i+1} - \delta_i) + K_{i,i-1}(\delta_i - \delta_{i-1}) \\ &- D_{i,i+1}(\Delta\omega_{i+1} - \Delta\omega_i) \\ &+ D_{i,i-1}(\Delta\omega_i - \Delta\omega_{i-1}) + D_i\Delta\omega_i\end{aligned} \tag{5.14}$$

$$\Delta\omega_i = \omega_i - \omega_i^* = 0 \tag{5.15}$$

$$v_{qs} = -r_s i_{qs} - \frac{\omega}{\omega_b} X_{ss} i_{ds} - \frac{\omega}{\omega_b} X_m i_{dr} \tag{5.16}$$

$$v_{ds} = \frac{\omega}{\omega_b} X_{ss} i_{qs} - r_s i_{ds} + \frac{\omega}{\omega_b} X_m i_{qr} \tag{5.17}$$

$$v_{qr} = -\frac{\omega - \omega_r}{\omega_b} X_m i_{ds} - r_r i_{qr} - \frac{\omega - \omega_r}{\omega_b} X_{rr} i_{dr} \tag{5.18}$$

$$v_{dr} = \frac{\omega - \omega_r}{\omega_b} X_m i_{qs} + \frac{\omega - \omega_r}{\omega_b} X_{rr} i_{qr} - r_r i_{dr} \tag{5.19}$$

$$-v_{qs} i_{qg} - v_{ds} i_{dg} + v_{qr} i_{qr} + v_{dr} i_{dr} = i_{qg}^2 R_g + i_{dg}^2 R_g \tag{5.20}$$

$$Q_g^{ref} = v_{qs} i_{dg} - v_{ds} i_{qg} \tag{5.21}$$

$$A_{ctrl} x_{ctrl} + B_{ctrl} u_{ctrl} = 0 \tag{5.22}$$

$$i_{ql} = i_{qs} + i_{qg} \tag{5.23}$$

$$i_{dl} = i_{ds} + i_{dg} \tag{5.24}$$

In (5.14)-(5.24), the steady-state mechanical shaft model is represented by (5.14)-(5.15), the steady-state DFIG model is represented by (5.16)-(5.19), the steady-state dc-link model is represented by (5.20)-(5.21), the steady-state converter controller model is represented by (5.22), and the steady-state grid integration relationship for DFIG-based wind turbine is represented by (5.23)-(5.24).

The steady state calculation problem for DFIG-based wind turbines can be formulated as solving the algebraic equations shown in (5.14)-(5.24), and obtaining steady-state values for all the state variables involved in the wind turbine dynamic model. In the following section, detailed discussions on known and unknown variables in (5.14)-(5.24) as well as analytical solutions to this problem are presented.

5.3.2 Calculating Steady State Operating Conditions for DFIG-based Wind Turbines

In order to solve the initialization problem formulated in the previous section, known variables (system inputs) and unknown variables in (5.14)-(5.24) should be identified. In this section, the following assumptions are made on the known variables:

Assumption 1. *There exists a valid power flow solution for the studied system before the steady state calculations. Steady-state bus voltage phasor $\vec{v_s}$ and current injection phasor $\vec{i_l}$ of the studied DFIG-based wind turbine can be obtained from the power flow analysis.*

Based on the above assumption, when DFIG is modeled under synchronous reference frame, its stator voltage and terminal current on q and d axes can be derived from $\vec{v_s}$ and $\vec{i_l}$. Therefore, v_{qs}, v_{ds}, i_{ql}, and i_{dl} can be considered as known variables for the initialization problem.

Assumption 2. *DFIG steady-state rotor speed ω_r^* has a one-to-one mapping with wind speed, and the value of $\omega_r^*(v_{wind})$ can be obtained under any wind speed condition, that is, ω_r^* is a known variable for the initialization problem.*

This assumption requires the rotor speed-wind speed curve $\omega_r^*(v_{wind})$ should be known and stored prior to the steady state calculations. In practical applications, this curve can be obtained using field test data or supervisory control objectives. It should be emphasized that generating $\omega_r^*(v_{wind})$ curve via supervisory control objectives is practically applicable for various types of wind turbine control systems. Key parameters for calculating the $\omega_r^*(v_{wind})$ relationship can be obtained through vendors of the wind turbines. Detailed discussion on wind turbine control systems can be found in [68]. An example for obtaining the steady-state reference signals for ω_r^* and the blade pitch angle of a series of General Electric (GE) wind turbines can be found in [73].

Based upon the above assumptions on known variables (v_{qs}, v_{ds}, i_{ql}, i_{dl}, ω_r^*) for the steady state calculation problem, the following sections present details of the proposed direct initialization approach.

5.3.2.1 Steady-State Operating Conditions of Mechanical Shaft Model

According to the steady-state relationship of mechanical shaft model shown in (5.14)-(5.15), the steady-state condition for state variable ω_r can be represented as (5.25):

$$\begin{cases} \omega_i = \omega_i^* \\ \omega_r = \omega_r^* \end{cases} \tag{5.25}$$

It should be noted that under steady-state condition, $\tau_i = -K_{i,i+1}(\delta_{i+1} - \delta_i)$.

It is clear from the above discussions that calculating steady-state ω_r involves no knowledge on state variables of DFIG, DC-link or converter controller model. Once steady-state ω_r is obtained, it can be considered as a known input variable for the steady-state analysis of the other three models.

5.3.2.2 Steady-State Operating Conditions of DFIG Model and DC-link Model

The DFIG and DC-link models together form a nonlinear dynamical system. Its steady-state relationship is represented in (5.16)-(5.21), (5.23)-(5.24). Since v_{qs}, v_{qs}, i_{ql}, i_{dl} are assumed to be known, and ω_r is considered as known input from the steady-state analysis of mechanical shaft model, the following 8 unknown variables need to be determined: v_{qr}, v_{dr}, i_{qr}, i_{dr}, i_{ds}, i_{qs}, i_{qg}, i_{dg}. In order to obtain these unknowns, nonlinear equations (5.16)-(5.21), (5.23)-(5.24) need to be solved. Since the total number of equations match with the total number of unknowns, this problem is fully determined.

Theorem 1. *The steady state calculation problem for DFIG model and DC-link model has at most one solution. If the solution exists, analytical formats of all the 8 unknowns are*

available.

Proof. It can be concluded that in the 8 equations shown in (5.16)-(5.21), (5.23)-(5.24), Equations (5.16)-(5.19), (5.21), (5.23)-(5.24) represent linear relationship among the 8 unknowns, while Equation (5.20) represents nonlinear (quadratic) relationship among the 8 knowns. Therefore, these equations can be grouped as (5.26) and (5.27):

$$A \begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_{i+1} \\ \vdots \\ x_8 \end{bmatrix} = bx_i \tag{5.26}$$

$$f([x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_8], x_i) = 0 \tag{5.27}$$

where (5.26) represents linear relationship shown in (5.16)-(5.19), (5.21), (5.23)-(5.24), and (5.27) represents nonlinear relationship shown in (5.20).

x_i denotes an arbitrary unknown variable selected from the 8 unknowns, i is an integer satisfying $1 \leq i \leq 8$, $[x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_8]^T$ denotes the vector of unknown variables except for x_i. A denotes the parameter matrix for vector $[x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_8]^T$, b denotes the parameter vector for unknown variable x_i.

According to the linear relationship of (5.26), Equation (5.28) can be obtained:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_{i+1} \\ \vdots \\ x_8 \end{bmatrix} = A^{-1}bx_i \tag{5.28}$$

Therefore, the nonlinear relationship of (5.27) can be re-written as (5.29):

$$f([A^{-1}bx_i]^T, x_i) = 0 \tag{5.29}$$

Since (5.27) is a quadratic function of the 8 unknowns, it can be concluded that (5.29) represents a quadratic function of single unknown variable x_i. The number of solutions to (5.29) lies within the following three situations:

a) There is 0 solution to (5.29): in this situation, solution to x_i does not exist. Therefore, according to (5.28), solutions to $[x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_8]^T$ do not exist either. The original problem has no solution.

b) There are 2 different solutions to (5.29): in this situation, there exists two different solutions to x, which correspond to two distinguished equilibrium points of the DFIG/DC-link model. However, the following paragraph shows at least one of the two equilibrium points is unstable.

