POWER SYSTEM ONLINE STABILITY ASSESSMENT USING SYNCHROPHASOR DATA MINING

A Dissertation

by

CE ZHENG

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Mladen Kezunovic
Committee Members,	Shankar P. Bhattacharyya
	Garng M. Huang
	Yu Ding
Head of Department,	Chanan Singh

May 2013

Major Subject: Electrical Engineering

ABSTRACT

Traditional power system stability assessment based on full model computation shows its drawbacks in real-time applications where fast variations are present at both demand side and supply side.

This work presents the use of data mining techniques, in particular the Decision Trees (DTs), for fast evaluation of power system oscillatory stability and voltage stability from synchrophasor measurements. A regression tree-based approach is proposed to predict the stability margins. Modal analysis and continuation power flow are the tools used to build the knowledge base for off-line DT training. Corresponding metrics include the damping ratio of critical electromechanical oscillation mode and MW-distance to the voltage instability region. Classification trees are used to group an operating point into predefined stability state based on the value of corresponding stability indicator. A novel methodology for knowledge base creation has been elaborated to assure practical and sufficient training data. Encouraging results are obtained through performance examination.

The robustness of the proposed predictor to measurement errors and system topological variations is analyzed. A scheme has been proposed to tackle the problem of when and how to update the data mining tool for seamless online stability monitoring. The optimal placement for the phasor measurement units (PMU) based on the importance of DT variables is suggested.

A measurement-based voltage stability index is proposed and evaluated using field PMU measurements. It is later revised to evaluate the impact of wind generation on distribution system voltage stability.

Next, a new data mining tool, the Probabilistic Collocation Method (PCM), is presented as a computationally efficient method to conduct the uncertainty analysis. As compared with the traditional Monte Carlo simulation method, the collocation method could provide a quite accurate approximation with fewer simulation runs.

Finally, we show how to overcome the disadvantages of mode meters and ringdown analyzers by using DTs to directly map synchrophasor measurements to predefined oscillatory stability states. The proposed measurement-based approach is examined using synthetic data from simulations on IEEE test systems, and PMU measurements collected from field substations. Results indicate that the proposed method complements the traditional model-based approach, enhancing situational awareness of control center operators in real time stability monitoring and control.

DEDICATION

This research endeavor is dedicated to

my wife, Yimai Dong, and to my son, Kevin Zheng.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my graduate studies advisor, Dr. Mladen Kezunovic, for his guidance and support throughout the course of this research. His understanding and patience are indispensable to the completion of this work. I also would like to thank Dr. Huang, Dr. Bhattacharyya, Dr. Lively, Dr. Ding and Dr. Cui for their valuable comments to this study and serving as my committee members since 2009.

Thanks to my friends and colleagues, Dr. Chengzong Pang, Dr. Jinfeng Ren, Mr. Yufan Guan, Ms. Biljana Matic Cuka, Ms. Papiya Dutta, Dr. Vuk Malbasa, Dr. Gurunath Gurrula, Mr. Po-Chen Chen, Ms. Qin Yan, and Ms. Bei Zhang for their help and support throughout my graduate study. I also want to extend my gratitude to the department faculty and staff, in particular Ms. Tammy Carda and Ms. Nancy Reichart, for making my time at Texas A&M University a great experience.

My research was mainly funded by financial resources from two projects. One is funded by U.S. Department of Energy (DOE): "Synchronized Sampling Uses for Real Time Monitoring and Control". The other is funded by NSF I/UCRC Power System Engineering Research Center (PSERC): "Data Mining to Characterize Signatures of Impending System Events or Performance Using PMU Measurements". I would like to acknowledge the financial support from all the sponsors.

Finally, I wish to thank my wife, my father and my mother for their perpetual love, patience, encouragement and support.

NOMENCLATURE

SCADA	Supervisory Control and Data Acquisition
PMU	Phasor Measurement Unit
WAMS	Wide Area Measurement System
IED	Intelligent Electronic Device
EMS	Energy Management System
RTU	Remote Terminal Unit
TxDOT	Texas Department of Transportation
PCM	Probabilistic Collocation Method
DT	Decision Tree
OSM	Oscillatory Stability Margin
DR	Damping Ratio
VSM	Voltage Stability Margin
CART	Classification and Regression Tree
OSM-RT	Regression Tree for Oscillatory Stability Margin Prediction
VSM-RT	Regression Tree for Voltage Stability Margin Prediction
CPF	Continuation Power Flow
VSI	Voltage Stability Index
ARMA	Autoregressive Moving Average

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1 Problem Statement

1.1.1 Enhancing EMS with Synchrophasor Measurements

Nowadays, power system is featured by numerous data collecting points and a variety of measurement devices. Beyond the traditional Supervisory Control and Data Acquisition (SCADA) system composed of RTUs, most substations are equipped with Intelligent Electronic Devices (IEDs) performing the task of data collection [1]. These modern IEDs can record and store a huge amount of data with a periodicity depending upon the intended purpose of the device: Digital Protective Relays and Digital Fault Recorders (DPRs and DFRs) capturing data during fault occurrence (termed as non-operational data), and Phasor Measurement Units (PMUs) capturing continuous time-synchronized phasors (termed as situational awareness data).

With the advancement of measurement techniques and growing availability of information captured by newly emerged IEDs such as PMUs, it is imperative to integrate the various data sources with traditional SCADA/Energy Management System (EMS) structure. Possible solutions for data integration and its applications in power system need to be addressed [2]. For example, present-day System Integrity Protection Schemes (SIPS) against voltage instability mostly rely on the detection of low voltage conditions. The under-voltage criteria allows simple and possibly distributed SIPS, for instance for load shedding. However, it essentially relies on the observation of already degraded operating conditions. The challenge is thus to demonstrate that synchrophasor-based

applications can offer better anticipation capabilities by detecting the inception of instability rather than its consequences.

SCADA system solutions have been used to provide real-time information about power system states since the late sixties. The implementation of SCADA solutions improved the performance of EMS functions. Although the newly emerged multifunctional IEDs are not standardized regarding the functions they perform, they indeed are necessary additions to the data recording infrastructure needed for performing a comprehensive analysis of substation equipment operation. The merging of various substation data with the data from SCADA solutions to enhance EMS functions has not yet been explored adequately.

Figure 1 shows the state-of-the-art data acquisition structure and its possible implementation in analyzing two types of power system stability status, i.e. oscillatory stability and voltage stability. Traditionally the data used for the stability analysis in electrical utilities are obtained from the SCADA system or state estimation functions, which, as shown in Figure 1, are refreshed on a time scale from several seconds to several minutes. The SCADA measured data does not have the characteristics needed to implement the new analysis and control tools due to the lack of time-synchronized sampled waveform data. In some cases the forecasted load pattern and unit commitment dispatch are used instead of actual data to predict system performance. When a disturbance occurs and immediate controls need to be initiated, traditional stability analysis using slowly updated or forecasted data can only provide very limited decision making support.

Compared to traditional SCADA system, synchrophasor IEDs such as PMUs enable a much higher data sampling rate and provide the synchronized phasor measurements across the network [3].

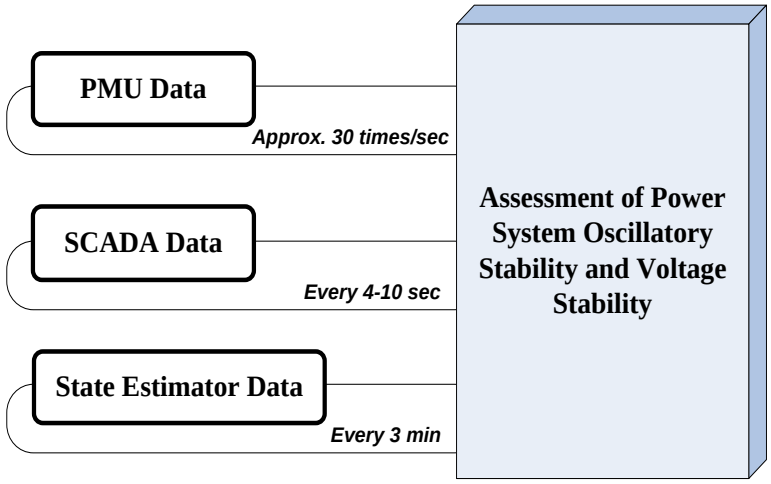


Figure 1 Power system stability analysis using data from various sources

1.1.2 Power System Stability Assessment Using Data Mining

From the control center operators' point of view, the fast assessment of power system oscillatory stability and voltage stability is of great importance for real-time operation. It is desirable that the impending system events can be immediately detected and that operators are provided with the updated information on whether or not a power system can maintain synchronism and acceptable voltage levels when subject to disturbances.

Traditionally, the method of time-domain simulation is used to analyze system stability status [4]. However two obstacles prevent the traditional method's application

in real-time monitoring and control. Firstly, the need of full system model computation makes the simulation method time-consuming. Considering the fast onset of an instability event, the traditional method may not be able to provide immediate event detection. Using a simplified system model could accelerate the simulations, but this brings concern over approximate analysis results leading to inaccurate decisions. Secondly, the data used for the stability analysis in electrical utilities are obtained from the SCADA system or state estimation functions, which are refreshed on a time scale from several seconds to several minutes. In some cases the forecasted load pattern and unit commitment dispatch are used instead of actual data to analyze system performance. When a disturbance occurs and immediate controls need to be initiated, traditional stability analysis using slowly updated or forecasted data can only provide limited decision making support.

To make the situation worse, in power system planning and on-line applications a complete model may not be readily available. This model is necessary for obtaining the linearized system description required by traditional oscillatory stability analysis [5]. Similar problems exist in the voltage stability assessment process [6]. Under such circumstances, the data mining techniques, benefiting from accurate generalization ability without detailed knowledge of all system parameters, becomes an attractive alternative.

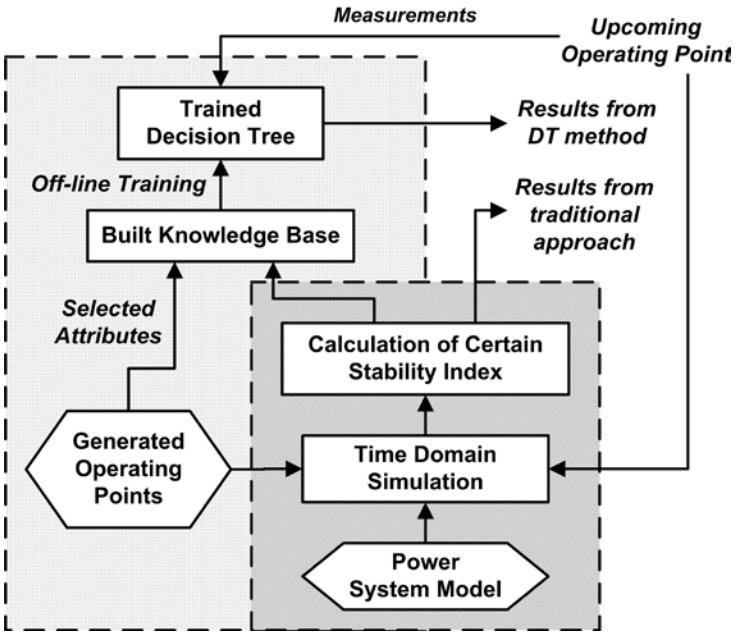


Figure 2 Comparison between conventional approach and the DT method

In particular, the relationship and difference between the conventional time-domain simulation approach and one of the data mining tools, the Decision Tree (DT) method [7], is shown in Figure 2. Compared with the traditional method, the advantage of DT method lies in its capability of fast analysis facilitated by fewer required inputs and straightforward model structure. By learning the system behavior from a known set of OPs, the DT model can predict system responses without detailed model computations. In addition, the DT method is appealing because it uses a white-box model, which makes the results easy to interpret. Based on the combination of splitting rules along a path of the tree, preventive and corrective control strategies could be formulated.

1.2 Current Research Efforts

1.2.1 Data Integration

A few studies have been reported that address this topic. In [1], the benefits of data integration and information extraction achieved by merging temporal and spatial considerations have been explored. Some other research primarily discuss how to extract desired information from field measurements and improve the quality of control center visualization tools in terms of good graphical representations [8]-[10]. In [8], the inefficiency of using numerical data to display system conditions is identified, and three key characteristics of effective graphical representations are illustrated. In [9] and [10], the concept of Geographic Data Views (GDVs) has been introduced, which uses dynamically created visualization tools to display a wider range of power system information than what is possible using the existing geographically based visualization tools. In particular, the research of [9] utilizes Global Positioning System (GPS)-based information provided by PMUs to display power system status in real-time. In some other research, high-resolution satellite images are used to provide 2-D and 3-D representations [11]-[12].

With the growth of its complexity, power systems are becoming more stressed and exposed to all kinds of disturbances. To adapt to the emerging situation, several new applications deploying integration of field measurements have been introduced both in research and practice [13]-[21]. They utilize data from SCADA RTUs, PMUs and other IEDs to analyze alarms, detect cascades and locate faults. In [8] it is shown that the merging of RTU, PMU and other substation IED data could effectively improve the

ability to detect cascades, efficiently process alarms, and accurately determine the location of faults.

1.2.2 Detecting Impending Events from Synchrophasor Measurements

In the past, the system stability was evaluated using very limited and slowly updated field measurements. With the development of data acquisition technique, especially for the time synchronized measurements at different locations, detection of the onset of instability became more accurate. Glavic and Cutsem in [22] demonstrated that synchrophasors, with the ability to capture system-wide long-term dynamics, opens new perspectives for real-time system stability monitoring. A sensitivity-based approach to compute voltage stability along the system trajectory using synchrophasor measurements was proposed in [23]. The general mathematical formulation proposed in [24] calculates the voltage stability margin by computing a Maximum Loadability Index using data from PMUs.

1.2.3 Decision Tree Approach

In the field of power systems, Wehenkel et al. first introduced the DT method to solve the transient stability assessment problems using SCADA data [25]-[26]. In [27]-[31], DTs were successively applied to assess system operational security by applying a pre-defined set of credible contingencies and enforcing an acceptable threshold criterion on system variables based on standard operating practices. Later, in [32], the system post-disturbance stability has been analyzed by DT using its fast evaluation capability. In

[33], a genetic algorithm was applied in feature selection to search for the best inputs to DT for oscillatory stability region prediction. In [34] and [35], Kamwa et al. showed that there is a trade-off between a data mining model's accuracy and its transparency. A review of literature reveals that the problem of using DT for stability margin monitoring from substation field measurements has not yet been fully explored.

The concept of decision tree comprises the Classification Tree and Regression Tree. While in previous works classification trees have been extensively studied to group an operating point (OP) into one of several pre-defined stability categories, the use of regression trees (RT) to predict the stability margin, i.e. how far the system is away from a possible instability event, has not yet been fully studied. With respect to its online use, the areas that remain unexplored include how fast the RT can process PMU measurements, how well the RT can deal with measurement errors, and how robust the RT is to the system topology changes. It is also imperative to develop a systematic approach to generating a sufficient and realistic knowledge base for off-line training of DT.

Several other data mining tools such as Neural Networks [36] and Support Vector Machines [37] have been used to evaluate the system stability status. Compared with some "black-box" tools, the DT piece-wise structure provides system operators with a clearer cause-effect relationship of how the system variables lead to the onset of an instability event. Using DTs it is possible to identify the critical variables and thresholds that need to be analyzed to gain insight into the stability margin of a system.

1.3 Proposed Research

A breakdown of the proposed research in this dissertation is as follows:

- Develop a methodology that takes use of the PMU collected synchrophasor measurements for online stability estimation and early detection of impending system instability events;
- Examine the prediction accuracy and robustness of DT for online assessment of system oscillatory stability and voltage stability status. The accuracy and efficiency between the DT and other data mining tools such as Support Vector Machine (SVM) and Neural Networks (NN) will be compared;
- The important issue of DT robustness with respect to PMU measurement errors and changes in system topology will be explored;
- Develop a methodology for optimal PMU placement. Check the performance of DT using synchrophasor measurements from a limited number of PMUs;
- Explore a measurement-based approach that directly applies data mining and signal processing techniques to field PMU measurements;
- Explore the impact of grid integration of wind generation on system stability. Investigate how data mining tools can help deal with the supply side uncertainty introduced by the stochastic nature of wind generation.

In this work both the model-based approach and measurement-based (depend on whether the system model data is used) approach will be explored. For the model-based approach, a knowledge base will be created through exhaustive simulations on known

system model parameters and then utilized to train the decision trees. For the situation where detailed model parameters are missing, a measurement-based approach will be deployed to estimate system stability status directly from the synchrophasor measurements collected at the PMU-equipped substations. In addition, the measurement-based approach is also applied to investigate the impact of wind integration on power system voltage stability. The efficacy of the measurement-based approach is going to be tested using field PMU measurements.