If i_{qg} is selected to be the single unknown variable x_i, (5.29) can be represented as (5.30):

$$ai_{qg}^2 + bi_{qg} + c = 0 \tag{5.30}$$

If within a small neighborhood of an equilibrium point, the system input i_{ql} is perturbed along the trajectory shown in (5.31):

$$\dot{i_{ql}} = \dot{i_{qs}} + (ai_{qg}^2 + bi_{qg} + c) \tag{5.31}$$

Then according to (5.23), within a small neighborhood of an equilibrium point, i_{qg} is perturbed along the trajectory shown in (5.32):

$$\dot{i_{qg}} = ai_{qg}^2 + bi_{qg} + c \tag{5.32}$$

When $a < 0$, Figure 5.4 represents the vector field for (5.32) along the direction of i_{qg}:

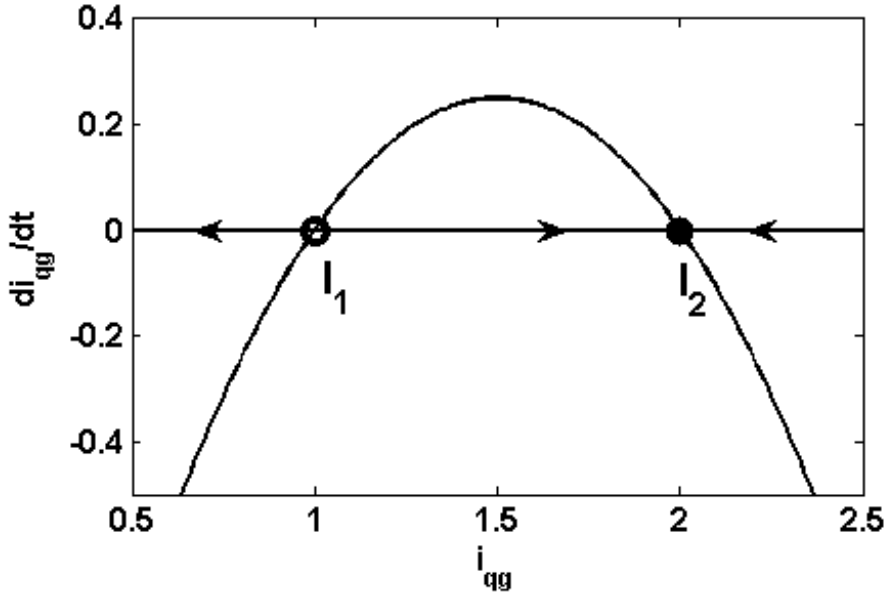

Figure 5.4: Vector field for (5.32) along the direction of i_{qg} [4].

It can be concluded from Figure 5.4 that, when system input i_{ql} is perturbed along the trajectory of (5.31), within the small neighborhood of equilibrium point I_1, i_{qg} will

leave its steady-state value I_1 with a local velocity $\dot{i_{qg}}$. According to the system algebraic relationship, state variables of the system will also leave their steady-state values. Therefore, within the small neighborhood of I_1, there exists a certain kind of small perturbation which will drive the system away from the neighborhood of I_1. This indicates that I_1 is an unstable equilibrium point of the system.

When $a > 0$, similar idea can be applied and I_2 can be shown to be unstable. Therefore, there is at most one stable equilibrium point in the system, and there is at most one valid solution to the original problem, which corresponds to the stable equilibrium point.

c) There is 1 solution to (5.29): in this situation, there exists a unique solution to x, which corresponds to an equilibrium point of the DFIG/DC-link model. Since there is a bifurcation happening in this situation (the total number of equilibrium points in the system is changing from 0 to 1 to 2), the system itself is structurally unstable, that is, a small perturbation in system parameter or input would affect dynamic behavior of the system [74]. Since it is not meaningful to start simulation or eigenvalue analysis from a structurally unstable system, this unique solution is not valid for the original problem.

Therefore, there is at most one valid solution x to the original problem, whose x_i can be obtained analytically by solving (5.29). Analytical formats of the remaining 7 unknowns $[x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_8]^T$ can be determined through (5.28). □

The above theorem and proof demonstrate an approach to calculate steady-state operating conditions for DFIG-based wind turbines in an analytical way. For the purpose of implementation, this approach can be realized by selecting i_{qg} as unknown variable x_i. The derivations shown in (5.33)-(5.41) represent the procedure of obtaining analytical solutions to the 8 unknowns:

From (5.21), i_{dg} can be expressed as (5.33):

$$i_{dg} = \frac{v_{ds}}{v_{qs}} i_{qg} + \frac{Q_g^{ref}}{v_{qs}} \tag{5.33}$$

Substitute (5.33) into (5.23) and (5.24):

$$i_{qs} = i_{ql} - i_{qg} \tag{5.34}$$

$$i_{ds} = i_{dl} - \frac{v_{ds}}{v_{qs}} i_{qg} - \frac{Q_g^{ref}}{v_{qs}} \tag{5.35}$$

Substitute (5.34) and (5.35) into (5.16) and (5.17):

$$\begin{aligned} i_{dr} &= (\frac{\omega_b r_s}{\omega X_m} + \frac{X_{ss} v_{ds}}{X_m v_{qs}}) i_{qg} \\ &\quad + (\frac{Q_g^{ref} X_{ss}}{X_m v_{qs}} - \frac{\omega_b}{\omega X_m} v_{qs} - \frac{\omega_b r_s}{\omega X_m} i_{ql} - \frac{X_{ss}}{X_m} i_{dl}) \\ &= a_1 i_{qg} + b_1 \end{aligned} \tag{5.36}$$

$$\begin{aligned} i_{qr} &= (\frac{X_{ss}}{X_m} - \frac{\omega_b r_s v_{ds}}{\omega X_m v_{qs}}) i_{qg} \\ &\quad + (\frac{\omega_b}{\omega X_m} v_{ds} - \frac{X_{ss}}{X_m} i_{ql} + \frac{\omega_b r_s}{\omega X_m} i_{dl} - \frac{Q_g^{ref} \omega_b r_s}{\omega X_m v_{qs}}) \\ &= a_2 i_{qg} + b_2 \end{aligned} \tag{5.37}$$

Substitute (5.34)-(5.37) into (5.18) and (5.19):

$$\begin{aligned} v_{qr} &= (\frac{(\omega - \omega_r) X_m v_{ds}}{\omega_b v_{qs}} - a_2 r_r - \frac{(\omega - \omega_r) X_{rr} a_1}{\omega_b}) i_{qg} \\ &\quad + (\frac{(\omega - \omega_r) Q_g^{ref} X_m}{\omega_b v_{qs}} - \frac{\omega - \omega_r}{\omega_b} X_m i_{dl} - b_2 r_r \\ &\quad - \frac{(\omega - \omega_r) X_{rr} b_1}{\omega_b}) \\ &= a_3 i_{qg} + b_3 \end{aligned} \tag{5.38}$$

$$
\begin{aligned}
v_{dr} &= (\frac{(\omega-\omega_r)X_{rr}a_2}{\omega_b} - \frac{(\omega-\omega_r)X_m}{\omega_b} - r_r a_1) i_{qg} \\
&+ (\frac{(\omega-\omega_r)X_{rr}b_2}{\omega_b} + \frac{(\omega-\omega_r)X_m}{\omega_b} i_{ql} - r_r b_1) \\
&= a_4 i_{qg} + b_4
\end{aligned} \tag{5.39}
$$

Substitute (5.33), (5.36) - (5.39) into (5.20):

$$
a_5 i_{qg}^2 + b_5 i_{qg} + c_5 = 0 \tag{5.40}
$$

where:

$$
\begin{aligned}
a_5 =& a_2 a_3 + a_1 a_4 - R_g - R_g \frac{v_{ds}^2}{v_{qs}^2} \\
b_5 =& - v_{qs} - \frac{v_{ds}^2}{v_{qs}} + a_3 b_2 + a_2 b_3 + a_4 b_1 \\
&+ a_1 b_4 - R_g \frac{2 Q_g^{ref} v_{ds}}{v_{qs}^2} \\
c_5 =& (-\frac{Q_g^{ref} v_{ds}}{v_{qs}} + b_2 b_3 + b_1 b_4 - \frac{R_g {Q_g^{ref}}^2}{v_{qs}^2})
\end{aligned}
$$

Through solving (5.40), analytical solution of unknown variable i_{qg} can be obtained using (5.41):

$$
\begin{cases}
i_{qg}^{(1)} = \dfrac{-b_5 - \sqrt{b_5^2 - 4a_5 c_5}}{2a_5} \\
i_{qg}^{(2)} = \dfrac{-b_5 + \sqrt{b_5^2 - 4a_5 c_5}}{2a_5}
\end{cases} \tag{5.41}
$$

Substituting (5.41) into (5.33)-(5.39), two set of solutions for the other 7 unknown variables can be obtained accordingly. When $a_5 < 0$, the unique candidate for the valid steady-state operating conditions corresponds to $i_{qg}^{(2)}$. When $a_5 > 0$, the unique candidate for the valid steady-state operating conditions corresponds to $i_{qg}^{(1)}$.