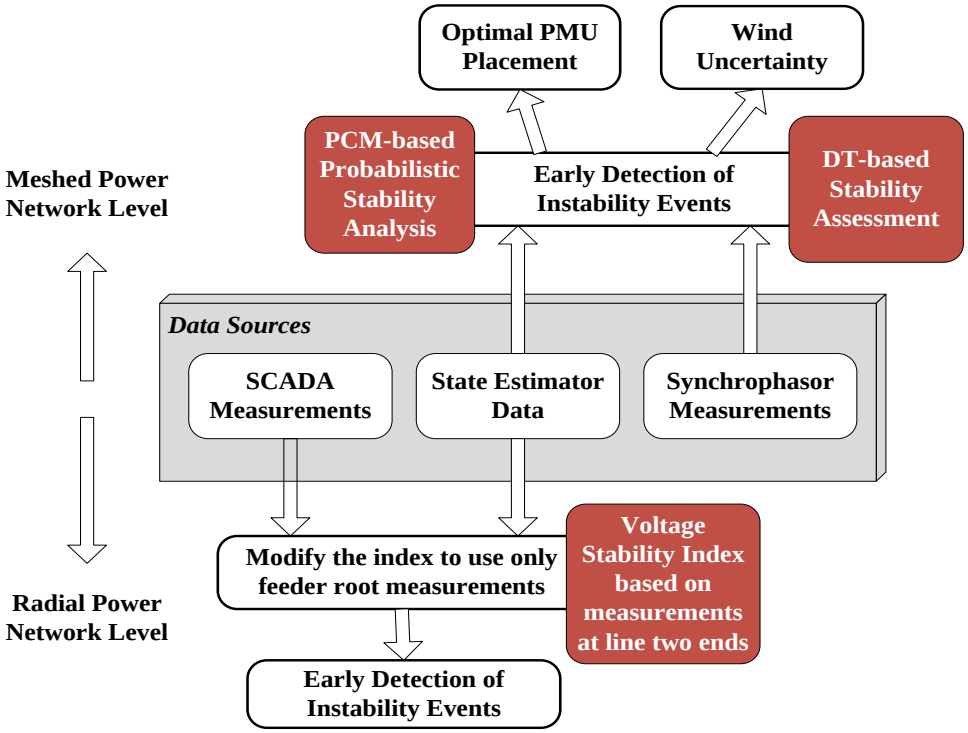


Figure 3 Proposed data integration and research framework

The proposed data integration and research framework is shown in Figure 3. The data mining techniques will be adopted for meshed network stability assessment from PMU measurements and state estimator solutions. For the distribution network, the proposed measurement-based voltage stability index will be further modified to conduct the stability assessment task. For both meshed and radial power networks, the impact of grid integration of wind generation will be explored.

1.4 Organization of the Dissertation

This dissertation starts with the classification analysis of power system stability status using a model-based approach. It is followed by model-based stability margin prediction using regression trees. A new approach of placing the PMUs at the most critical substations is proposed next. The variable importance derived from decision trees has been adopted to rank the importance of different substations in monitoring system stability. A measurement-based approach has been developed to estimate in real time the transmission system voltage stability using field PMU measurements received from utility companies. Later the algorithm has been modified to adapt to the situation of distribution system. The impact of wind farms integrated to distribution network on voltage stability has been explored using the proposed measurement-based approach. At last, a novel data mining technique, the Probabilistic Collocation Method, has been utilized for uncertainty analysis of grid integration of wind generation.

2. POWER SYSTEM STABILITY ASSESSMENT USING CLASSIFICATION TREES*

Conventional time-domain simulation based on system modeling has been used as the primary tool to analyze power system stability. This method is straightforward and accurate as long as adequate system model and measurements are used. However, two challenges have prevented the simulation method from being used for real-time applications: 1) it is computationally involved; 2) it raises concerns over approximate analysis results when a simplified model is used. As the importance of real-time stability monitoring and early detection of system events has been increasingly emphasized recently, an alternate approach based on data mining technique, especially the DT method, was attempted in this research [38]. The proposed approach of using classification tree for model-based stability assessment will be detailed in this section.

The proposed assessment scheme (using DT and PMU measurements) is shown in Figure 4 and compared with existing analytical method (using model simulation and SCADA data). As mentioned before, classification and regression trees are trained to emulate system behavior and predict system stability status. An abnormal operating point with insufficient damping or MW-distance can be immediately identified. Compared with the traditional time-domain simulation approach that requires full model

* Part of the material in this section is reprinted from "A fast stability assessment scheme based on classification and regression tree" by Ce Zheng, Vuk Malbasa, and Mladen Kezunovic, IEEE International Conference on Power System Technology (POWERCON), Auckland, New Zealand, Oct. 2012, DOI: 10.1109/PowerCon.2012.6401453 ©2012 IEEE, with permission from IEEE.

computation each time a new OP has emerged, the DT method is faster since repetitive model computations are avoided, and accurate if it is appropriately updated.

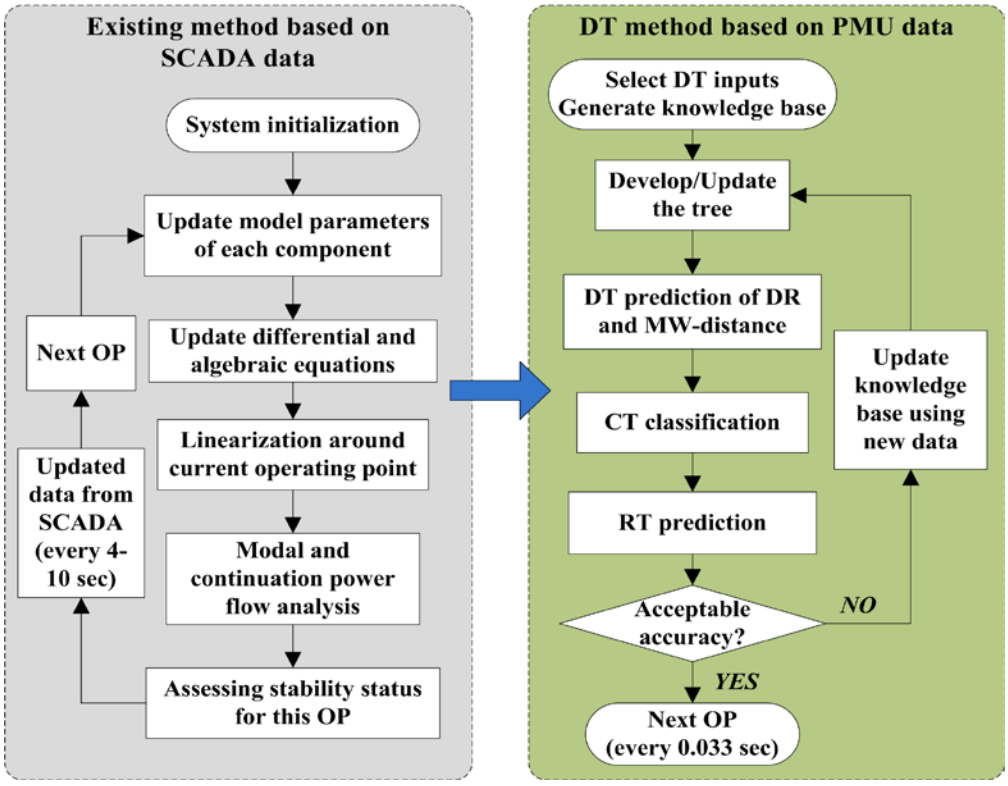


Figure 4 Comparison between existing method and the proposed scheme

2.1 Theoretical Formulation

Two important aspects of system operational performance, namely oscillatory stability and voltage stability, are targeted for monitoring. First the definition of an instability event is revisited:

- Oscillatory stability is related to Hopf bifurcation. An instability event occurs whenever, following a small disturbance, the damping torques are insufficient to

bring the system to a steady-state operating condition which is identical or close to the pre-disturbance condition [39]-[40].

- Voltage stability is related to saddle-node bifurcation. Voltage instability occurs when the load attempts to step beyond the capability of the combined transmission and generation system [6] [41].

2.1.1 Oscillatory Stability Assessment (OSA)

Modern power systems have evolved into systems of very large size. Initially separated systems are interconnected to each other. Different areas with larger generation capacity and inertia are added. Due to the deregulation and the difficulty of transmission expansion today, system operators are often forced to operate the system close to its stability limits, which leads to the reoccurrence of small-signal oscillation problem. As a consequence in large interconnected power systems small signal stability, especially inter-area oscillatory stability, become increasingly important.

The oscillatory stability is always analyzed by modal analysis. A power system can be described as a set of non-linear differential algebraic equations (DAE):

$$\begin{cases} \dot{x} = f(x, y, u) \\ 0 = g(x, y, u) \end{cases} \tag{2.1}$$

where x is the state vector, y is the algebraic vector, and u is the input vector. The DAEs in (2.1) are formulated by detailed modeling of each network component. By linearizing the non-linear equations in (2.1) at a particular system operating point, the following equations are derived:

$$\begin{cases} \Delta\dot{x} = A\Delta x + B\Delta u \\ \Delta y = C\Delta x + D\Delta u \end{cases} \tag{2.2}$$

The matrices *A*, *B*, *C*, and *D* in (2.2) provide a linearization around the system equilibrium point. Each pair of complex conjugate eigenvalues of matrix A corresponds to an oscillation mode of the system. The *A* matrix can be further decomposed as:

$$A = \Phi\Lambda\Psi \tag{2.3}$$

In (2.3), Λ represents the diagonal eigenvalue matrix, and Φ and Ψ are left and right eigenvector matrices respectively. For the i^{th} oscillation mode with the following conjugate eigenvalues:

$$\lambda_i = \sigma_i \pm j\omega_i \tag{2.4}$$

The oscillation frequency is given by:

$$f_i = \omega_i / 2\pi \tag{2.5}$$

The mode damping ratio (DR) can be calculated by:

$$\xi_i = \frac{\sigma_i}{\sqrt{\sigma_i^{\,2} + \omega_i^{\,2}}} \tag{2.6}$$

The critical oscillation mode that is insufficiently damped should be closely monitored. Assume DR_{crit} is the damping ratio of the critical mode, the scheme shown in Figure 5 is proposed for OSA.

As shown in Figure 5, three oscillatory stability states, namely "Stable" (including "Good" and "Fair"), "Alert" and "Unstable', are defined according to the value of DR_{crit}. Here we specify the thresholds of DR_{crit} at 6% and 3%. The generic thresholds for oscillatory stability assessment will be discussed in Section 7. A

classification tree (CT) is used to assign a system operating point (OP) into one of the above stability states.

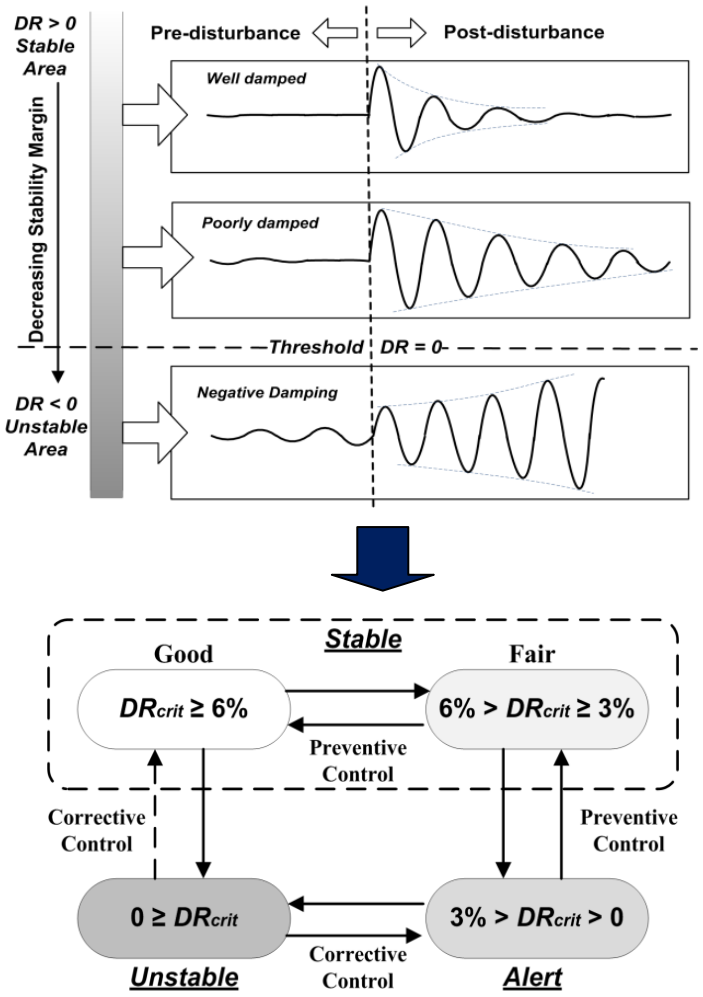


Figure 5 Proposed oscillatory stability assessment scheme

2.1.2 Voltage Stability Assessment (VSA)

Voltage instability occurs when the load attempts to step beyond the capability of the combined transmission and generation system, i.e. crosses the maximum deliverable

power [6]. Various methodologies for voltage stability analysis have been proposed. Among them the Continuation Power Flow (CPF) method [42] is able to provide a reliable measure of how far the system can move away from its current OP and still remain secure.

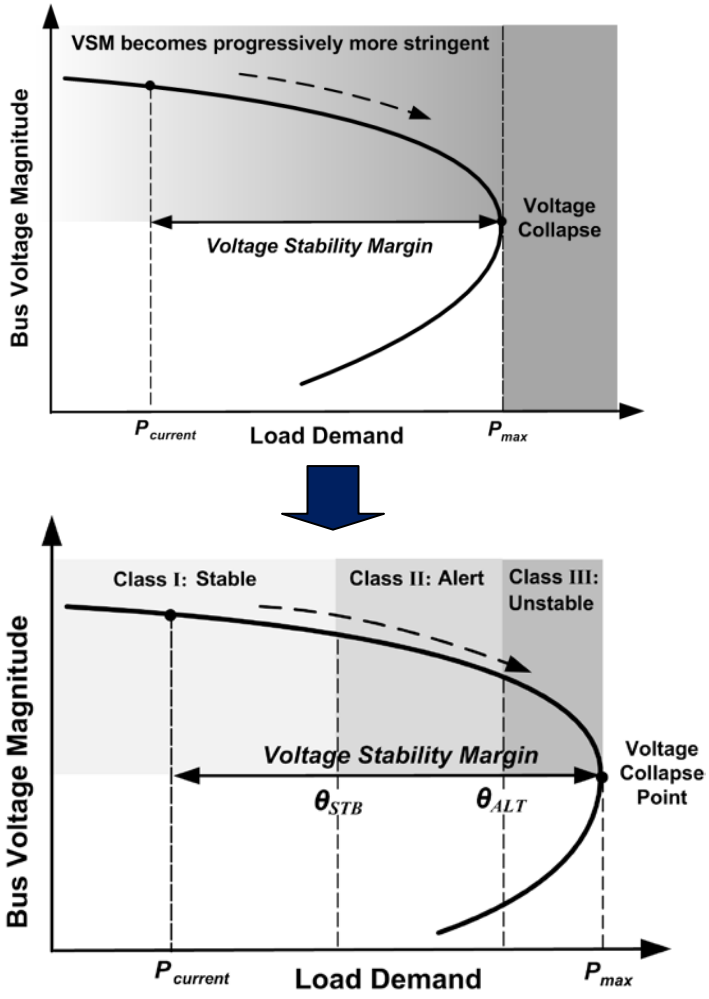


Figure 6 Proposed voltage stability assessment scheme

Assuming a constant load power factor, the variation of the voltage magnitude with the increase of load demand is plotted in Figure 6. In this work the idea of CPF

algorithm is explored, and the MW-distance from current OP to the critical voltage collapse point is used as the voltage stability indicator:

$$\mathrm{P_{distance}} = \mathrm{P_{collapse}} - \mathrm{P_{current}} \tag{2.7}$$

$$VS^{i}_{margin} = \frac{MW^{i}_{distance}}{P^{i}_{\max}} \times 100\% \tag{2.8}$$

Where P_{max} is the maximum deliverable power, and $P_{current}$ is the active load demand of current OP. VS^{i}_{margin} represents the voltage stability margin of the i^{th} OP.

For the given voltage stability thresholds $\boldsymbol{\theta}_{\mathbf{STB}}$ and $\boldsymbol{\theta}_{\mathbf{ALT}}$ ($\boldsymbol{\theta}_{\mathbf{STB}} > \boldsymbol{\theta}_{\mathbf{ALT}}$), OPs will be labeled as "Stable" as long as they satisfy $VS^{i}_{margin} \geq \boldsymbol{\theta}_{\mathbf{STB}}\mathrm{VS_{margin}} > 40\%$; and 'Unstable' when $\boldsymbol{\theta}_{\mathbf{ALT}} \geq VS^{i}_{margin}$. The remaining OPs are labeled as "Alert".