5.3.2.3 Steady-State Operating Conditions of Converter Controller Model

It is obvious that at steady state, all the state variables of converter controller should match their reference values. Therefore, the steady-state results shown in (5.42) can be obtained:

$$x_{ctrl} = x_{ctrl}^{ref} \tag{5.42}$$

where x_{ctrl}^{ref} denotes the reference values for controller state variables. Different sets of reference values can be specified in order to study wind turbine performance under different steady-state operating conditions.

For wind turbines with complex nonlinear control structures, the steady-state operating conditions of the nonlinear control elements (such as the saturation, hysteresis, and relays) can be obtained through standard initialization techniques for generator controllers. It should be noted that these nonlinear control elements are standard elements for not only DFIG-based wind turbines but also other types of generators (such as synchronous generators that ware widely modeled in power system simulation tools). The initialization problem for these nonlinear control structures can be solved using existing techniques, and therefore is out of the scope of this section. In [73], one possible realization of the initialization technique for the nonlinear control structures of a series of GE wind turbines is discussed. This realization can be generalized to obtaining the steady-state operating conditions of various wind turbine control structures.

It should be noted that the proposed approach is not suitable for initializing the wind turbine generic models implemented in most commercial software for power system electromechanical simulations. In the generic models developed by the Western Electricity Coordinating Council (WECC) Renewable Energy Modeling Task Force [75, 76], the detailed DFIG electro-magnetic dynamic model is ignored and replaced by a simplified generator/converter model. The steady-state operating conditions of the simplified genera-

tor/converter model can be obtained using existing initialization techniques for transient simulation software, without applying the proposed approach.

5.4 Sensitivity-Based Analysis of Wind Farm Spatial Distribution Impact on System SSO

The multi-machine model proposed in the previous section describes a wind farm with different types of wind turbines at different physical locations. This structure indicates an heterogeneous and spatially-distributed wind farm. It may have unique impact on system SSO which cannot be fully revealed by single-machine or homogeneous wind farm models. Specifically, the coupling relationship among spatially-distributed wind turbines with identical and different parameter settings may affect SSO behavior of the nearby system. In order to explore the coupling relationship among wind turbines, quantitative eigenvalue sensitivity analysis with respect to the parameters of each wind turbine as well as operating conditions of the wind farm is performed based on the proposed heterogeneous and spatially-distributed wind farm model. The following technique is adopted for the sensitivity analysis:

Let A be an $n \times n$ matrix, and let λ be eigenvalues of A with corresponding right eigenvector x and left eigenvector y. Suppose A is perturbed to $A^{new} = A + \delta A$, and consequently λ is perturbed to $\lambda^{new} = \lambda + \delta\lambda$. If $\|\delta A\|_2 = \epsilon$ is sufficiently small, then [77]

$$\delta\lambda = \frac{y^T(\delta A)x}{y^T x} + O(\epsilon^2) \tag{5.43}$$

The above equation provides information on sensitivities of eigenvalues of matrix A with respect to small perturbations in the matrix elements. For the analysis of wind farm SSO, A is considered to be the linearized state matrix of dynamical system described by (5.13), which can be fully determined by system parameters and operating conditions at a particular moment. Therefore, through perturbation of wind turbine parameters and

operating conditions, the eigenvalue sensitivities corresponding to the perturbed parameter or operating condition can be obtained.

In order to study the coupling relationship among wind turbines with identical or different parameter settings, and explore its potential impact on system SSO, eigenvalue sensitivities with respect to all the parameters for each turbine are computed under a particular operating condition. If there exists a particular SSO mode, whose corresponding eigenvalue is highly sensitive to parameters of a group of wind turbines, then this group of turbines are expected to have strong coupling relationship on this SSO mode. If eigenvalue of certain SSO mode is highly sensitive to parameters of only one wind turbine, then this turbine is expected to be weakly coupled with other turbines in the system.

5.5 Case Studies and Discussions

The above modeling and analysis techniques are applied to the following test systems in order to study the impact of wind farm spatial distribution on system SSO.

5.5.1 Practical Two-Wind-Farm System Model

5.5.1.1 System Description

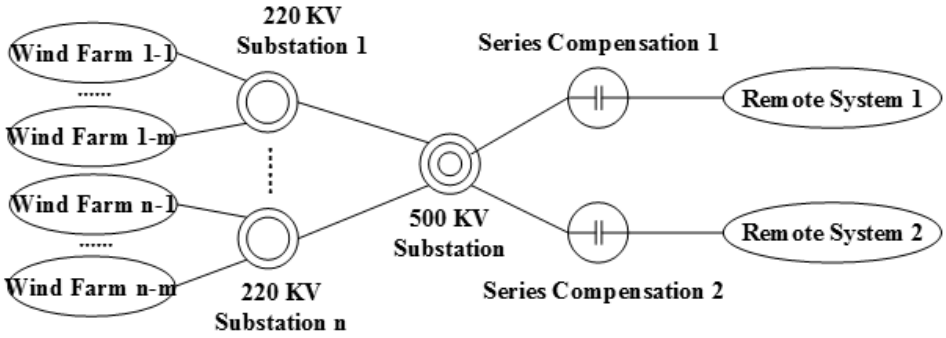

Figure 5.5: Structure of wind energy integration in China Jibei Power Grid [3].

This part of study is conducted on a two-wind-farm system model developed from practical wind-integrated power system in China Jibei Power Grid. Structure of the original Jibei wind-integrated system is shown in Figure 5.5, where multiple wind farms are connected to two remote bulk power systems through a tree-structured network with series compensations on both 500kV transmission lines. However, if the system is modeled in detail with all the wind farms taken into consideration, too much computational complexity will be introduced due to increment of system dimension. In order to release computational burden as well as study the impact of spatial distribution of wind turbines with different parameters, a model presented in Figure 5.6 is adopted for sensitivity study. In the proposed model, all the wind turbines with the same parameters and nearby physical locations are grouped together and represented by one integrated DFIG wind turbine model. Two integrated wind farms are modeled with different parameter settings, and are connected to two remote systems through transmission lines with different length. In the reduced model shown in Figure 5.6, each wind farm model represents the integration of a certain number of DFIG-based wind turbines of the same type (with exactly the same parameters), and different wind farm models have different wind generator parameters. Meanwhile, different wind farm models also have different number of wind turbines. Besides, two remote power systems are modeled in the proposed system structure. In power flow analysis, one of the remote systems is modeled as slack bus, representing a remote generating area, the other one is modeled as PV bus with negative real power injection to the network, representing a remote loading area, and the wind farms are modeled as PQ buses with positive power injections to the network, representing wind-integrated generating area, which agree with the typical operating condition settings in China Jibei Power Grid. In order to study the impact of physical locations of compensation devices, the series capacitor and shunt capacitor/inductor are modeled at both low-voltage side and high-voltage side of the system. Applying the proposed model in Figure 5.6, original multi-wind-farm model is reduced to

a two-wind-farm model with integrated wind turbine parameters, while the tree structure network is preserved. This approach keeps the model simple enough for implementation and computation while still suitable for exploring the influence of wind farm spatial distribution, wind turbine parameter difference, and compensation device physical location on analysis and control results of wind-induced SSO.