2.1.3 Proposed DT-based Predictive Model

As shown in Figure 2, the relationship and difference between the conventional time-domain simulation approach and the CT method is demonstrated. Compared with the traditional method, the advantage of CT method lies in its capability of fast analysis facilitated by fewer required inputs and straightforward model structure. By learning the system behavior from a known set of OPs, the CT model can predict system responses without detailed model computations. In addition, the CT method is particularly appealing because it uses a white-box model, which makes the results easy to interpret. Based on the combination of splitting rules along a path of the tree, preventive and corrective control strategies could be formulated.

2.2 Knowledge Base Generation

The knowledge base is a database used for off-line training of the CT-based predictive model. It is composed of a number of instances, and each instance represents a system operating point and is labeled with corresponding stability states. The larger the system is, the more attributes are needed to characterize the OP. These attributes comprise voltage and current phasors, active/reactive power flow, and some composite attributes.

Typically, the DT-based predictive model will gain more generalization power if a larger number of instances are included in the knowledge base. However the database generation process should be correctly designed, otherwise it will not capture sufficient information from the entire problem space.

Both the voltage stability and oscillatory stability are closely related to the load/generation composition of a power system, and their increase/decrease trend at a certain system snapshot [40]. If the load/generation composition varies, different OPs are formed. The change in the load demand and generation output can be described as:

$$P_G = P_G^0 + \Delta P_G \qquad Q_G = Q_G^0 + \Delta Q_G$$

$$P_L = P_L^0 + \Delta P_L \qquad Q_L = Q_L^0 + \Delta P_L \times Q_L^0 / P_L^0 \tag{2.9}$$

where P_G and Q_G are active/reactive power outputs of all the generators except the slack bus generator, P_L and Q_L are vectors of active/reactive power delivered to the loads. Superscript 0 represents base case OP. The vectors ΔP_G, ΔQ_G, ΔP_L and ΔQ_L stand for the variations in power.

In this work, the commercial software PSS/E is used for iteratively solving load

flows, and deriving characteristic matrix *A* at different OPs through numerical perturbation. Python and MATLAB [43] programs are developed to automate the PSS/E simulations, perform modal analysis, conduct the CPF-based voltage stability analysis, compute stability margins, and establish the knowledge base. The pseudo-code for knowledge base creation is illustrated below:

1) Initialize PSS/E in Python. Import system model parameters: Number of Generation Buses = i, Number of Load Buses = j, Number of buses with shunt capacitor = k
2) Let u ($u \in N$) be the iteration index with a step change of $C_{G/L/S}\%$

   Suppose G_1 is slack bus. Repeat:

   for $A_2=0 \rightarrow u_2$ do

   Scale the output of G_2 to: $P_{G2} = P_{G2}^0 (1 + A_2 \times C_{G2}\%)$

   …

   for $A_i=0 \rightarrow u_i$ do

   Scale the output of G_i to: $P_{Gi} = P_{Gi}^0 (1 + A_i \times C_{Gi}\%)$

   for $A_{i+1}=0 \rightarrow u_{i+1}$ do

   Scale load 1 to: $P_{L1} = P_{L1}^0 (1 + A_{(i+1)} \times C_{L1}\%)$

   …

   for $A_{i+j}=0 \rightarrow u_{i+j}$ do

   Scale load j to: $P_{Lj} = P_{Lj}^0 (1 + A_{(i+j)} \times C_{Lj}\%)$

   for $A_{i+j+1}=0 \rightarrow u_{i+j+1}$ do

Scale shunt 1 to: $Q_{S1} = Q_{S1}^{0}(1 + A_{(i+j+1)} \times C_{S1}\%)$

…

for $A_{i+j+k}=0 \rightarrow u_{i+j+k}$ do

Scale shunt k to: $Q_{Sk} = Q_{Sk}^{0}(1 + A_{(i+j+k)} \times C_{Sk}\%)$

Solve the load flow at: $\{P_{G2},...,P_{Gi},P_{L1},...,P_{Lj},Q_{S1},...,Q_{Sk}\}$

If this OP is unsolvable: eliminate

Oscillatory Stability Analysis:

Import system model dynamic data. Derive the *A* matrix.

Voltage Stability Analysis:

Derive the voltage collapse point via continuation-based method

Export computed features of current OP

End Loops

3) Repeat: for i=0→number of OPs do

Modal analysis of *A* matrix using (3)-(5): *DR* (ζ_i)

Compute voltage stability index using (6)-(7): VS^{i}_{margin}

Export computed stability margins

End Loop

2.3 Features Available to DT for Prediction

With respect to the input attributes of a decision tree, it is reported that different attribute combinations may result in different data mining accuracies [33]. In order to

accelerate the prediction process, it is desirable to use the least number of attributes as DT inputs while keeping an acceptable level of overall prediction accuracy. Typically the input attributes are selected using engineering insight and empirical evidence.

In this work we consider the basic measurements from a PMU. The involved DT input attributes are as follows:

- VM_i and VA_i: positive sequence voltage magnitude and phase angle at Bus i;
- IM_i_j and IA_i_j: positive sequence current magnitude and phase angle from Bus i to Bus j.

The commercial software CART [44] is used to develop CTs for evaluation of oscillatory stability and voltage stability.

2.4 Performance Examination of Classification Tree

2.4.1 Description of Test Systems

Two test systems, namely the IEEE 3-machine 9-bus system [39] and the IEEE 10-machine 39-bus system [46], are used to implement the proposed scheme. The one-line diagrams of these two test systems are shown in Figure 7.

2.4.2 Knowledge Base Preparation

The knowledge base generated for these two systems are summarized in Table 1.

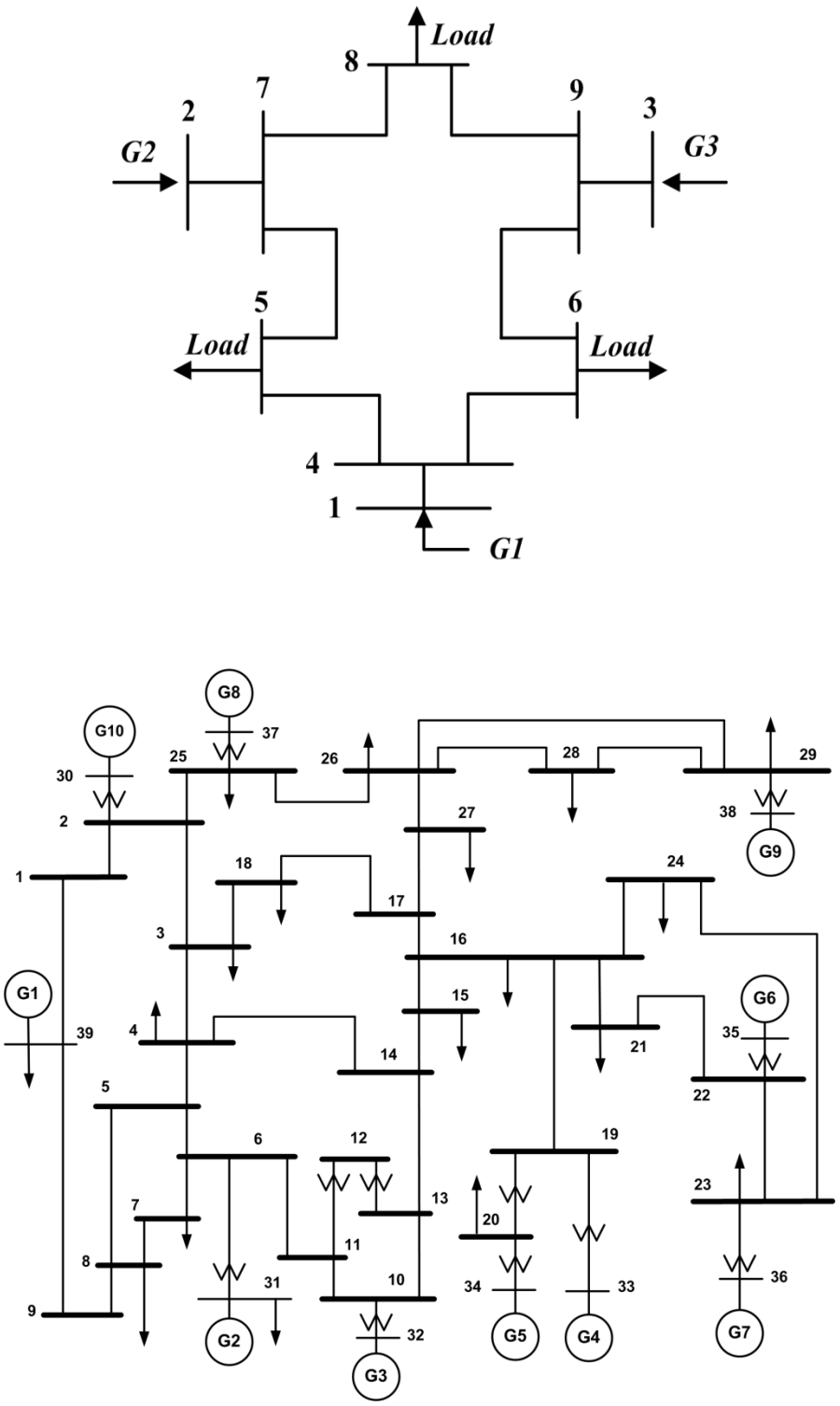


Figure 7 One-line diagrams of the IEEE 9-bus and 39-bus test systems

Table 1 Procedures for Knowledge Base Generation

System	Instances included in OSA Knowledge Base			Total
	Stable	Alert	Unstable	
9-Bus	663 (61.90%)	358 (33.43%)	50 (4.67%)	1071
39-Bus	2549 (71.30%)	962 (26.91%)	64 (1.79%)	3575
	Instances included in VSA Knowledge Base			
9-Bus	707 (51.68%)	495 (36.18%)	166 (12.13%)	1368
39-Bus	2206 (60.21%)	1175 (32.07%)	283 (7.72%)	3664

2.4.3 Adjustment of Priors and Selection of Attributes

It can be observed from Table 1 that the number of instances included in each class is highly unbalanced. Compared with some other data mining tools that do not perform well when dealing with unbalanced data, the classification tree integrated in CART has the strength to assure that every class will be treated equally regardless of its size. This is achieved by specifying the *Priors* for each class. In this work the *Prior* for the "Unstable" class has been adjusted to be slightly higher than that of other classes. The objective is to put more emphasis on the detection of unstable instances.

2.4.4 Performance of Classification Tree

The theoretical background of developing a CT in CART can be found in [7]. Each of the above generated knowledge bases has been randomly split into two data sets: 80% of the instances are used as training set; the remaining 20% serve the purpose of independent testing. Due to the stochastic nature of the splitting process, slight differences may occur between the derived CTs and their performance. Therefore in this work, the process of knowledge base splitting, tree training and testing has been replicated at least 10 times until the mean value and standard deviation of new case testing accuracy become stable.

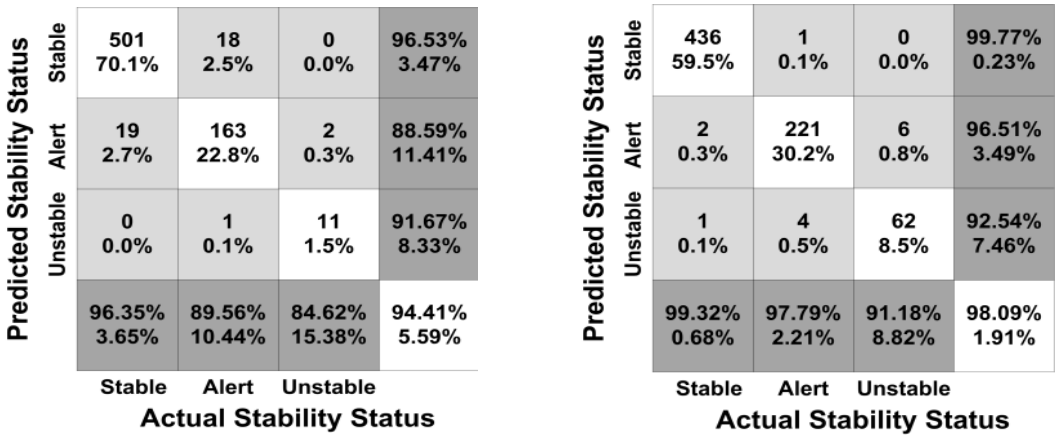
The *Entropy* method is adopted to grow the CTs in CART. The performance of CTs in new case testing is summarized in Table 2.

Table 2 Performance of the Classification Tree

System	Method	Accuracy of New Case Testing	
		OSA	VSA
9-Bus	*Entropy*	98.63%	99.56%
39-Bus	*Entropy*	94.38%	97.95%

The CT new case testing results of one replication for the IEEE 39-bus system are shown in Figure 8. An interesting observation from Table 2 and Figure 4 is that the CT performance for OSA is less accurate than that of VSA. This is because the system oscillatory stability behavior is highly non-linear. In order to reach certain prediction

accuracy, a larger training dataset is needed by OSA-CT compared with VSA-CT. In this work, more instances could be generated if we set the *Stopping Criterion (b)* with a higher accuracy requirement.



a) Testing results of 39-bus OSA b) Testing results for 39-bus VSA

Figure 8 CT stability assessment for the 39-bus system in one replication

The classification tree can be developed using different methodologies, e.g. *Gini*, *Twoing*, and *Entropy*. Another important setting is the minimum cases a parent node should have, which may impact the size of resulted CT. In this work the tree settings are varied to explore their impact on the assessment accuracy. The results are shown in Figure 9.

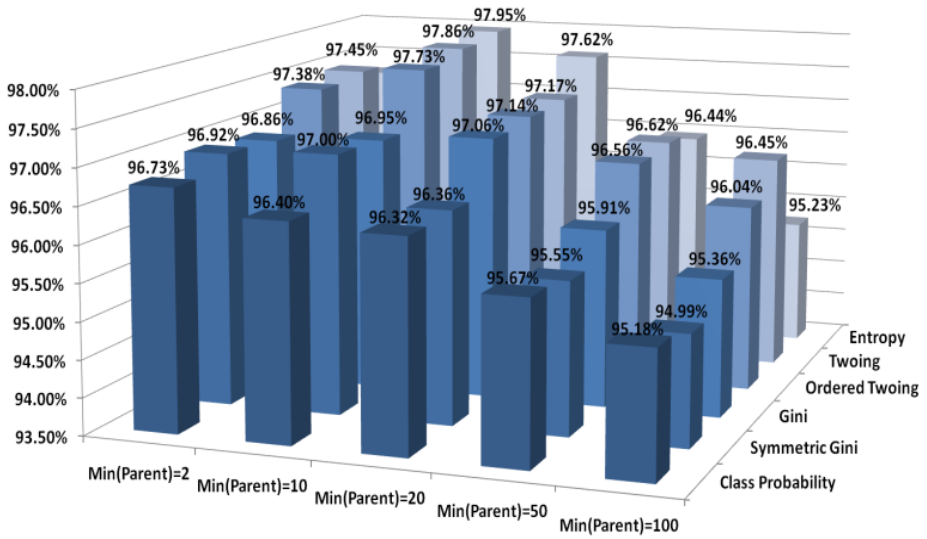


Figure 9 Classification tree performance using different tree growing methods

Two conclusions could be made from Figure 9: 1) the CT performance for the stability assessment problem is related to how the tree is developed. In this case the *Entropy* method achieved the best classification accuracy; 2) the setting for minimum parent node cases can alter the shape of the resulted tree as well as its performance. In general, the more cases a parent node is required to have, the less terminal nodes the derived CT may possess. This experiment demonstrated that there is a trade-off between tree complexity and accuracy. A large-sized tree may encounter the over-fitting problem, whereas a small-sized tree that is not adequately developed may produce less accurate classification results. A trial and comparison process is needed to find the best CT, and this can typically be accomplished by nested cross-validation.

2.5 Summary

This section explores the use of classification trees for fast evaluation of oscillatory stability and voltage stability. The following conclusions were reached:

- Two reliable stability metrics have been deployed and several stability states were defined. They provide an accurate assessment of the stability status of each OP;
- A systematic methodology for knowledge base generation has been proposed. Stopping criteria were elaborated to assure a sufficient dataset for CT training. Encouraging results were obtained through performance examination using the generated knowledge base;
- The CT classification accuracy is related to how the tree is developed, and the setting for minimum parent node cases can alter the shape of the resulted tree as well as its performance.

3. POWER SYSTEM STABILITY MARGIN PREDICTION USING REGRESSION TREES*

3.1 Theoretical Formulation

The oscillation modes that carry significant amount of energy but with insufficient DR are critical among all modes and need to be keenly monitored. Occurrence of an instability event is possible when a poorly damped mode is excited by a small or large disturbance. In this work the DR of critical oscillation mode is used as the oscillatory stability margin (OSM) indicator [45]. As shown in Figure 5, the OSM becomes progressively more stringent as the value of critical mode DR decreases.

The damping ratio is not an index from the parameter space, so strictly speaking it may not be proper to term it as 'margin'. In this work DR is selected as the OSM indicator in the sense that it provides smooth movement trajectory, clear partition between stable/unstable states, and an explicit distance from unstable point.

The variation of load bus voltage magnitude with different load demand is plotted as the P-V curve shown in Figure 6. The MW-distance from the current operating point to the voltage collapse point ('Knee' point), where the load demand equals the maximum deliverable power, provides a reasonable measure of system voltage stability margin. The VSM referred here corresponds to system long-term voltage stability, which

---

* Part of the material in this section is reprinted from "Regression tree for stability margin prediction using synchrophasor measurements," by Ce Zheng, Vuk Malbasa, and Mladen Kezunovic, *IEEE Trans. Power Syst.*, Vol. 28, No. 2, pp. 1978-1987, May 2013. DOI: 10.1109/TPWRS.2012.2220988 ©2012 IEEE, with permission from IEEE.

cannot be used to capture the short-term voltage stability.

(a) Generate *n* different OPs

(b) For each OP, determine the maximum deliverable power by means of the CPF technique

(c) Calculate the voltage stability margin for the i^{th} OP using the following index:

$$VS^{i}_{margin} = \frac{P^{i}_{\max} - P^{i}_{current}}{P^{i}_{\max}} \times 100\% \tag{3.1}$$

(d) Train the RT off-line using selected features from the *n* OPs and their corresponding VS_{margin}

(e) Use the trained RT to predict VSM in real time

3.2 Proposed Research

3.2.1 Regression Tree Method

Compared with the traditional time domain simulation approach that requires full model computation each time a new OP has emerged, the advantage of RT method lies in its simplified model structure and fast OP analysis facilitated by fewer required inputs. Figure 10 provides a simple example of RT structure. The unfolding OP is related to its stability margin through a unique top-down path. The splitting rule at each node that belongs to a given path represents an operational threshold. Based on the combination of splitting rules along the path, preventive and corrective control strategies could be formulated and initiated.

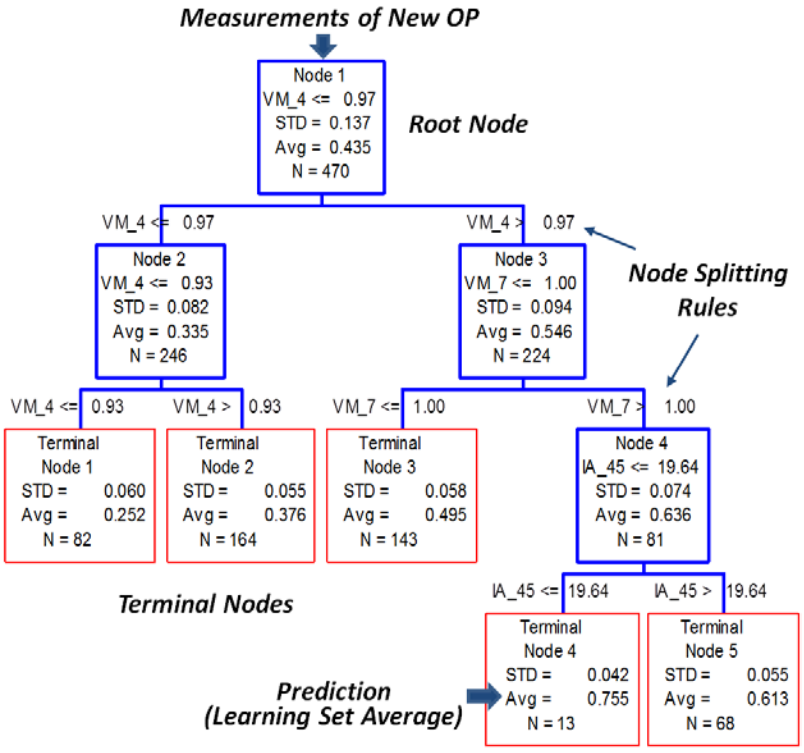


Figure 10 An example of the RT model structure

In regression analysis, a case consists of instance (x, y) where x is the vector of attributes and y is the target. The relationship between x and y is usually described by a regression function, through which it is possible to estimate how the target y changes when x is varied. In our proposed approach, the regression function is replaced by a binary tree structure, where x is the synchrophasor measurements and y is the system stability margin, i.e. the damping ratio or MW-distance. CART is used to develop OSM-RT and VSM-RT used for evaluating oscillatory and voltage stability margins respectively.

The approach to build a RT entails three steps: 1) tree growing using learning

dataset; 2) tree pruning using test dataset or cross-validation; 3) selection of the best pruned tree. Experimental tests show that there is a trade-off between the tree complexity and its accuracy: a small-sized tree cannot capture enough system behavior, and a large-sized tree usually leads to imprecise prediction due to its over-fitting model. In this work the rule of minimum cost regardless of size to search for the best pruned RT commensurate with accuracy is adopted. The complexity cost parameter in CART has been set to equal to zero. The RT growing, node splitting, tree pruning and optimal tree selection algorithms are detailed in Appendix A.

3.2.2 Proposed Approach

The proposed framework for RT-based stability margin prediction and event detection is shown in Figure 11. PMU measurements from different substations are collected and time-aligned by the Phasor Data Concentrator (PDC). The synchrophasor measurements are then delivered to the Wide Area Measurement System (WAMS) server located at the central control facility. At the control center operator room, the RTs for monitoring OSM (OSM-RT) and VSM (VSM-RT) are trained and updated periodically. The PMU data of an upcoming OP is dropped down the respective tree until it reaches a terminal node. Then the predicted stability margin is the average value of the learning set samples falling into that terminal node. Any OP with insufficient stability margin will be detected immediately by checking corresponding thresholds. Operators are alerted with the possible event and preventive control strategies can be initiated in a timely manner.

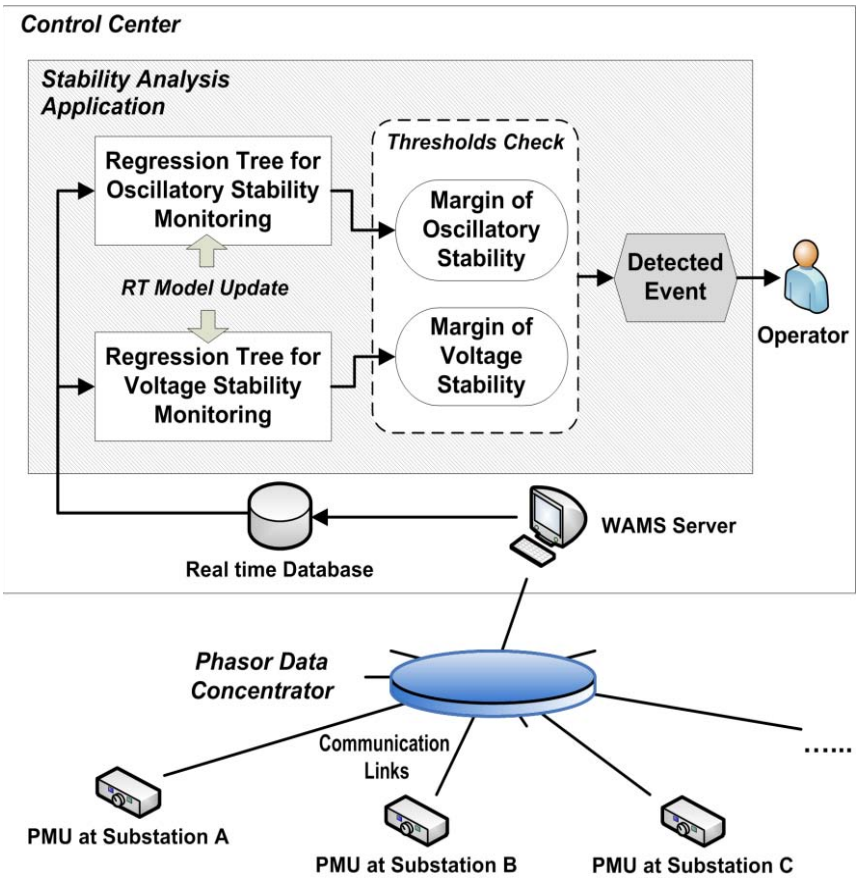


Figure 11 Proposed framework of the RT-based stability margin prediction and event detection

3.3 Knowledge Base Generation

Using the same approach illustrated in Section 2, the power supply at generation buses, demand at load buses, and the output of shunt capacitors were systematically varied. A total of 1071 OPs with corresponding OSMs, and 1153 OPs with corresponding VSMs have been produced for the 9-bus system. The number of records generated for the 39-bus system knowledge base is 4276 and 3664 respectively.

In addition, in this work the generator active/reactive power limits have been taken into account to reflect the practical stability margin. This has significant impact on the computation of VSM: when the load demand increases, a feasible load flow solution may not exist due to the limited generation capacity, even before the maximum loadability of the transmission system is reached. Therefore the derived P_{max} may be somewhere on the top half of the PV curve before the "Knee" point shown in Figure 6.

In order to build a sufficiently large knowledge base, in this work two stopping criteria are followed:

1) Each generator/load/shunt should be varied at least 4 times ($u \geq 4$) and the total variation should be at least 30% of the base value ($u \times C_{G/L/S} \geq 30$). The goal is to capture the most system behavior from the problem space;
2) The RT training and testing accuracy converges. The R^2 is used to measure the prediction accuracy and will be detailed in next section.

The trajectory of the 39-bus system stability margin is shown in Figure 12. Corresponding stability thresholds are shown as the flat planes dividing each margin space into two halves: an instability event will be immediately identified in the top half. For this power system the voltage stability threshold is put at VS_{margin}=30%. This value can be further adjusted according to the real-time operational needs.

As it can be observed from Figure 12, a high imbalance in size between the stable and unstable cases exists. This is a very practical issue in power system operation since most of the time the system is in its stable state. From the classification point of view, compared with some other data mining tools that do not perform well when dealing with

unbalanced data, the decision tree implemented in the CART software has the property of assuring that every class is treated equally regardless of its size. This is achieved by specifying the *Prior* of each class. From the regression point of view, there is no need to set *Priors* because each case will be treated as an equal point on the continuous stability margin space. Because of the least squares loss function for regression, as implemented in CART, large mistakes are penalized more than smaller ones, thus large errors at any OP are emphasized be they on the stable or unstable part of the stability margin space. Once the relationship between input and output is identified, the regression model defines a mapping of an OP to its stability margins regardless of the state/class the OP belongs to.

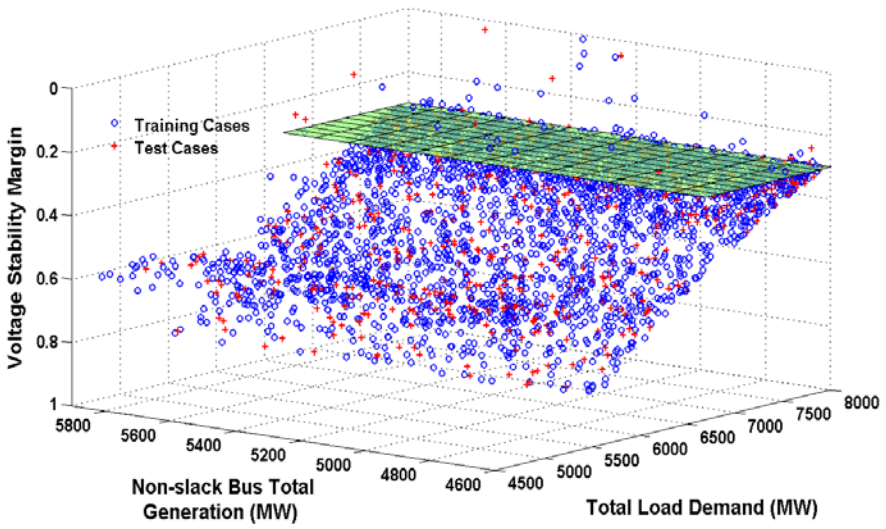


Figure 12 Trajectory of voltage and oscillatory stability margins of the IEEE 39-bus (New England) test system

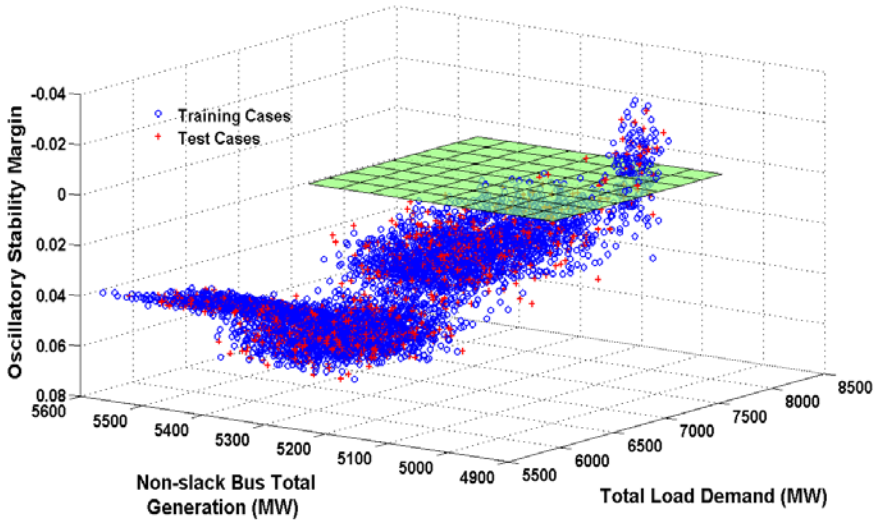


Figure 12 Continued

3.4 Off-line Training and New Case Testing

Each knowledge base is split into two independent data sets: 80% of the records are randomly selected for training of OSM-RT and VSM-RT; the remaining 20% of the records will serve the purpose of RT testing. The 10-fold *cross validation* method is adopted to grow the RT in CART. In experiments because of the random nature of the splitting process, slight differences may occur between the performances of each derived RT. Therefore in this work, the process of knowledge base splitting, tree training and testing has been replicated 10 times until the Mean and Standard Deviation of RT accuracy become stable.

In contrast with a classification tree for which the accuracy could be directly derived from the misclassification rate, the performance of a regression tree is measured through a statistical index, termed *Residuals Squared Error* (R^2) [47]. We report the accuracy of a RT model as follows:

$$R^2 = 1 - \frac{\sum_{(x_i, y_i) \in TS} [y_i - d(x_i)]^2}{\sum_{TS} (y_i - \bar{y}_{root})^2} \tag{3.2}$$

where TS is the set of training samples, x_i is input, $\bar{y}_i$ is the actual stability margin, $d(x_i)$ is the RT predicted value, and y_{root} is the mean of y_i in the tree root node.

In general the closer the value of R^2 is to 1, the better the prediction is. However in practice, how good an R^2 is depends on the particular application and the way it is measured [48]. Experimental results from this work show that a quite acceptable value of R^2>0.90 can be achieved.

Sometimes the R^2 alone may not be sufficient, especially in the case when the typical difference between values predicted by RT and the actual stability margins is desired. Therefore another measure, the *Root-Mean-Square* (RMS), is utilized:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} [y_i - d(x_i)]^2}{n}} \tag{3.3}$$

where n is the number of test cases. The numerator stands for the sum of squared deviations of the actual stability margins around the RT predictions. The value of RMS error depends on the base magnitude of the target stability margin to be predicted. In the proposed scheme, a typical value of OSM is in the range of -0.01 to 0.1, and the VSM is usually ranging from 0.05 to 1.0. Hence the RMS errors of VSM-RT are usually several times larger than that of the OSM-RTs.

Once the training is complete, the derived RTs were evaluated using the unseen test cases. Much more emphasis must be put on the accuracy of new case testing because, for real-time applications, a predictive model which cannot fit the unseen

system behavior well is unacceptable, even if high accuracy is obtained during the off-line training, as it lacks generalization power. The corresponding training and new case testing accuracy is summarized in Table 3. In addition, the results of new case testing were reported separately in terms of *Security Test* and *Reliability Test*. While the security test examines how well the stable OPs are predicted, the reliability test checks if all unstable OPs are correctly identified.