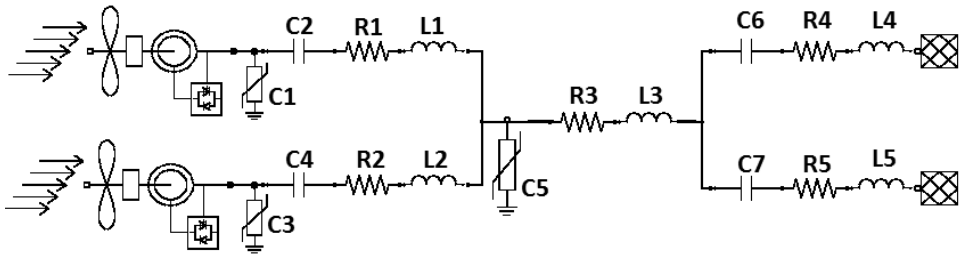

Figure 5.6: Structure of proposed two-wind-farm system model [3].

The proposed modeling and aggregation approach is generalizable towards other wind-integrated systems with radial connection structures. Through representing wind turbines with different parameter settings via different aggregated models, this modeling approach does not need to assume that all the wind turbines in the studied system have identical parameter settings, which would benefit detailed SSO analysis for multi-machine wind farms.

5.5.1.2 Parameter Sensitivity Analysis

Quantitative Parameter Sensitivity Analysis is conducted on the proposed two-wind-farm model. The following operating condition is studied: system synchronous frequency = 50 Hz; wind speed = 4 m/s; series compensation level on high-voltage side = 50 %; series compensation level on low-voltage side = 60 %; shunt compensation rated reactive power output on high-voltage side and low-voltage side = 0.1 pu; wind farm power factor

Table IX: Complex Eigenvalues and Natural Frequencies of Two-Wind-Farm System [3]

Complex Eigenvalue	Natural Frequency	Complex Eigenvalue	Natural Frequency
$-150.56\pm10266i$	1633.9 Hz	$-163.5\pm9626i$	1532.0 Hz
$-59.19\pm5817i$	925.8 Hz	$-12.27\pm4601i$	732.3 Hz
$-12.14\pm5229i$	832.2 Hz	$-83.03\pm5190i$	826.0 Hz
$-19.29\pm549.9i$	87.5 Hz	$-11.60\pm541.6i$	86.2 Hz
$-11.77\pm536i$	85.3 Hz	$-16.47\pm314.4i$	50.0 Hz
$-3.43\pm314.5i$	50.1 Hz	$0\pm314.2i$	50.0 Hz
$-18.64\pm78.8i$	12.5 Hz	$3.22\pm102.3i$	16.3 Hz
$-11.77\pm92.33i$	14.7 Hz	$-5.65\pm89.31i$	14.2 Hz
$0.97\pm101.8i$	16.2 Hz	$0\pm6.29i$	1.00 Hz
$0\pm2.57i$	0.41 Hz		

= 0.98. Under this operating condition, 19 oscillation modes are found and listed in Table IX, where 7 of them are within sub-synchronous frequency range. From Table IX, it can be seen that the 16.3 Hz and 16.2 Hz oscillation modes are undamped with positive real parts of the eigenvalues. Since their frequencies are within the frequency range of common wind farm SSOs, this suggests that there could be potential SSO events in this system under the studied operating condition.

Eigenvalue sensitivities corresponding to the 16.3 Hz and 16.2 Hz unstable oscillation modes with respect to system parameters and inputs are presented in Figure 5.7 and Figure 5.8, respectively. It could be seen that the 16.3 Hz oscillation mode is highly sensitive to parameters of DFIG subsystem of the first wind farm, converter controller subsystem of the first wind farm, as well as network subsystem. Also, the 16.2 Hz oscillation mode is highly sensitive to parameters of DFIG subsystem of the second wind farm, converter controller subsystem of the second wind farm, as well as network subsystem. Furthermore,

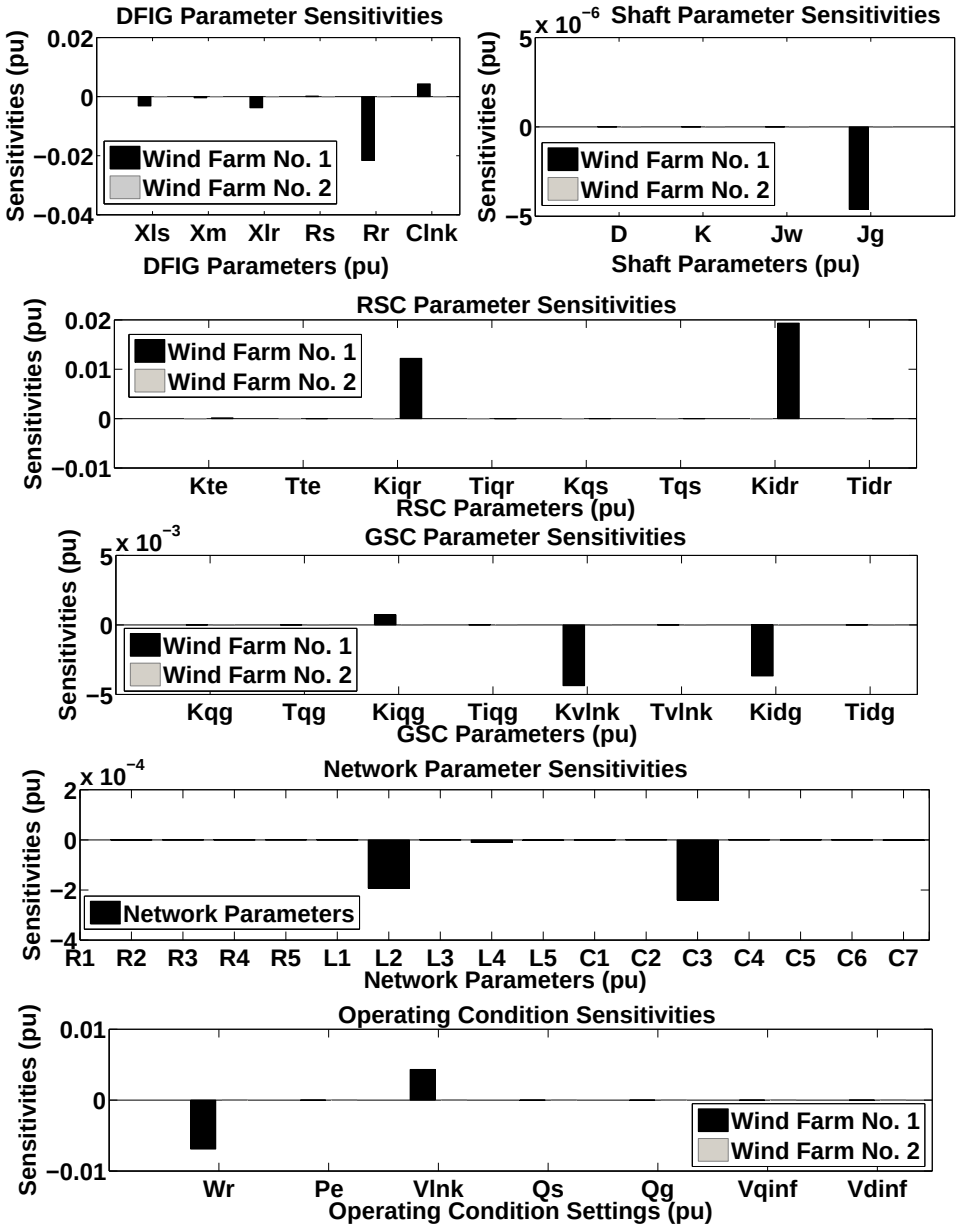

Figure 5.7: Eigenvalue sensitivities of two-wind-farm system (16.3 Hz mode) [3].

DFIG rotor speed, which has one-to-one relationship with wind speed, has a significant impact on both oscillation modes. Parameters with the highest sensitivity lies in DFIG converter controllers of the two wind farms. The shunt compensation is shown not to have a significant influence on this oscillation mode compared with other network parameters.