Table 3 Performance of the Regression Trees

System	Oscillatory Stability Margin (OSM-RT)				
	Train R^2	Unseen OPs Overall Accuracy		Reliability and Security Test (RMS)	
		R^2	RMS	Reliability	Security
9-bus	0.9984	0.9858	0.0023	0.00083	0.00235
39-bus	0.9617	0.9519	0.0034	0.00386	0.00328
System	Voltage Stability Margin (VSM-RT)				
	Train R^2	Unseen OPs Overall Accuracy		Reliability and Security Test (RMS)	
		R^2	RMS	Reliability	Security
9-bus	0.9928	0.9791	0.0184	0.03357	0.01480
39-bus	0.9941	0.9694	0.0211	0.02736	0.01965

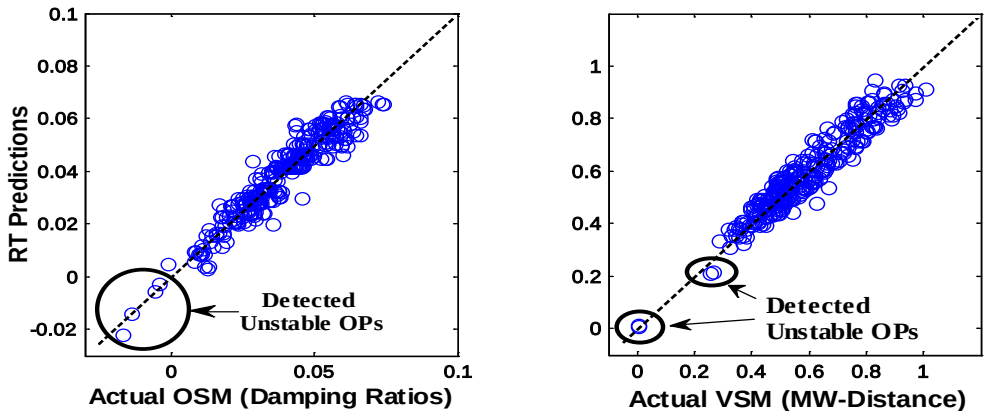


Figure 13 RT predicted margins versus the actual stability margins of the IEEE 39-bus system. Left: OSM-RT performance; Right: VSM-RT performance

The prediction for 300 new OPs of the 39-bus system is shown in Figure 13. The RT-based approach has exhibited encouraging capability for system stability margin prediction.

The performances of differently sized OSM-RTs are summarized in the relative error curve shown in Figure 14. Among these trees, a 13-node subtree pruned from the 45-node "optimal" tree is shown in Figure 15(a), and the "Largest" tree with 465 nodes is shown in Figure 15(b).

Compared with the optimal tree, numerical results show that although the 465-node tree has boosted the training accuracy from 0.9617 to 0.9872 R^2, its accuracy in new case testing actually dropped from 0.9520 to 0.9407 R^2. This is because an over-developed tree may perform well in training, but it will lose the generalization power in predicting unseen instances. The optimal tree with the lowest relative cost has the best generalization power and should be selected.

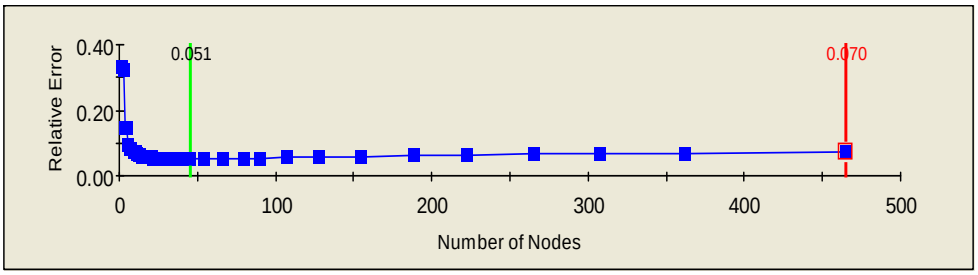
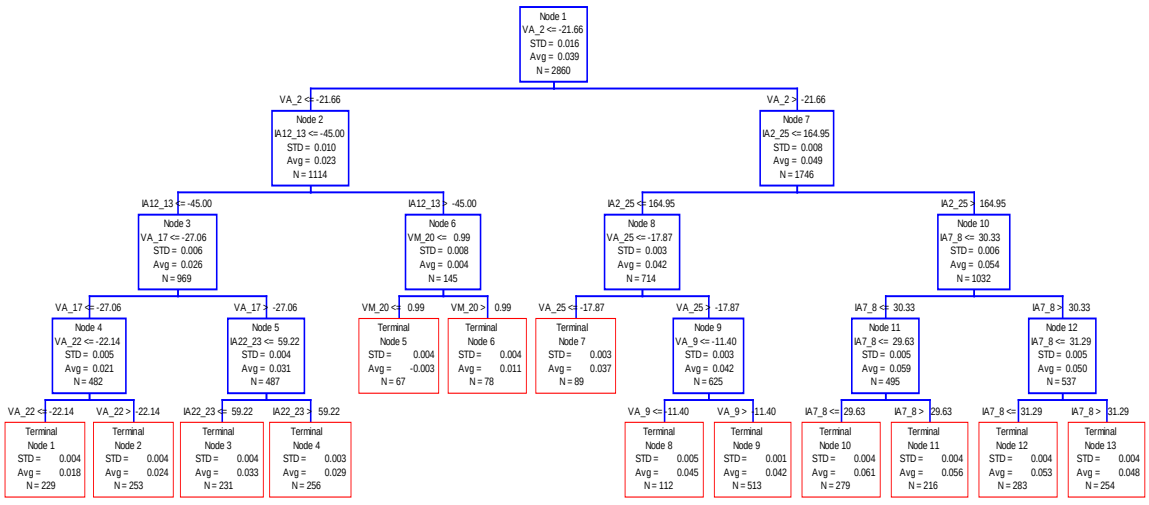


Figure 14 Relative cost of a series of differently sized RTs

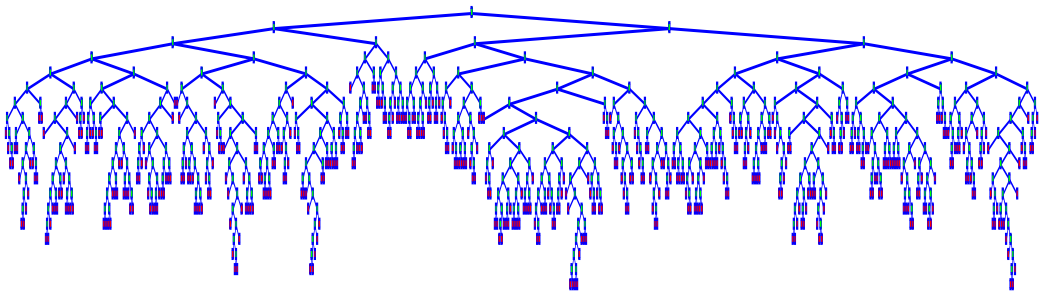
3.5 Comparison with Other Data Mining Tools

In this work the performance of RT has been compared with two widely used data mining tools: Support Vector Machine (SVM) and Neural Network (NN). The R^2 accuracy of different data mining tools for the 39-bus system is summarized in Table 4.



(a) 13-node tree pruned from the optimal OSM-RT

Figure 15 Regression trees for oscillatory stability margin prediction



(b) Largest RT with 465 terminal nodes

Figure 15 Regression trees for oscillatory stability margin prediction

Table 4 New Case Testing Accuracy using Different Data Mining Tools for the 39-Bus System

Tools	Testing R^2 of OSM	Testing R^2 of VSM
RT	0.9519	0.9694
SVM	0.9591	0.9811
NN	0.9579	0.9572

According to the results, the RT-based model achieved almost identical prediction accuracy as other data mining tools. Compared with some "black-box" tools, the DT piece-wise structure as shown in Figure 15(a) provides system operators with a clearer cause-effect relationship of how the system variables lead to the onset of an instability event. It is possible to identify the critical variables and thresholds that need to be analyzed to gain insight into the stability margin of a system.

3.6 Application to a Larger System

3.6.1 Description of the WECC Equivalent System

The RT-based predictive model has been applied to the Western Electric Coordinating Council (WECC) equivalent system shown in Figure 16 [49]. This network consists of 179 buses, 29 generators, 42 shunts, and 104 loads.

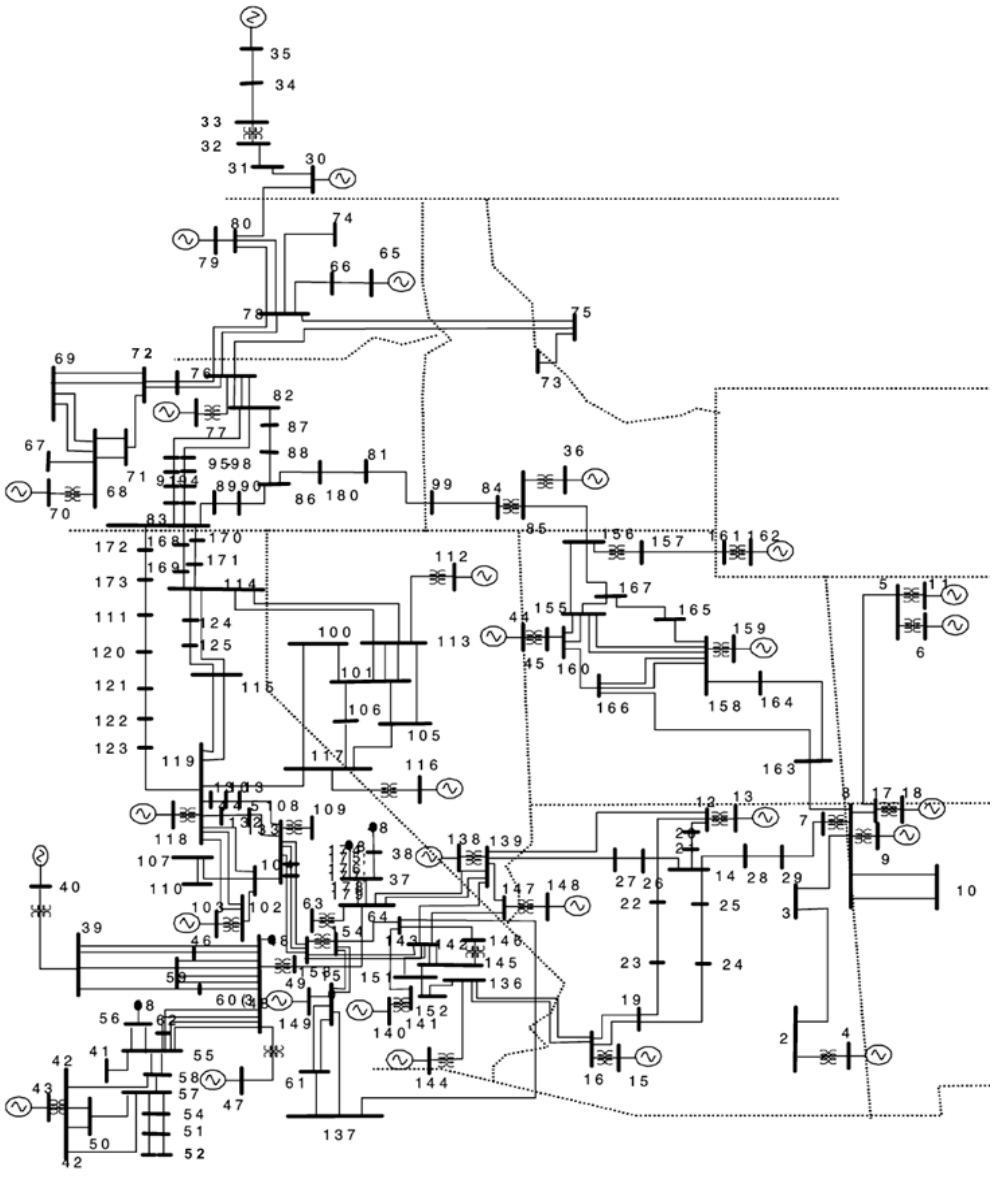


Figure 16 One-line diagram of the WECC 179-bus equivalent system

3.6.2 Knowledge Base Generation and RT Performance

The same methodology of creating the knowledge base for the 9-bus and 39-bus systems is adopted. In addition, two practical issues have been considered: 1) the real/reactive power output limit of each generator is more stringent in this larger system and should be complied with strictly; 2) it is computationally too expensive to generate the database by varying only one component each time. For instance, if the iteration index u is set to be 4, a total of 4^{175} OPs will need to be analyzed. It may be more practical to group the loads and generators according to their geographical locations. Seven areas are formed and it is assumed that the loads/generators within each area will increase/decrease at the same rate.

A total of 12572 records have been generated for the OSM-RT and 15303 records for the VSM-RT. The impact of the size of training set on the performance of resulted RT is examined: 100%, 50%, 20%, 10%, 5%, and 2% of the training cases are used to derive RT for each task. All experiments have been replicated 10 times and the Mean of new case prediction accuracy is summarized in Figure 17. It clearly shows that the prediction accuracy increased when more cases were used to train the RTs.

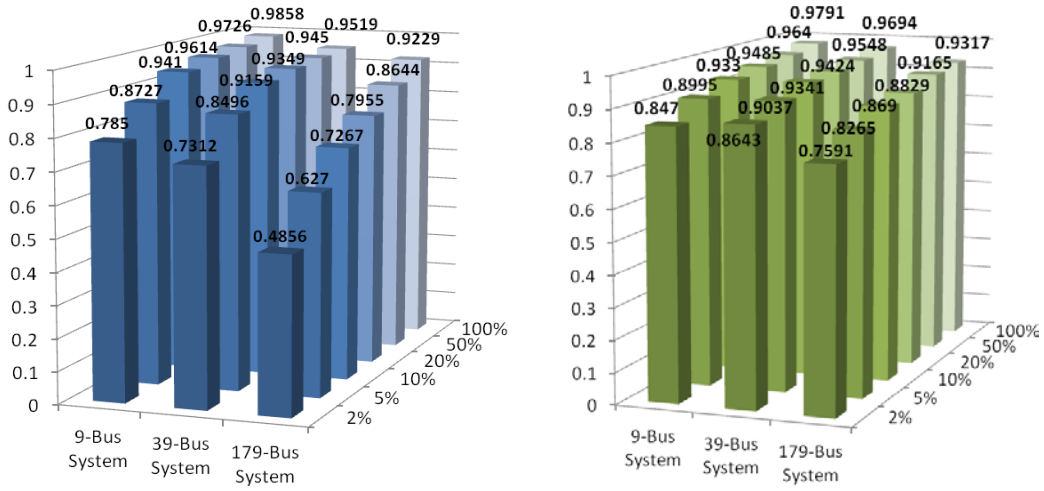


Figure 17 New case prediction accuracy of RTs trained with differently sized data sets. Left: OSM-RT; Right: VSM-RT

In order to embed the RT model into an actual online process, three aspects need to be examined and corresponding requirements must be satisfied: 1) Eligibility for high speed analysis; 2) Robustness to measurement error; 3) Capability to accommodate topology change.

3.6.3 Data Processing Speed

Traditionally the data used for the stability analysis in electrical utilities are obtained from the SCADA system or state estimation functions, which are refreshed on a time scale from several seconds to several minutes. These slowly updated data can only provide limited decision making support under the new situation where fast variations are present at both demand side and supply side. The capability to take advantage of the fast updated PMU data is critical in real-time applications.

In practice, the PMU measurements are updated very fast, most likely at least 30

times per second. In order to evaluate the system stability status at each snapshot, the processing of PMU data must be less than 1/30=0.033 second.

Table 5 Computational Speed of Regression Trees

Type of Regression Models	IEEE 39-bus System		WECC 179-bus System	
	Off-line Training	New Case Prediction	Off-line Training	New Case Prediction
OSM-RT	36.01 s (3421 cases)	about 3 s (855 cases)	164.97 s (10058 cases)	about 5 s (2514 cases)
VSM-RT	31.38 s (2931 cases)	about 2 s (733 cases)	195.45 s (12242 cases)	about 7 s (3061 cases)

The data processing speed of RTs is summarized in Table 5. The computational time is estimated using the built-in clock of CART executed on an Intel Pentium IV 3.00-GHz CPU with 2 GB of RAM. It can be seen that the derived OSM-RT or VSM-RT can assess 1000 new OPs in less than 4 s for the 39-bus system, and 3000 new OPs in less than 8 s for the WECC 179-bus system. According to the results, the RTs satisfy the speed requirement of real-time applications.

3.6.4 Impact of Measurement Errors

The phasor estimation process may introduce errors. PMUs manufactured by multiple vendors can also yield inaccurate readings. In real-time application, the PMU measurement errors of the i^{th} OP can be expressed as:

$$VM_i^{meas} = VM_i^{real} + \Delta\varepsilon_{VMi} \quad VA_i^{meas} = VA_i^{real} + \Delta\varepsilon_{VAi}$$

$$IM_i^{meas} = IM_i^{real} + \Delta\varepsilon_{IMi} \qquad IA_i^{meas} = IA_i^{real} + \Delta\varepsilon_{IAi} \tag{3.4}$$

where the superscript *real* means actual values of the phasor, and *meas* stands for measured values.

Table 6 Performance of the 179-Bus Regression Trees Considering PMU Measurement Error

Type of Regression Models	Add Noise Only to the Test Cases			
	Security Test		Reliability Test	
	R^2	RMS	R^2	RMS
OSM-RT	0.7906	0.00106	0.7403	0.00121
VSM-RT	0.8091	0.02785	0.7629	0.03010
Type of Regression Models	Add Noise to Both Training and Test Cases			
	Security Test		Reliability Test	
	R^2	RMS	R^2	RMS
OSM-RT	0.9170	0.00068	0.8994	0.00071
VSM-RT	0.9266	0.01789	0.9045	0.01940

According to the IEEE C37.118 "*Standard for Synchrophasors for Power Systems*" [50], PMUs that are Level 1 compliant with the standard should provide a Total Vector Error (TVE) less than 1%. This implies that the following constraints must be satisfied:

$$1\% > \left| \frac{VM_i^{meas} \angle VA_i^{meas} - VM_i^{real} \angle VA_i^{real}}{VM_i^{real} \angle VA_i^{real}} \right|$$

$$1\% > \left| \frac{IM_i^{meas} \angle IA_i^{meas} - IM_i^{real} \angle IA_i^{real}}{IM_i^{real} \angle IA_i^{real}} \right| \quad (3.5)$$

Considering (3.4) and (3.5), random noise $\Delta\varepsilon$ has been added to the original phasor magnitudes and angles of the WECC 179-bus system knowledge base. In Table 6 two scenarios were tested. While in both scenarios errors were added to the test cases, it is shown that the RTs trained with measurement error had much better performance than the ones without the error taken into account in the training data set.

3.6.5 Impact of Topology Variation

In this work the robustness of RT to certain system topology changes was examined. The scenarios that were evaluated and RT performances are summarized in Table 7.

It can be seen that OSM-RTs were able to provide somewhat acceptable predictions with low RMS errors, even under situations the network topology had changed. On the other hand, VSM-RTs appear to be less robust and the performance varied case by case: the *N-1* test in the 9-bus system had a significant impact on the VSM prediction due to the small size of the system; acceptable predictions were achieved for the case of generator outage in the 39-bus system; the *N-2* scenario in the 39-bus system was too severe for the VSM-RT to handle. More case studies were conducted on the 179-bus system VSM-RT: low RMS errors were observed in

experiments where slight topology changes are made, such as one of the double-circuit transmission lines out of service.

Table 7 Regression Tree Performance under System Topological Variations

Scenarios of Topology Change	Type	RMS Error of OSM-RT	RMS Error of VSM-RT
Line 8-9 taken out	9 BUS N-1	0.00880	0.15481
G10 out of service	39 BUS N-1	0.00417	0.04089
G10 and Line 26-28 taken out	39 BUS N-2	0.00726	0.20702
Line 1 of 76-82 out of service	179 BUS N-1	0.00337	0.03046
Line 1 of 90-156 out of service	179 BUS N-1	0.00421	0.02654
Line 1 of 95-98 out of service	179 BUS N-1	0.00385	0.03198
Line 81-180 out of service	179 BUS N-1	0.00552	0.08325
Line 1 of 90-156 and Line 1 of 76-82 out	179 BUS N-2	0.00473	0.04830
G63-1 and Line 1 of 95-98 out of service	179 BUS N-2	0.00574	0.03792
G63-1 and Line 81-180 out of service	179 BUS N-2	0.00588	0.10736

3.7 Discussion

3.7.1 Ability of RTs to Handle Evolving System Conditions

The problem of how to sustain the prediction accuracy of RT under the evolving system operating conditions is critical for its online implementation. In fact this is also a problem in all data mining tools. In general, the change of system operating conditions can be categorized into two types:

- The variation of system load/generation patterns;
- The variation of system topology due to contingencies, scheduled maintenance, and system dispatch.

The work reported in Sections 3.4, 3.5 and part of 3.6 tackles the first type of variation. As illustrated in the knowledge base creation process, the generator/load/shunt has been widely varied in a systematical way to capture the most system behavior from the problem space.

The change in system topology is a major reason that causes a data mining tool to fail in real-time applications. The results shown in Table 7 indicate that the RT sensitivity to topology changes becomes less distinct in large sized network and under mild changes in topology. This obviously helps in making RTs useful even under topology changes. It is also observed that RTs are not able to accommodate certain severe contingencies, e.g. the line 81-180 out of service. In the field of data mining and machine learning, the so-called ‘concept change’ describes methodology for dealing with such type of topology variation. A literature search reveals that there is not a generally effective way for the data mining tool to cope with the concept change

incrementally, although some work has shown results [51]. Most of the time a re-train using the updated knowledge base is necessary to reflect new topology condition.

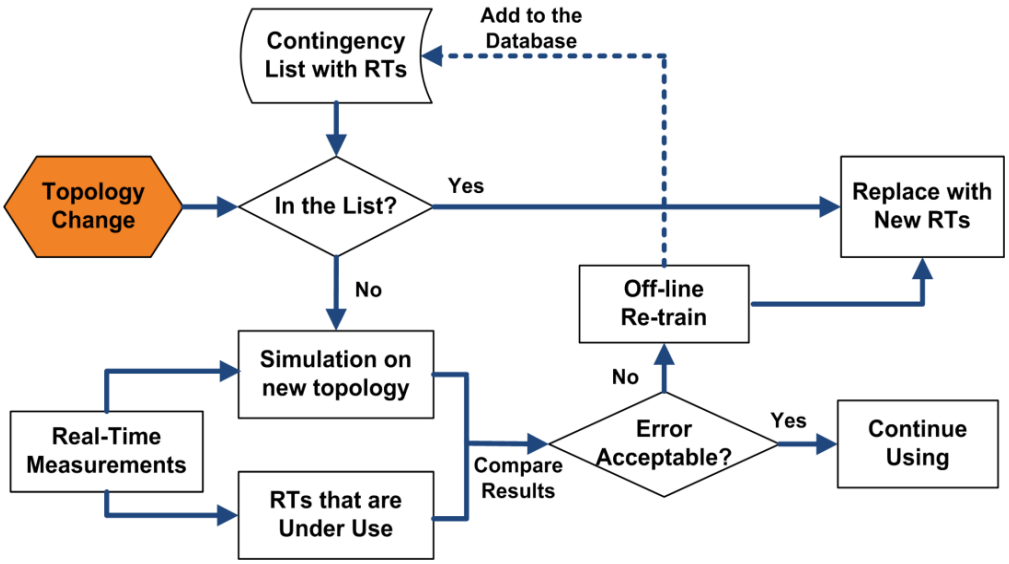


Figure 18 Scheme for RTs to handle system topology change

3.7.2 When and How to Update the RTs

To re-train an RT model whenever it is obsolete is time-consuming and may not satisfy the requirement of seamless on-line monitoring. An effective solution may be to prepare a knowledge base for each of the credible contingencies beforehand, and train a series of candidate RTs accordingly. Figure 18 shows the proposed scheme. The list of credible contingencies is usually readily available at utility companies. If in online application an unseen contingency occurs and RT fails to provide accurate predictions, a new RT will be trained and deployed. The new contingency scenario and RTs will be added to the historical database. With the increase of contingency scenarios accumulated

in database, fewer unseen topology conditions will be encountered. The obsolete models can be quickly replaced by the candidate RTs corresponding to the post-contingency condition.

3.8 Summary

In this work the approach of using regression tree to predict power system stability margins is explored and the following conclusions have been reached:

- Synchronized voltage and current phasors have been used as RT input feature. With a sufficiently large knowledge base, the RT model can predict the system oscillatory and voltage stability behavior with high accuracy;
- According to the test results, the RT model is fast enough to process PMU measurements, and it is robust to handle measurement errors that are within 1% TVE;
- The RT sensitivity to system topology variation becomes less distinct in large sized network and under mild changes in topology.

4. OPTIMAL PLACEMENT OF PHASOR MEASUREMENT UNITS*

In previous sections, the RTs were fed with voltage and current phasors measured at all buses. An underlying assumption is that almost every substation is equipped with a PMU. In practice, this is not economically feasible since the installation of PMUs and corresponding telecommunication path is very costly. A reasonable approach may be to install only a limited number of PMUs at the most critical substations. The problem of finding the optimal PMU location is equivalent to selecting the best reduced set of RT input features without a significant degradation in RT performance [45].

4.1 Combined Bus Ranking

Ideally, the optimal solution could be obtained through an exhaustive trial and comparison of all possible feature combinations. However it is computationally too involved to do so. In this work we are proposing a different approach and the idea comes from a unique property of the RT model structure. The topology of the 9-bus system OSM-RT derived in Section 3 is shown in Figure 19. It is interesting to track the actions of the tree. Each node has been split by an input variable, and the variable is selected as the splitter because it is the most powerful variable among all candidate features that can

* Part of the material in this section is reprinted from "Regression tree for stability margin prediction using synchrophasor measurements," by Ce Zheng, Vuk Malbasa, and Mladen Kezunovic, *IEEE Trans. Power Syst.*, Vol. 28, No. 2, pp. 1978-1987, May 2013, DOI: 10.1109/TPWRS.2012.2220988 ©2012 IEEE, with permission from IEEE.

best split the node. The variables gain credit towards their importance by serving as primary splitters that actually split a node, or as back-up splitters (surrogates) to be used when the primary splitter is missing [52]. By summarizing the variables' contribution to the overall tree when all nodes are examined, the *Variable Importance* (*VI*) can be obtained.

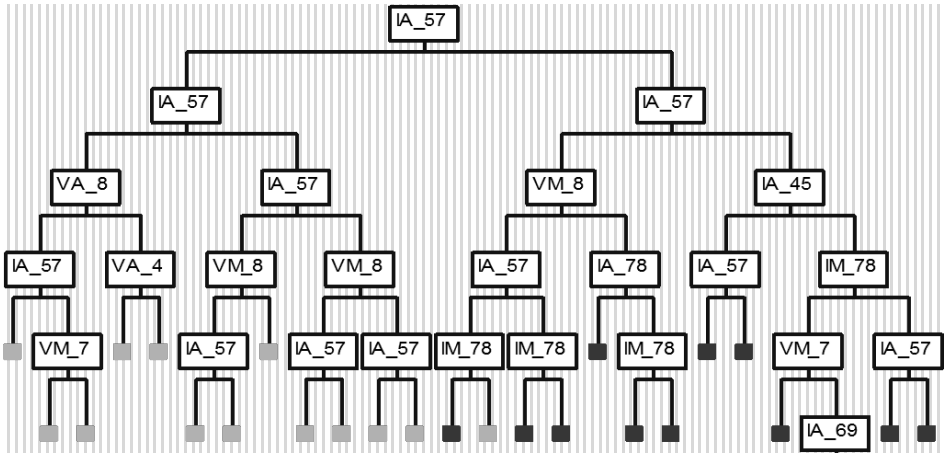


Figure 19 OSM-RT topology and node splitters of the 9-bus system

To calculate the *VI*, search all splits $s \in S$ on variable x_m at each tree node $t \in T$, and find the split s^*_m that gives the largest decrease in regression R [7]:

$$\Delta R(s^*_m, t) = \max_{s \in S} \Delta R(s,t) \tag{4.1}$$

Suppose s^* is the best of s^*_m, and $\tilde{s}_m$ is the split on variable x_m that has the best agreement with s^* in terms of partitioning cases, the measure of importance of variable x_m is defined as:

$$VI(x_m) = \sum_{t \in T} \Delta R(\tilde{s}_m, t) \tag{4.2}$$

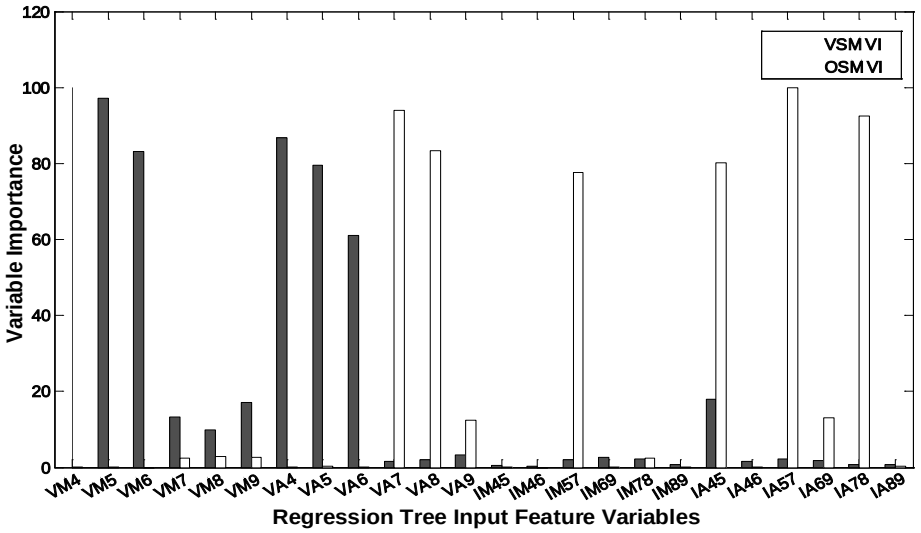


Figure 20 IEEE 9-bus system VSM-RT and OSM-RT variable importance

Figure 20 shows the computed *VI* for the OSM-RT and VSM-RT of the 9-bus system derived in Section 3. The actual measures of importance have been normalized so that the most important variable has a *VI* of 100.

The idea of *Combined Bus Ranking* (CBR) is as follows: The overall contribution of each bus to the oscillatory and voltage stability evaluation can be quantified by combining the importance of variables measured at this bus.

Mathematically the CBR of Bus i can be expressed as:

$$CBR_i = \sum_{x \in i} VI_{OSM-RT}(x) + \sum_{x \in i} VI_{VSM-RT}(x) \tag{4.3}$$

where *X* is the vector of RT input variables, *x* is the individual variable belong to *X*, and

VI(x) is its importance. By specifying $x \in i$, only the variables measured at Bus *i* will be counted.

4.2 Optimal PMU Locations

A ranking list of the bus contributions can be obtained by sorting the CBR values from high to low. The optimal PMU locations will be suggested by selecting the top ranked buses from the list. In this work the CBR of top ranked buses were computed by considering only the primary splitters, because the surrogate variables that appear to be important but rarely split nodes are almost certainly highly correlated with the primary splitters and contain similar information. Once the top ranked buses were selected, the standard *VI* considering both primary and surrogate splitters were used to rank the remaining buses. In Table 8, the CBR for the WECC 179-bus system was calculated and top 10 buses are listed. Also shown in the table are the 10 buses with the lowest CBR.

Suppose that a number of 4 to 20 PMUs will be installed in the WECC system. By placing them at the top ranked buses of Table 8, the resulted RT prediction accuracy for unseen OPs are summarized in Figure 21. The RT performance using the measurements from the lowest ranked buses is also presented for the purpose of comparison.

As shown in Figure 21, in contrast with the RTs fed with measurements from the lowest ranked buses, those constructed using the measurements from top ranked buses have exhibited better performances. Another conclusion could be made by comparing the R^2 of Figure 21 with Figure 17: almost identical RT prediction R^2 was achieved by

using the reduced set of measurements from the PMU locations suggested by CBR. Last but not least, there is a huge decrease of the complexity in RT training since much fewer features are used. The training time of the 179-bus RTs has been reduced from about 3 minutes to less than 30 seconds.