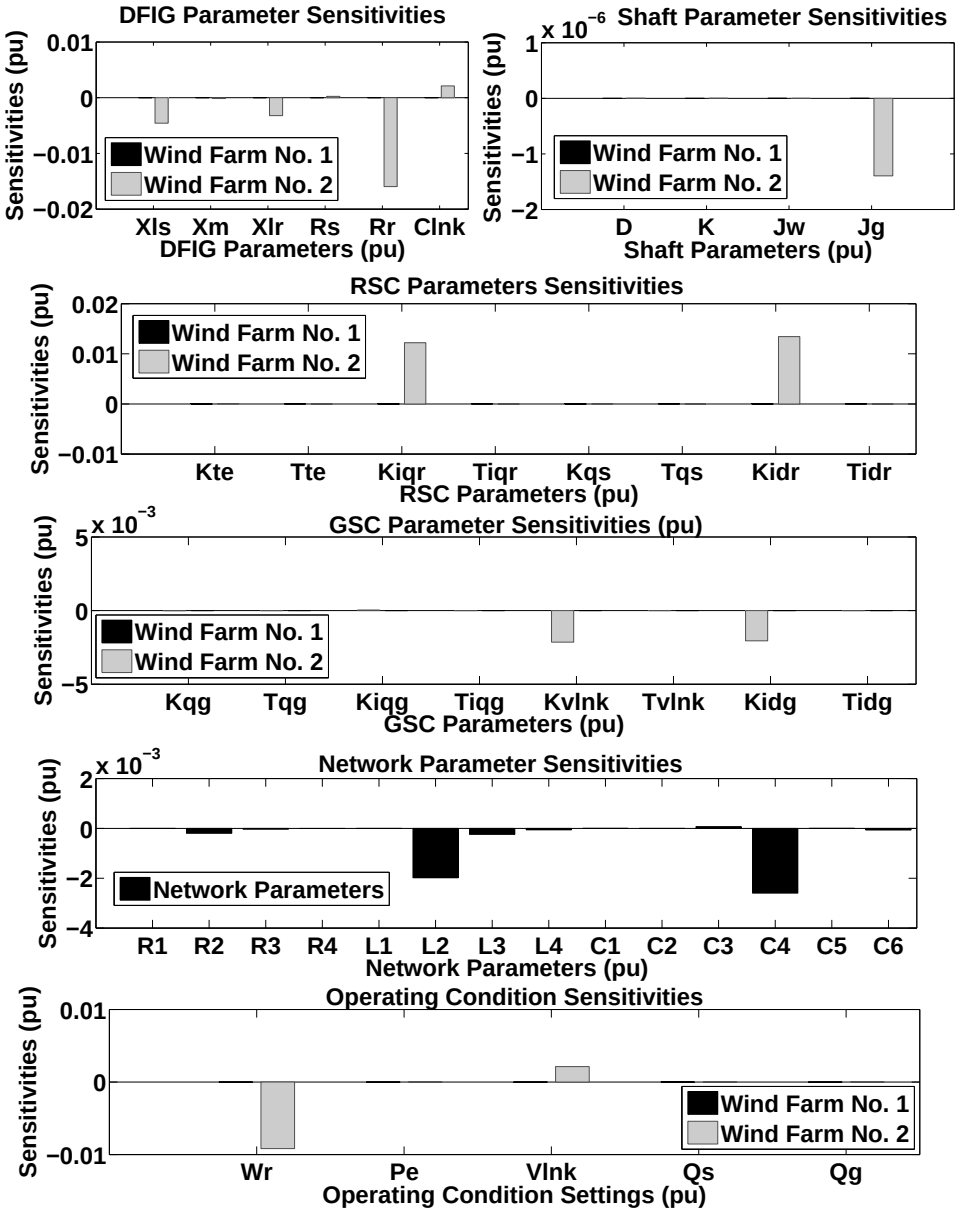

Figure 5.8: Eigenvalue sensitivities of two-wind-farm system (16.2 Hz mode) [3].

Although different parameters are assigned to the wind farms, SSO modes of the two wind farms tend to be similar. Moreover, according to the sensitivity results, no SSO mode is shown to have strong correlation with both wind farms, indicating weak coupling among spatially-distributed wind turbines when they have different sets of parameters. On the

other hand, network components located close to wind farms has higher impact on the unstable oscillation modes compared with those located far from the wind farms.

The above sensitivity results are verified through a parameter adjustment process. The converter controller gains K_{Iqr} and K_{Idr} of the first wind farm are both tuned from 0.002 to 0.001, and controller gains K_{Iqr} and K_{Idr} of the second wind farm are both tuned from 0.02 to 0.01. Simulation and FFT results are presented in Figure 5.9 and Figure 5.10, which show that both unstable oscillations are effectively damped out after control. It can be seen from Figure 5.9 and Figure 5.10, that although the coupling among wind farms with different parameters tends to be weak on the studied SSO modes, the two wind farm with different parameters still behave coherently in both un-damped and damped SSO. These phenomena could be caused by the similar oscillation frequencies of the two unstable SSO modes in the studied system. Although the parameters of the two wind farms are different, under the same operating condition including wind speed condition, the two unstable oscillation modes have very similar oscillation frequencies. Besides, it has been shown in [56, 59] that the damping of SSO modes are highly related to wind speed, which can be directly mapped to DFIG rotor speed, and it has been observed in China Jibei Power Grid that the wind farm SSO tends to happen under particular wind speed condition, therefore, it is reasonable when both wind farms experience the same wind speed, the oscillation phenomena of both wind farms tend to behave coherently with very similar oscillation frequencies and different oscillation magnitudes.

5.5.2 Practical Five-Wind-Farm System Model

In order to further study the coupling relationship among wind turbines with the same or different parameters, the above two-wind-farm model is extended to a five-wind-farm model with five integrated wind farm models connected to series-compensated transmission lines. Modal analysis is performed on the following two scenarios: 1) the five wind

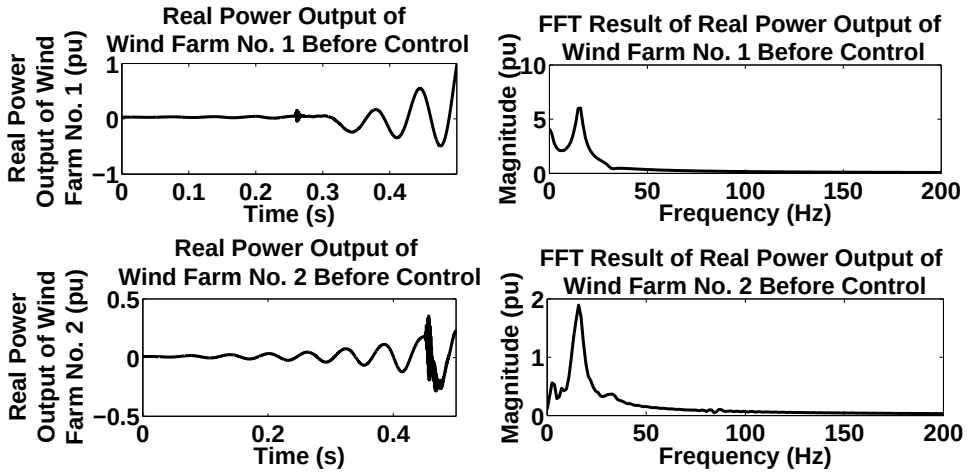

Figure 5.9: Simulation and FFT results for two-wind-farm system before mitigation control [3].

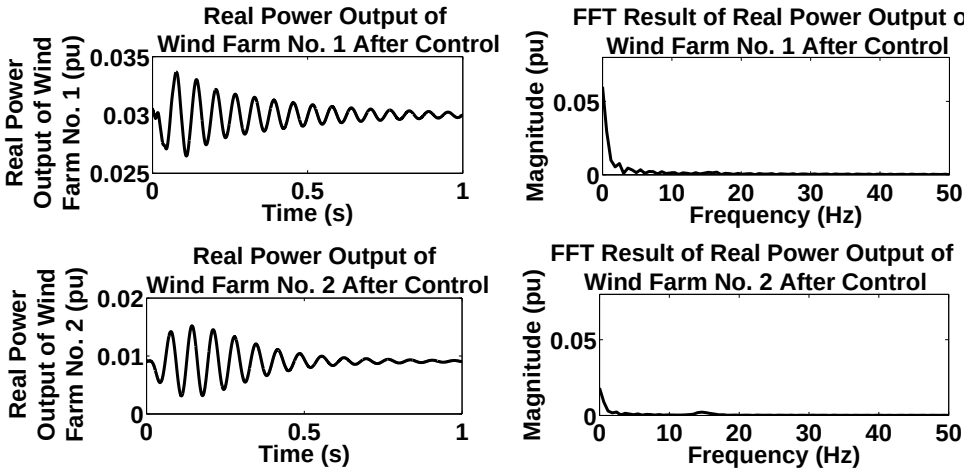

Figure 5.10: Simulation and FFT results for two-wind-farm system after mitigation control [3].

farms have different parameters and number of wind turbines; 2) the five wind farms have the same parameters and number of wind turbines. Under the above situations, eigenvalues corresponding to the five unstable SSO modes are shown in Table X and Table XI, respectively. The participation factors of wind farm state variables to the five unstable oscillation

Table X: Unstable SSO Modes of Five-Wind-Farm Model with The Same Wind Farm Parameters [3]

Complex Eigenvalue	Natural Frequency	Complex Eigenvalue	Natural Frequency
$10.67 \pm 100.78i$	16.04 Hz	$10.67 \pm 100.78i$	16.04 Hz
$10.67 \pm 100.78i$	16.04 Hz	$10.67 \pm 100.78i$	16.04 Hz
$10.67 \pm 100.78i$	16.04 Hz		

Table XI: Unstable SSO Modes of Five-Wind-Farm Model with Different Wind Farm Parameters [3]

Complex Eigenvalue	Natural Frequency	Complex Eigenvalue	Natural Frequency
$25.92 \pm 93.60i$	14.90 Hz	$51.64 \pm 82.02i$	13.05 Hz
38.70 ± 87.62	13.94 Hz	$10.67 \pm 100.77i$	16.04 Hz
12.88 ± 98.03	15.60 Hz		

modes are presented in Figure 5.11 and Figure 5.12, respectively.The modal analysis results are summarized as follows: i) in the test case where all the wind farms have different parameters, the five unstable oscillation modes have similar oscillation frequencies, each mode is highly relevant to state variables of one wind farm; ii) in the test case where all the wind farms have the same parameters, the five unstable oscillation modes have the same oscillation frequencies, each mode is highly relevant to state variables of more than one wind farm. This indicates that when the wind farms have different parameters, their coupling on the studied sub-synchronous modes tends to be weak, while when the wind farm have the same parameters, their coupling on the studied sub-synchronous modes tends to be strong.

The summarizing observation suggests that, for wind turbines with identical set of pa-

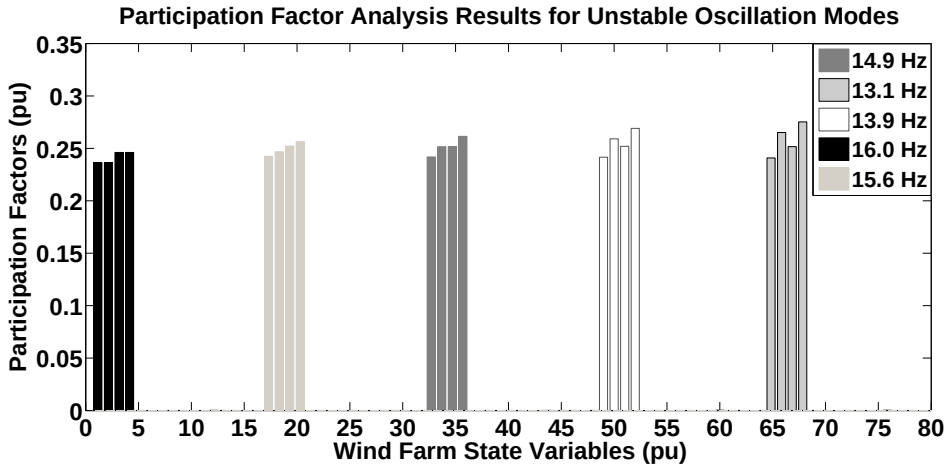

Figure 5.11: Participation factors of wind farm states to SSO modes when wind farms have different parameters (for 5-wind-farm system).Note: in this figure, States 1-16 belong to Wind Farm No. 1; States 2-32 belong to Wind Farm No. 2; States 32-48 belong to Wind Farm No. 3, States 48-64 belong to Wind Farm No. 4, States 64-80 belong to Wind Farm No. 5 [3].

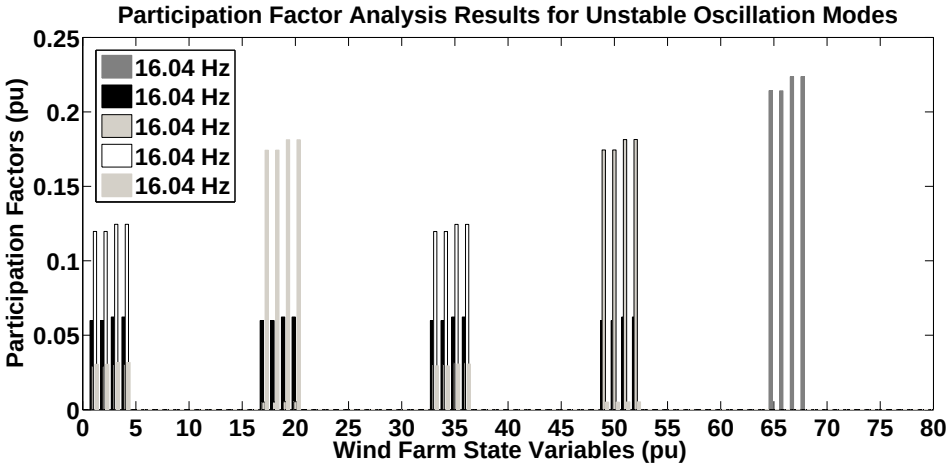

Figure 5.12: Participation factors of wind farm states to SSO modes when wind farms have the same parameters (for 5-wind-farm system). Note: in this figure, States 1-16 belong to Wind Farm No. 1; States 2-32 belong to Wind Farm No. 2; States 32-48 belong to Wind Farm No. 3, States 48-64 belong to Wind Farm No. 4, States 64-80 belong to Wind Farm No. 5 [3].

rameters and identical operating condition, their SSO modes tend to be strongly coupled with each other; for wind turbines with different set of parameters and identical operating condition, their SSO modes tend to be weakly coupled with each other. In both situations, all the wind turbines tend to oscillate in a coherent way with similar oscillation frequencies.

5.6 Section Conclusion

This section proposes modal analysis as a means of investigating the impact of wind farm spatial distribution on recently-discovered SSO incidents in wind-integrated power systems. Multi-machine wind farm models derived from practical wind-integrated system are proposed and integrated with the network model. Based on the integrated model, a thorough sensitivity analysis of SSO with respect to parameters of spatially-distributed wind turbines, wind conditions, etc. is conducted. Groups of parameters (e.g., DFIG controller parameters and network series compensation level) are shown to have significant impact on the occurrence of SSO. According to the analysis and simulation results, for wind turbines with identical set of parameters and identical operating condition, their SSO modes tend to be strongly coupled with each other; for wind turbines with different set of parameters and identical operating condition, their SSO modes tend to be weakly coupled with each other. In both situations, all the wind turbines tend to oscillate in a coherent way with similar oscillation frequencies.

Based on these findings, the system operators could introduce more robust system standards when faced with many dispersed wind farms. Future work could also be done in designing countermeasures to mitigate such SSO in real-time operations.