Table 8 WECC 179-Bus System Combine Bus Ranking

Top Ranked Buses			Lowest Ranked Buses		
Rank	Location	CBR	Rank	Location	CBR
# 1	Bus 90	100.82	# 170	Bus 162	0.31
# 2	Bus 100	100.23	# 171	Bus 163	0.28
# 3	Bus 95	38.27	# 172	Bus 172	0.24
# 4	Bus 96	18.47	# 173	Bus 168	0.12
# 5	Bus 97	13.99	# 174	Bus 85	0.11
# 6	Bus 67	12.73	# 175	Bus 50	0.02
# 7	Bus 12	12.52	# 176	Bus 92	0.02
# 8	Bus 11	8.48	# 177	Bus 94	0.01
# 9	Bus 9	8.44	# 178	Bus 165	0.01
# 10	Bus 20	8.24	# 179	Bus 171	0.00

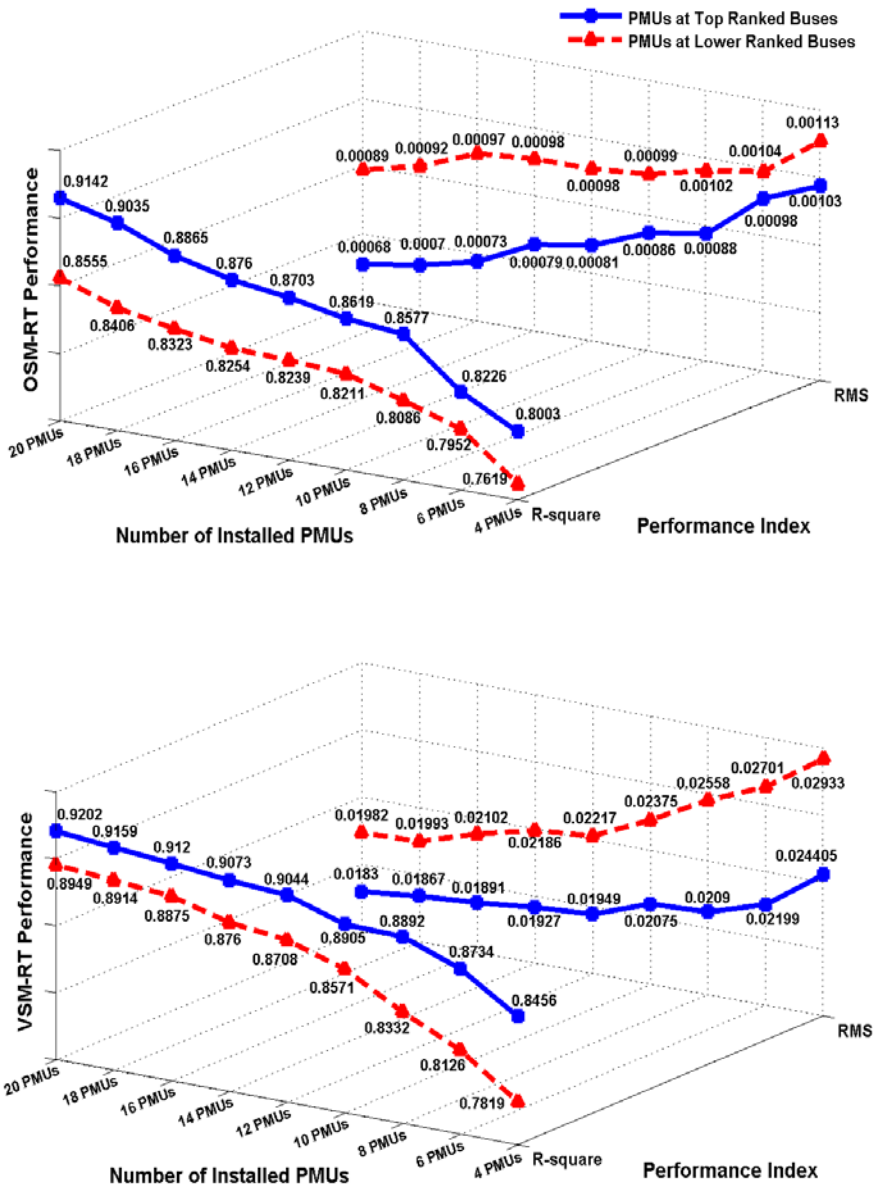


Figure 21 RT performance considering different PMU placements in the 179-bus system

4.3 Summary

A novel methodology for optimal placement of PMUs in a power network is proposed. The variable importance of each DT input features can be derived from CART

and utilized to rank the importance of network substations for stability assessment applications. The combined bus ranking derived from RT variable importance is used to suggest optimal PMU locations. The performance of DTs using synchrophasor measurements from a limited number of PMUs was checked. Test results show that the measurements from reduced locations can still lead to satisfactory RT prediction accuracy.

5. A MEASUREMENT-BASED VOLTAGE STABILITY INDEX AND ITS USE IN IMPACT ANALYSIS OF WIND INTEGRATION*

The objective of this section is to develop a measurement-based methodology for power system voltage stability assessment. The desired algorithm should be accurate and easy to utilize online. The proposed stability measures will be tested using actual PMU measurements to evaluate their eligibility for real-time application. In particular, the stability indices are derived using a "model-free" approach, i.e. the oscillatory or voltage stability margin is computed directly from synchrophasor measurements. A distinct advantage of the proposed measurement-based method is that it does not rely on any model parameter values. The measurement-based approach is by no means proposed to replace the existing approach based on system model parameters. It will serve as complement to the model-based approach.

As the installed capacity of wind power generation is increasing continuously, its impacts on system voltage stability have been intensively studied in recent days. Wind gusts will produce output power spikes and cause poor feeder voltage regulation, which may lead to voltage collapse when there is no sufficient reactive power support. An accurate voltage stability index is needed to quantify the voltage stability margin in real time. This section addresses such an index using a measurement-based approach. Next, the proposed stability index has been modified and implemented in simulations with

---

* Part of the material in this section is reprinted from "Distribution system voltage stability analysis with wind farms integration," by Ce Zheng and Mladen Kezunovic, North American Power Symposium (NAPS), 2010 DOI: 10.1109/NAPS.2010.5618978 ©2010 IEEE, with permission from IEEE.

wind farms integrated to distribution network are treated, where the required measurements are available only at feeder roots. Simulation results are presented to examine the performance of both fixed-speed wind turbine and variable speed wind turbine. STATCOM is also included to demonstrate its capability for voltage stability improvement.

5.1 Problem Formulation

Recently, wind farms continue to expand both in size and in capacity. It is expected in Texas that between 8% and 11% of the state electricity will come from wind resources by 2015. The Department of Energy (DOE) is exploring an energy scenario in which 20% of U.S. electricity is provided by wind by 2030 [53]-[54].

Wind farms connected to transmission grid and distribution network post a great challenge on the stability of an existing power system. Sudden wind gusts produce output power spikes and affect the stability of node voltages. If the node voltage stays too low, wind farms connected to that node may be tripped due to AC under-voltage protection. Further, without sufficient reactive power support, voltage instability may lead to voltage collapse [6]. Immediate warning of voltage instability is significant in maintaining normal operation of both wind farm and the whole system.

An intensive study of voltage stability indices started in 1990s. At that time, the system voltage stability was evaluated using very limited measurements [55]. Later, with the development of data acquisition technique, especially the time synchronized measurements between different locations, detection of the onset of voltage stability

became more accurate [23] [56]. A sensitivity-based approach to compute voltage stability along the system trajectory using PMU measurements was proposed in [23]. The methods proposed in [56] and [57] calculate the voltage stability margin by computing the Maximum Loadability Index (MLI) using data from PMUs.

Today, the importance of voltage stability analysis has been reemphasized because of the large integration of renewable resources, especially the wind power. A few studies that explore the interaction between wind farms and the node voltage stability have been reported [58]-[61]. Most of the papers are concentrated on the voltage stability of transmission systems.

Traditionally the wind farms utilize fixed-speed wind turbines, mostly the Squirrel Cage Induction Generator (SCIG), to produce power. The SCIG-based wind farms don't serve the function of voltage regulation and absorb reactive power from the utility grid. Now, variable speed wind turbines have become more common than traditional fixed-speed wind turbines. They had about 60% market share around the world in 2004 [63]. The variable speed wind turbines are either Doubly-Fed Induction Generators (DFIG) or full power converters. A built-in AC-DC-AC voltage source converter isolates the wind turbine from the outside system and functions like a STATCOM to provide required reactive power. The use of variable speed wind turbines usually enhances the voltage stability [64] and [65].

This section investigates the impacts of different types of wind turbines on the distribution system voltage stability. We have proposed a simple and accurate measurement-based voltage stability index (VSI) which, after modification, could adapt

to the situation of distribution system, where the measurements are available only at feeder root. The local measurements are used to compute the VSI in real time. The method accommodates the time skew of SCADA data. No time synchronization or information exchange between the monitoring locations is needed for the index calculation. Using the voltage stability index, the work has examined the performance of both fixed-speed and variable speed wind turbines in a distribution network. The effectiveness of the index is also demonstrated by simulation results.

5.2 Proposed Voltage Stability Index

The proposed voltage stability index will be formulated in this section. Consider a transmission line within a power system of any number of buses. The line equivalent model is represented in Figure 22.

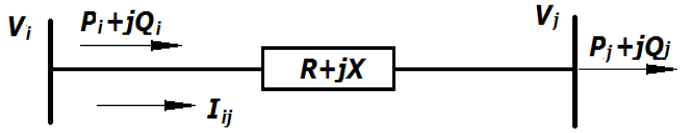


Figure 22 Equivalent model of a two-bus transmission system

Assume Bus i is the sending end and Bus j is the receiving end. The sending end voltage could be computed as below

$$V_i = V_j + I \cdot (R + jX)$$

$$= V_j + \frac{S_j^*}{V_j^*} \cdot (R + jX)$$

$$= V_j + \frac{P_j - jQ_j}{V_j^*} \cdot (R + jX)$$

$$= \frac{|V_j|^2 + P_j R + Q_j X + j(P_j X - Q_j R)}{V_j^*} \tag{5.1}$$

Substitute the voltage by its magnitude, (5.1) is rewritten as

$$|V_i| = \frac{\sqrt{(|V_j|^2 + P_j R + Q_j X)^2 + (P_j X - Q_j R)^2}}{|V_j|} \tag{5.2}$$

Rearrange (5.2) and the following is obtained

$$|V_j|^4 + \left[2(P_j R + Q_j X) - |V_i|^2\right] \cdot |V_j|^2 + (P_j R + Q_j X)^2$$

$$+ (P_j X - Q_j R)^2 = 0 \tag{5.3}$$

To guarantee that (5.3) is solvable, the following inequality constraint should be satisfied

$$|V_i|^4 - 4|V_i|^2(P_j R + Q_j X) - 4(P_j X - Q_j R)^2 \geq 0 \tag{5.4}$$

With the increase of the receiving end power demand $P_j + jQ_j$, the left side of (5.4) approaches zero, and the two-bus network reaches its maximum power transfer limit. From the steady-state/long-term voltage stability point of view, the maximum power transfer limit represents the "nose point" of the PV curve, at which the node voltage becomes unstable. Thus a voltage stability index could be extracted from (5.4).

To simplify the problem, a constant power factor at Bus j is assumed. $P_j + jQ_j$ could be then substituted by $(P_j + jQ_j) \cdot VSI$. VSI is the proposed voltage stability index having the following characteristics:

- VSI is greater than 1 when the transmission line is within its power transfer limit;
- VSI is equal to 1 when the transmission line is reaching maximum power transfer

capability;

- When VSI is less than 1, maximum power transfer limit is violated, and voltage becomes unstable.

Replacing $P_j + jQ_j$ by $(P_j + jQ_j) \cdot VSI$ in (4), we get

$$|V_i|^4 - 4|V_i|^2\left(P_jR + Q_jX\right) \times VSI$$

$$-4\left[\left(P_jX - Q_jR\right) \times VSI\right]^2 \geq 0 \tag{5.5}$$

By making the left side of (5.5) equal to zero, and extracting VSI, a measurement-based voltage stability index for a general two-bus transmission system can be obtained:

$$VSI = f(V_i, I_i, V_j, I_j) \tag{5.6}$$

Another task of this section is to analyze the distribution system voltage stability with wind farms integrated. To accomplish this, the network shown in Figure 22 has been considered as a distribution feeder, of which Bus *i* is the feeder root and Bus *j* is load bus. Further, modification of the VSI in (5.6) is needed to adapt to the situation of distribution system.

Since in distribution system the monitoring devices are usually installed at the feeder root (substation), we need to replace the receiving end power quantities in (5.6) using the sending end measurements. In Figure 22, Bus *j* powers could be replaced by Bus *i* powers using (5.6)

$$P_j + jQ_j = P_i + jQ_i - \left(R + jX\right)\left|I\right|^2 \tag{5.7}$$

Integrating (5.6) and (5.7), one gets

$$VSI = \pm \frac{\sqrt{\left[4|V_i|^2\left(P_jR + Q_jX\right)\right]^2 + 16|V_i|^4\left(P_jX - Q_jR\right)^2}}{8\left(P_jX - Q_jR\right)^2}$$

$$- \frac{4|V_i|^2\left(P_jR + Q_jX\right)}{8\left(P_jX - Q_jR\right)^2}$$

(Because the value of VSI is always greater or equal to zero, here we neglect the negative solution)

$$VSI = \frac{\sqrt{\begin{matrix}[4|V_i|^2((P_i - |I|^2R)R + (Q_i - |I|^2X)X)]^2 + \\ 16|V_i|^4((P_i - |I|^2R)X - (Q_i - |I|^2X)R)^2\end{matrix}}}{8((P_i - |I|^2R)X - (Q_i - |I|^2X)R)^2}$$

$$- \frac{4|V_i|^2\left((P_i - |I|^2R)R + (Q_i - |I|^2X)X\right)}{8\left(P_jX - Q_jR\right)^2} \quad (5.8)$$

After rearranging (5.8), the proposed voltage stability index for distribution system is obtained as below

$$VSI = \frac{|V_i|^2\sqrt{[(P_i - |I|^2R)^2 + (Q_i - |I|^2X)^2](R^2 + X^2)}}{2(P_iX - Q_iR)^2}$$

$$- \frac{|V_i|^2\left[(P_i - |I|^2R)R + (Q_i - |I|^2X)X\right]}{2(P_iX - Q_iR)^2} \quad (5.9)$$

5.3 Transmission System Voltage Stability Assessment

5.3.1 Field PMU Measurements

The PMU measured voltage magnitude at the Jojoba Station (JOA) in Salt River Project (SRP) system is studied. The data includes two 10-minute PMU measurements: one from 9:29:00 to 9:39:00, and the other from 14:08:00 to 14:18:00, all on the day of

July 21, 2011. Each of the two 10-minute measurements was related to a certain system disturbances.

Included in the morning measurements were two disturbances occurred at 9:29:05 and at 9:34:15 respectively. With respect to the afternoon measurements, there were two disturbances occurred at 14:08:45 and at 14:14:15 respectively.

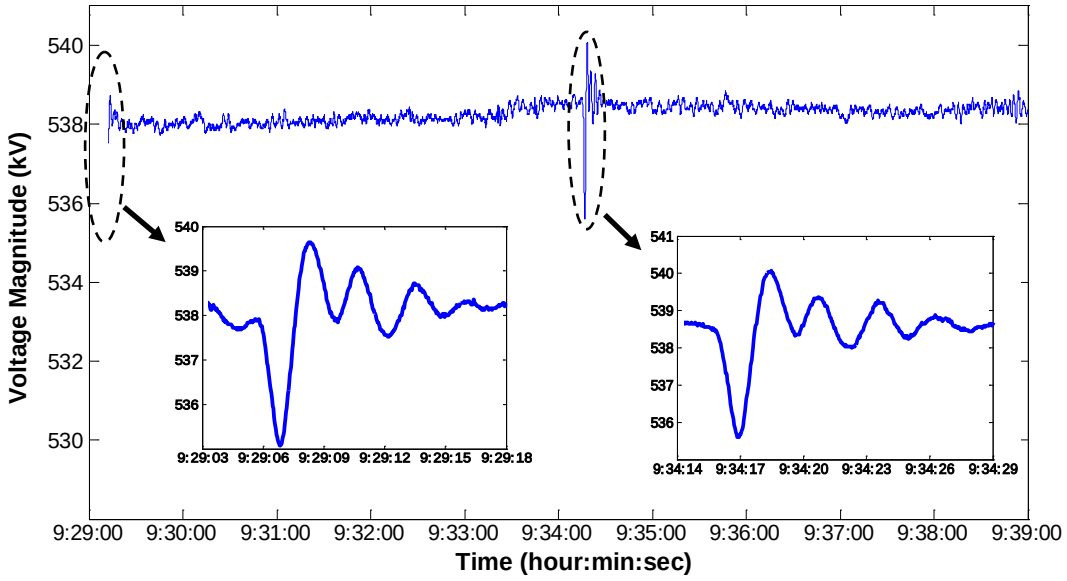


Figure 23 PMU measured voltage magnitude at the Jojoba Station (JOA): from 9:29:00 to 9:39:00

5.3.2 Voltage Stability Assessment

Using Equation (5.6) and the 10-minute morning measurements from the Jojoba Station (JOA) and the Kyrene Station (KY), the voltage stability index is calculated and shown in Figure 24. Figure 25 shows the computed voltage stability index using the 10-minute afternoon measurements.