6. CONCLUSIONS

6.1 Summary

In this dissertation, a framework is proposed for the enhanced planning and operations of modern power systems in the data-rich environment. Motivated by the major challenges involved in data collection and pre-processing, situational awareness, decision making process, as well as post-event analysis, this dissertation presents combined physics-based and data-driven analytics to solve the problems in modern power system planning and operations. The framework is introduced through a series of concrete examples for solving the above critical challenges using combined physics-based and data-driven techniques. To be specific, the following examples are presented to demonstrate the proposed frameworks:

1) A data collection and pre-processing platform: a purely data-driven approach is proposed to detect bad metering data in the PMU monitoring systems, and ensure the overall PMU data quality.

2) A situational awareness platform: a physics-based voltage stability assessment is presented to improve the situational awareness of system voltage instabilities.

3) A decision making platform: a combined physics-based and data-driven framework is proposed to support the decision making process of PMU-based power plant model validation.

4) A post-event analysis platform: a physics-based post-event analysis is presented to identify the root causes of the sub-synchronous oscillations induced by wind farm integrations.

The major contributions of this dissertation are suggested as follows:

1. A framework is proposed for integrating dynamic data into large physical systems modeled using differential-algebraic equations.

2. A data-driven algorithm is developed for improving the data quality of PMU measurement systems.

3. A comprehensive approach is proposed for exploring the power ïňĆow solvability problem for the purpose of strengthened voltage stability monitoring of modern power grids.

4. A decision support framework is proposed for automating the entire process of PMU-based power plant model validation, with the capability of batch power plant model validation and automatic diagnosis of power plant modeling errors.

5. An enhanced analysis is presented for identifying the root causes of the wind-farm-induced sub-synchronous oscillations.

6.2 Future Work

In the future work, the following two areas would be explored: 1) applying deep learning techniques for the feature extraction task of the PPMV diagnosis framework; 2) improving the power flow solution boundary visualization using advanced optimization techniques.

In the long run, the proposed framework would be expanded in the following aspects: 1) applying data-driven approaches for improving the real-time/predictive situational awareness of large-scale power systems; 2) developing advanced computational tools for the security assessment of power systems with deep renewable penetration; 3) exploring modern optimization techniques for enhancing the electricity market operations.

REFERENCES

[1] M. Wu and L. Xie, "Online detection of low-quality synchrophasor measurements: A data-driven approach," *IEEE Transactions on Power Systems*, vol. 32, pp. 2817–2827, July 2017.

[2] M. Wu, W. Huang, Q. Zhang, X. Luo, S. Maslennikov, and E. Litvinov, "Power plant model verification at iso new england," in *2017 IEEE PES General Meetings*, (Chicago, IL), July 2017.

[3] M. Wu, L. Xie, L. Cheng, and R. Sun, "A study on the impact of wind farm spatial distribution on power system sub-synchronous oscillations," *IEEE Transactions on Power Systems*, vol. 31, pp. 2154–2162, May 2016.

[4] M. Wu and L. Xie, "Calculating steady-state operating conditions for dfig-based wind turbines," *IEEE Transactions on Sustainable Energy*, (accepted, to appear).

[5] E. Ghahremani and I. Kamwa, "Dynamic state estimation in power system by applying the extended kalman filter with unknown inputs to phasor measurements," *Power Systems, IEEE Transactions on*, vol. 26, pp. 2556–2566, Nov 2011.

[6] D. Kosterev, "Hydro turbine-governor model validation in pacific northwest," *Power Systems, IEEE Transactions on*, vol. 19, pp. 1144–1149, May 2004.

[7] I. Kamwa, R. Grondin, and Y. Hebert, "Wide-area measurement based stabilizing control of large power systems-a decentralized/hierarchical approach," *Power Systems, IEEE Transactions on*, vol. 16, pp. 136–153, Feb 2001.

[8] J. Bertsch, C. Carnal, D. Karlson, J. McDaniel, and K. Vu, "Wide-area protection and power system utilization," *Proceedings of the IEEE*, vol. 93, pp. 997–1003, May 2005.

[9] California ISO, “Five year synchrophasor plan,” tech. rep., California ISO, Nov 2011.

[10] W. Qi, “Comparison of differences between SCADA and WAMS real-time data in dispatch center,” in *12th International Workshops on Electric Power Control Centers (EPCC)*, Jun 2013.

[11] S. Ghiocel, J. Chow, G. Stefopoulos, B. Fardanesh, D. Maragal, B. Blanchard, M. Razanousky, and D. Bertagnolli, “Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring,” *Power Systems, IEEE Transactions on*, vol. 29, pp. 881–888, March 2014.

[12] K. Jones, A. Pal, and J. Thorp, “Methodology for performing synchrophasor data conditioning and validation,” *Power Systems, IEEE Transactions on*, vol. 30, pp. 1121–1130, May 2015.

[13] K. Martin, “Synchrophasor data diagnostics: detection & resolution of data problems for operations and analysis,” in *Electric Power Group Webinar Series*, Jan 2014.

[14] Q. Zhang and V. Venkatasubramanian, “Synchrophasor time skew: Formulation, detection and correction,” in *North American Power Symposium (NAPS), 2014*, pp. 1–6, Sept 2014.

[15] S. Dutta and T. Overbye, “Information processing and visualization of power system wide area time varying data,” in *Computational Intelligence Applications In Smart Grid (CIASG), 2013 IEEE Symposium on*, pp. 6–12, April 2013.

[16] M. Wang, J. Chow, P. Gao, X. Jiang, Y. Xia, S. Ghiocel, B. Fardanesh, G. Stefopolous, Y. Kokai, N. Saito, and M. Razanousky, “A low-rank matrix approach for the analysis of large amounts of power system synchrophasor data,” in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pp. 2637–2644, Jan 2015.

[17] L. Liu, M. Esmalifalak, Q. Ding, V. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *Smart Grid, IEEE Transactions on*, vol. 5, pp. 612–621, March 2014.

[18] M. Wu and L. Xie, "Online identification of bad synchrophasor measurements via spatio-temporal correlations," in *2016 Power Systems Computation Conference (PSCC)*, pp. 1–7, June 2016.

[19] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, (New York, NY, USA), pp. 93–104, ACM, 2000.

[20] P. Kansal and A. Bose, "Bandwidth and latency requirements for smart transmission grid applications," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1344–1352, Sept 2012.

[21] S.-H. Huang, J. Schmall, J. Conto, J. Adams, Y. Zhang, and C. Carter, "Voltage control challenges on weak grids with high penetration of wind generation: Ercot experience," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–7, July 2012.

[22] C. L. DeMarco and T. J. Overbye, "Low voltage power flow solutions and their role in exit time based security measures for voltage collapse," in *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 2127–2131 vol.3, Dec 1988.

[23] I. Dobson, "Observations on the geometry of saddle node bifurcation and voltage collapse in electrical power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 39, pp. 240–243, Mar 1992.

[24] I. Dobson and H.-D. Chiang, “Towards a theory of voltage collapse in electric power systems,” *Systems & Control Letters*, vol. 13, no. 3, pp. 253 – 262, 1989.

[25] H.-D. Chiang, A. J. Flueck, K. S. Shah, and N. Balu, “Cpflow: a practical tool for tracing power system steady-state stationary behavior due to load and generation variations,” *IEEE Transactions on Power Systems*, vol. 10, pp. 623–634, May 1995.

[26] I. A. Hiskens and R. J. Davy, “Exploring the power flow solution space boundary,” *IEEE Transactions on Power Systems*, vol. 16, pp. 389–395, Aug 2001.

[27] D. K. Molzahn, I. A. Hiskens, and B. C. Lesieutre, “Calculation of voltage stability margins and certification of power flow insolvability using second-order cone programming,” in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 2307–2316, Jan 2016.

[28] I. Dobson and L. Lu, “New methods for computing a closest saddle node bifurcation and worst case load power margin for voltage collapse,” *IEEE Transactions on Power Systems*, vol. 8, pp. 905–913, Aug 1993.

[29] T. Jung, K. Kim, and F. Alvarado, “A marginal analysis of the voltage stability with load variations,” in *Power Systems Computation Conference*, 1990.

[30] J. B. Lasserre, “Global optimization with polynomials and the problem of moments,” *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[31] P. A. Parrilo, “Semidefinite programming relaxations for semialgebraic problems,” *Mathematical Programming*, vol. 96, no. 2, pp. 293–320, 2003.

[32] M. Kojima, “Sos and sdp relaxation of polynomial optimization problems,” November 2010.

[33] NERC, “Reliability guideline - power plant dynamic model verification using pmus.”

[34] NERC, “Standard mod-026-1 - verification of models and data for generator excitation control system or plant volt/var control functions.”

[35] NERC, “Standard mod-027-1 - verification of models and data for turbine/governor and load control or active power/frequency.”

[36] NERC, “Standard mod-033-1 - steady-state and dynamic system model validation.”

[37] Y. Li, R. Diao, R. Huang, P. Etingov, X. Li, Z. Huang, S. Wang, J. Sanchez-Gasca, B. Thomas, M. Parashar, G. Pai, S. Kincic, and A. Ning, “An innovative software tool suite for power plant model validation and parameter calibration using pmu measurements,” in *2017 IEEE Power and Energy Society General Meeting*, (Chicago, IL), August 2017.

[38] A. R. Borden and B. C. Lesieutre, “Variable projection method for power system modal identification,” *IEEE Transactions on Power Systems*, vol. 29, pp. 2613–2620, Nov 2014.

[39] Y. Cheng, W. J. Lee, S. H. Huang, and J. Adams, “Dynamic parameter identification of generators for smart grid development,” in *2011 IEEE Power and Energy Society General Meeting*, pp. 1–7, July 2011.

[40] Powertech Labs Inc., *User Manual of PSAT Power flow & Short-circuit Analysis Tool*, April 2011.

[41] Powertech Labs Inc., *User Manual of TSAT Transient Security Assessment Tool*, April 2011.

[42] P. Sauer and M. Pai, *Power System Dynamics and Stability*. Stipes Publishing L.L.C., 2006.

[43] E. J. Keogh and M. J. Pazzani, *Derivative Dynamic Time Warping*, pp. 1–11.

[44] R. Bellman and R. Kalaba, "On adaptive control processes," *IRE Transactions on Automatic Control*, vol. 4, pp. 1–9, November 1958.

[45] C. Myers, L. Rabiner, and A. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, pp. 623–635, Dec 1980.

[46] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, Feb 1978.

[47] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[48] L. Xie, P. M. S. Carvalho, L. A. F. M. Ferreira, J. Liu, B. Krogh, N. Popli, and M. Ilic, "Wind integration in power systems: Operational challenges and possible solutions," *Proceedings of the IEEE*, vol. 99, pp. 214–232, Jan 2011.

[49] P. Belkin, "Event of 10-22-09," in *CREZ Technical Conference, January*, Jan 2010.

[50] Y. H. Wan, "Synchronized phasor data for analyzing wind power plant dynamic behavior and model validation," Tech. Rep. NREL/TP-5500-57342, National Renewable Energy Laboratory, CO, Jan 2013.

[51] M. Wu, R. Sun, L. Cheng, and L. Xie, "Parameter sensitivity analysis for subsybchronous control interactions in wind-integrated power systems," in *CIGRE Grid of the Future Symposium*, (Houston, TX), Oct 2014.

[52] L. Wang, X. Xie, Q. Jiang, H. Liu, Y. Li, and H. Liu, "Investigation of SSR in practical DFIG-based wind farms connected to a series-compensated power system," *Power Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–8, 2014.

[53] P. M. Anderson, B. L. Agrawal, and J. E. Van Ness, *Subsynchronous Resonance in Power Systems*. NY: IEEE Press, 1990.

[54] ABB Study Team, "ERCOT CREZ reactive power compensation study," Tech. Rep. E3800-PR-00, ABB Inc., NC, Nov 2010.

[55] A. P. Masssimo Bongiorno and E. Agneholm, "The impact of wind farms on subsynchronous resonance in power systems," tech. rep., Elforsk, Swedish, April 2011.

[56] L. Fan, C. Zhu, Z. Miao, and M. Hu, "Modal analysis of a DFIG-based wind farm interfaced with a series compensated network," *Energy Conversion, IEEE Transactions on*, vol. 26, pp. 1010–1020, Dec 2011.

[57] R. Varma, S. Auddy, and Y. Semsedini, "Mitigation of subsynchronous resonance in a series-compensated wind farm using facts controllers," *Power Delivery, IEEE Transactions on*, vol. 23, pp. 1645–1654, July 2008.

[58] M. El-Moursi, B. Bak-Jensen, and M. Abdel-Rahman, "Novel statcom controller for mitigating ssr and damping power system oscillations in a series compensated wind park," *Power Electronics, IEEE Transactions on*, vol. 25, pp. 429–441, Feb 2010.

[59] L. Fan, R. Kavasseri, Z. L. Miao, and C. Zhu, "Modeling of DFIG-based wind farms for SSR analysis," *Power Delivery, IEEE Transactions on*, vol. 25, pp. 2073–2082, Oct 2010.

[60] D. Suriyaarachchi, U. Annakkage, C. Karawita, and D. Jacobson, "A procedure to study sub-synchronous interactions in wind integrated power systems," *Power Systems, IEEE Transactions on*, vol. 28, pp. 377–384, Feb 2013.

[61] A. Ostadi, A. Yazdani, and R. Varma, "Modeling and stability analysis of a DFIG-based wind-power generator interfaced with a series-compensated line," *Power Delivery, IEEE Transactions on*, vol. 24, pp. 1504–1514, July 2009.

[62] R. Wies, J. Pierre, and D. Trudnowski, "Use of arma block processing for estimating stationary low-frequency electromechanical modes of power systems," *Power Systems, IEEE Transactions on*, vol. 18, pp. 167–173, Feb 2003.

[63] N. Zhou, D. Trudnowski, J. Pierre, and W. Mittelstadt, "Electromechanical mode online estimation using regularized robust RLS methods," *Power Systems, IEEE Transactions on*, vol. 23, pp. 1670–1680, Nov 2008.

[64] H. Ghasemi and C. Canizares, "On-line damping torque estimation and oscillatory stability margin prediction," *Power Systems, IEEE Transactions on*, vol. 22, pp. 667–674, May 2007.

[65] D. Trudnowski, "Estimating electromechanical mode shape from synchrophasor measurements," *Power Systems, IEEE Transactions on*, vol. 23, pp. 1188–1195, Aug 2008.

[66] D. Trudnowski and J. Pierre, "Overview of algorithms for estimating swing modes from measured responses," in *Power Energy Society General Meeting, 2009. PES '09. IEEE*, pp. 1–8, July 2009.

[67] L. Dosiek, D. Trudnowski, and J. Pierre, "New algorithms for mode shape estimation using measured data," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pp. 1–8, July 2008.

[68] Z. Lubosny, *Wind Turbine Operation in Electric Power Systems: Advanced Modeling*. Berlin, Germany: Springer-Verlag, 2003.

[69] P. C. Krause, O. Wasynczuk, and S. D. Sudhoff, *Analysis of Electric Machinery*. NY: IEEE Press, 1995.

[70] P. Ledesma and J. Usaola, “Doubly fed induction generator model for transient stability analysis,” *Energy Conversion, IEEE Transactions on*, vol. 20, pp. 388–397, June 2005.

[71] J. Padron and A. Lorenzo, “Calculating steady-state operating conditions for doubly-fed induction generator wind turbines,” *Power Systems, IEEE Transactions on*, vol. 25, pp. 922–928, May 2010.

[72] L. Holdsworth, X. Wu, J. Ekanayake, and N. Jenkins, “Direct solution method for initialising doubly-fed induction wind turbines in power system dynamic models,” *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 150, pp. 334–342, May 2003.

[73] N. W. Miller, W. W. Price, and J. Sanchez-Gasca, “Dynamic modeling of ge 1.5 and 3.6 wind turbine-generators, geąłpower system energy consulting,” *Tech. Rep. Version 3.0*, 2003.

[74] J. J. Thomsen, *Vibrations and stability: advanced theory, analysis, and tools*. Springer Science & Business Media, 2003.

[75] WECC, “Wecc second generation wind turbine models,” January 2014.

[76] WECC, “Wecc type 3 wind turbine generator model Íc phase ii,” January 2014.

[77] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical mathematics*, vol. 37. Springer, 2007.