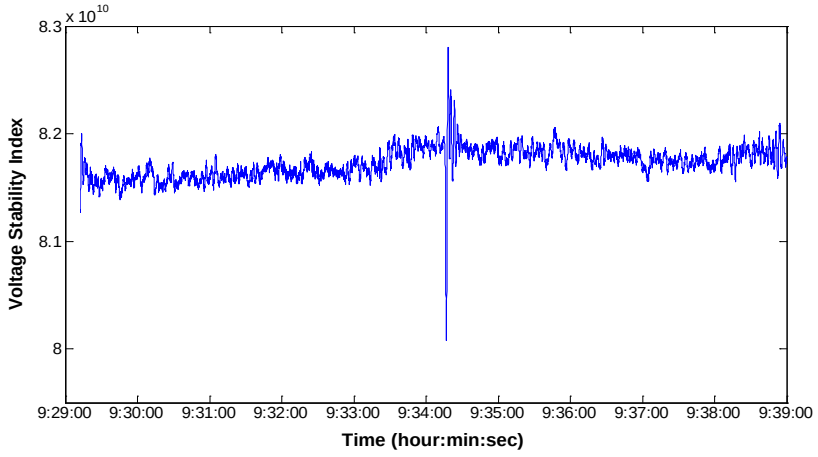


Figure 24 Voltage stability index calculated using the 10-minute morning PMU measurements from JOA and KY

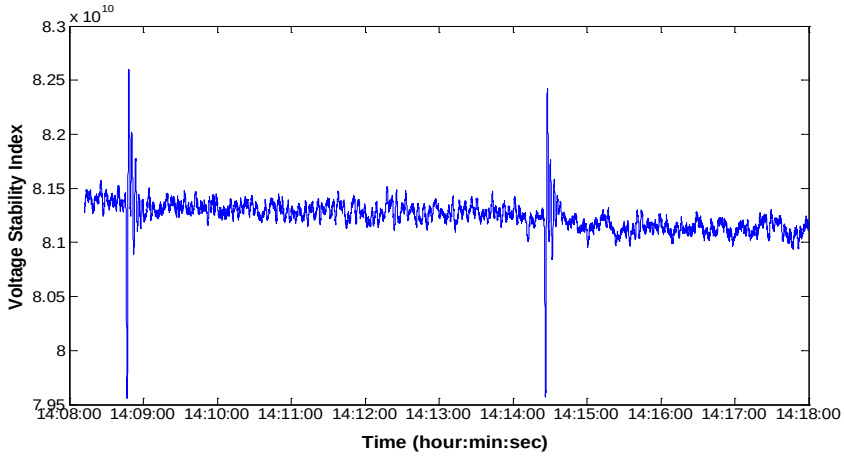


Figure 25 Voltage stability index calculated using the 10-minute afternoon PMU measurements from JOA and KY

It can be observed from Figure 24 and Figure 25 that during the two 10-minute

measurements, the computed voltage stability index stays well above zero, even under the situations when disturbances were occurred. This indicates that the transmission line between JOA and KY is far from its maximum power transfer limit for the given two periods of time, and the voltage profile in the vicinity of this small two-bus network is stable.

5.4 Impact Analysis of Wind Integration in Distribution System

5.4.1 Description of Test System

Figure 26 shows the single line diagram of a test distribution system. The switchable load at Bus 3 is fed by two power sources: one is the transmission system through a distribution feeder; the other is a local wind farm.

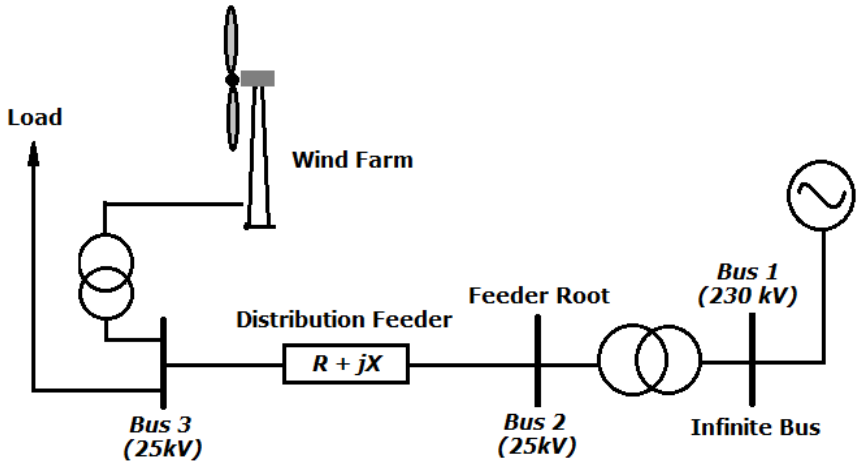


Figure 26 Single line diagram of the test system with wind farm integration

The feeder root is connected to the transmission system through a transformer

connected at Bus 1. The outside transmission system is considered as a power source (infinite bus) with a constant voltage of 230 kV. The voltage level of Bus 2 and Bus 3 is 25 kV. The feeder has a length of 25 km with an impedance of $2.88 + j9.896\ \Omega$.

The wind farm is connected to Bus 3 through a step-up transformer. To evaluate the impact of wind farm on local voltage stability, the cases of using two types of wind turbines will be used to perform the analysis: one is the conventional SCIG wind turbines and the other is DFIG wind turbines.

For the first case (turbine type), the wind farm contains three pairs of SCIG wind generators (WG). They are WG#1, WG#2 and WG#3. Each of the six turbines has a rated real power generation capacity of 1.5 MW. The wind farm has a rated total capacity of 9 MW. Besides the capacitor banks already integrated inside the wind farm, a switchable STATCOM is connected to the Point of Common Coupling (PCC) at Bus 3 for additional reactive power needs. The test scheme for this type of wind farm is shown in Figure 27.

For the second type, the wind farm is arranged the same as in Figure 27 except that the capacitor banks and STATCOM are removed. This is because the DFIG is able to control the terminal voltage through its electronic-based AC/DC voltage source converter, and no additional reactive power compensation devices are needed.

For both types of wind farm, the VSI is computed using local measurements at feeder root (Bus 2) in order to monitor the voltage stability of Bus 3.

Further, when configuring the test system, the control scheme of “Voltage Regulation” of the DFIG wind turbines is selected to maintain the terminal voltage level

at Bus 3.

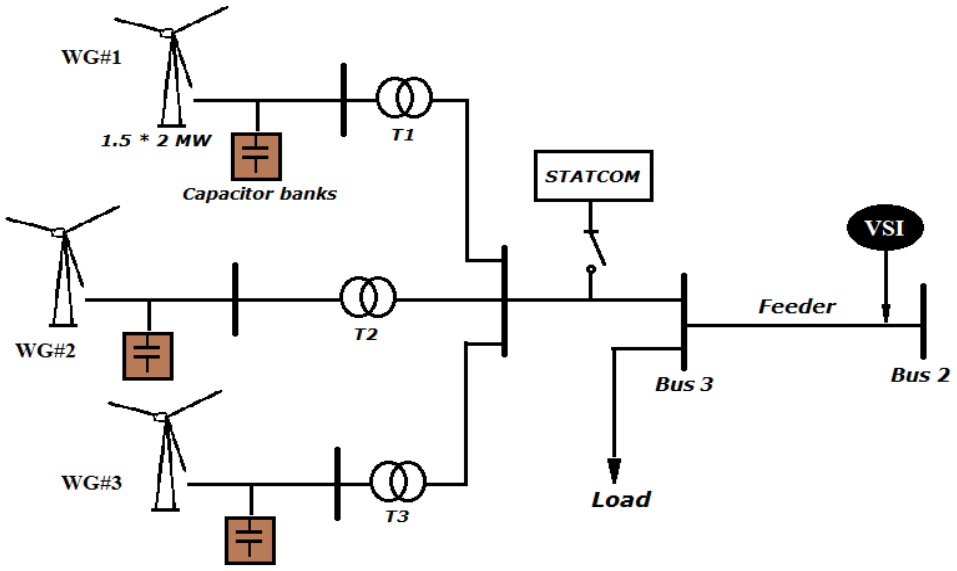


Figure 27 Details of the SCIG-based wind farm and reactive power compensators

5.4.2 Modeling of VSI

In our work, the Matlab/Simulink is used to establish the test system shown in Figure 26 and Figure 27. The block diagram of the proposed voltage stability index is shown in Figure 28. The real-time calculation of the index could be divided into three consecutive steps.

Firstly, the real-time bus measurements, as well as the feeder resistance and inductance are imported as inputs for the model to use. In this case, the required bus measurements include the voltage, current, as well as real and reactive power at Bus 2. Secondly, the data measurements are processed according to (5.8). The final procedure is to output the real-time value of the voltage stability index.

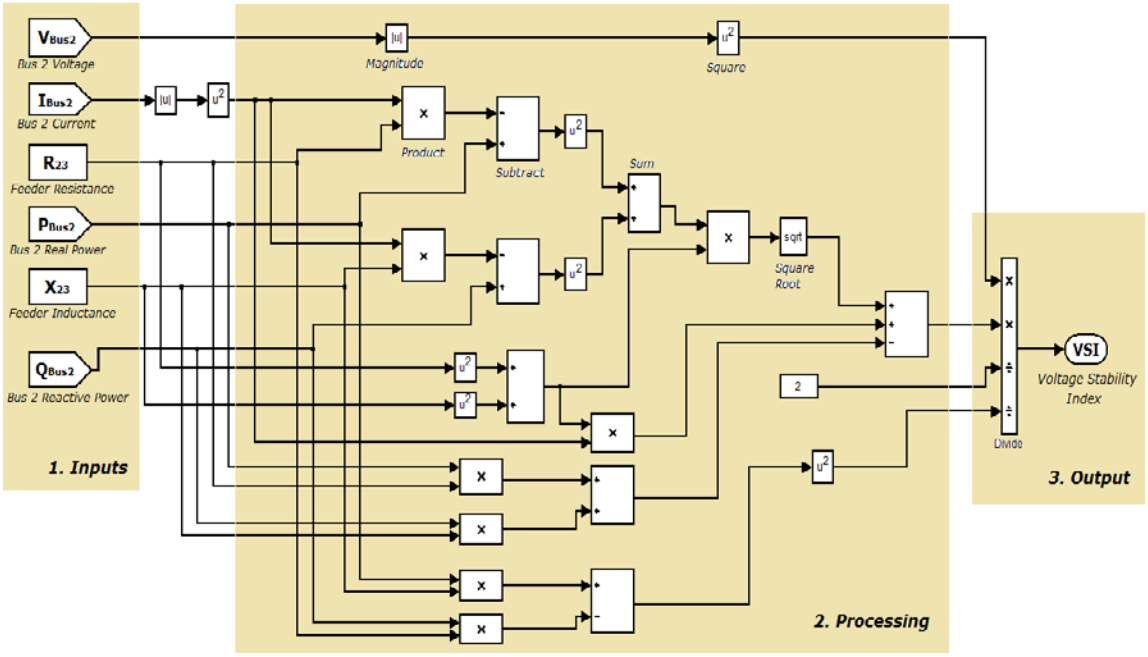


Figure 28 Voltage stability index implemented in Matlab/Simulink

5.4.3 Case Study: SCIG-based Wind Farm without STATCOM

This case will study the impact of SCIG-based wind farm on voltage stability. A wind gust is assumed to occur at t=2s. The wind speed at WG#1 increases from 8m/s to 11/s within 3 seconds. The same wind gust is applied to WG#2 and WG#3, respectively with 1.5 seconds and 3 seconds delays.

STATCOM are switched out first. Also disabled is the AC undervoltage protection. Simulation is implemented with the load switching in at t=12s, with a constant power demand of 20+j10 kVA. Simulation results are shown in Figure 29.

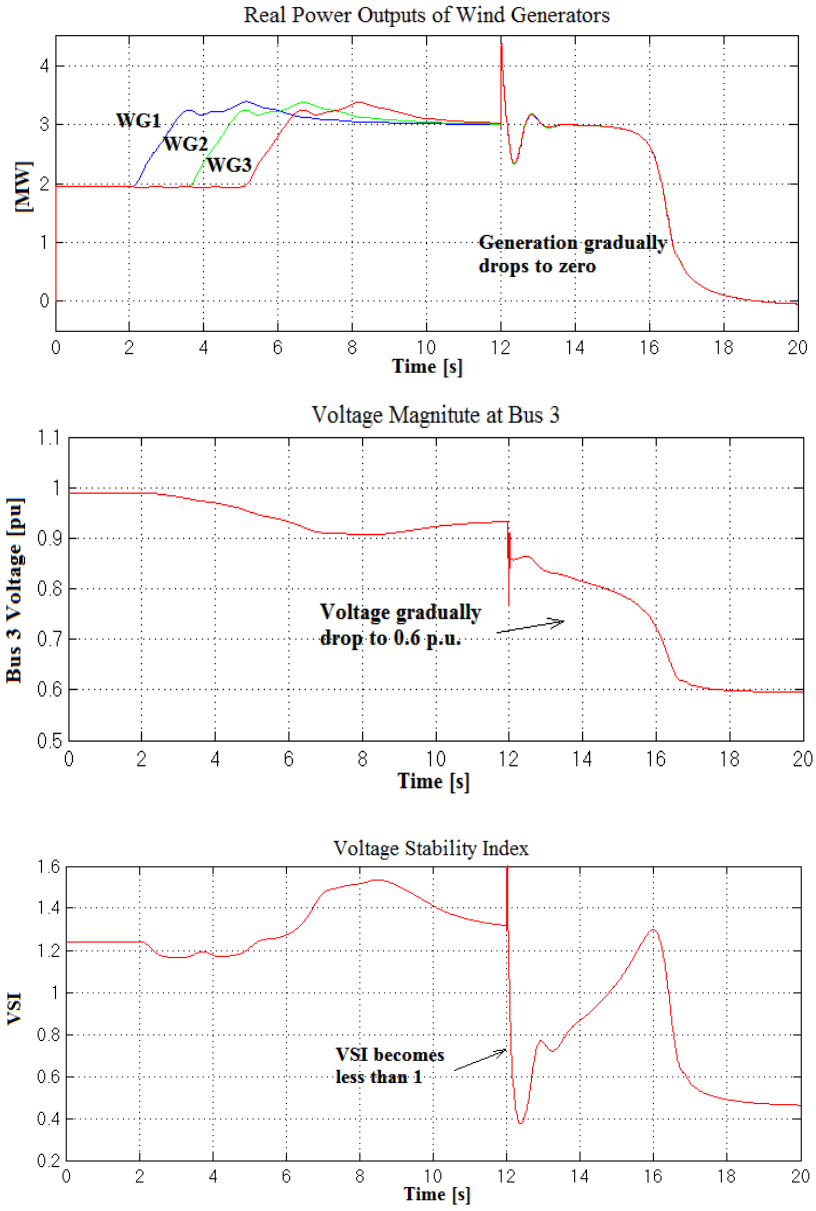


Figure 29 Simulation results of SCIG-based wind farm with the AC under-voltage with protection disabled

As is shown in Figure 29, the VSI becomes smaller than 1 soon after the load is switched in, which indicates that the voltage at Bus 3 is unstable. This could be verified

by checking the voltage measurement at Bus 3. The Bus 3 voltage magnitude has dropped to 0.6 p.u. within 8 seconds. This case demonstrates that the VSI could effectively detect the onset of voltage instability.

Another simulation is implemented by enabling the AC under-voltage protection scheme for wind turbines. Simulation results are shown in Figure 30.

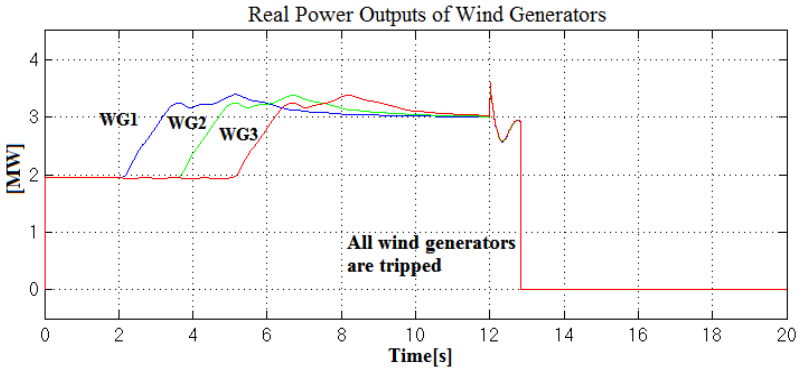


Figure 30 Simulation results of SCIG-based wind farm with the AC under-voltage protection enabled

As shown in Figure 30, about 0.9 seconds after the load is switched in, all of the wind generators have been tripped due to under-voltage protection. If one looks at the Bus 3 voltage magnitude, it can be found that the voltage magnitude has dropped to as low as 0.76 p.u. before the wind generators are tripped.

The main reason for the voltage drops in the two simulations shown above is that the SCIG-based wind farm absorbs reactive power from the utility grid while generating real power, which makes the wind farm a reactive power “sink”.

5.4.4 Case Study: SCIG-based Wind Farm with STATCOM

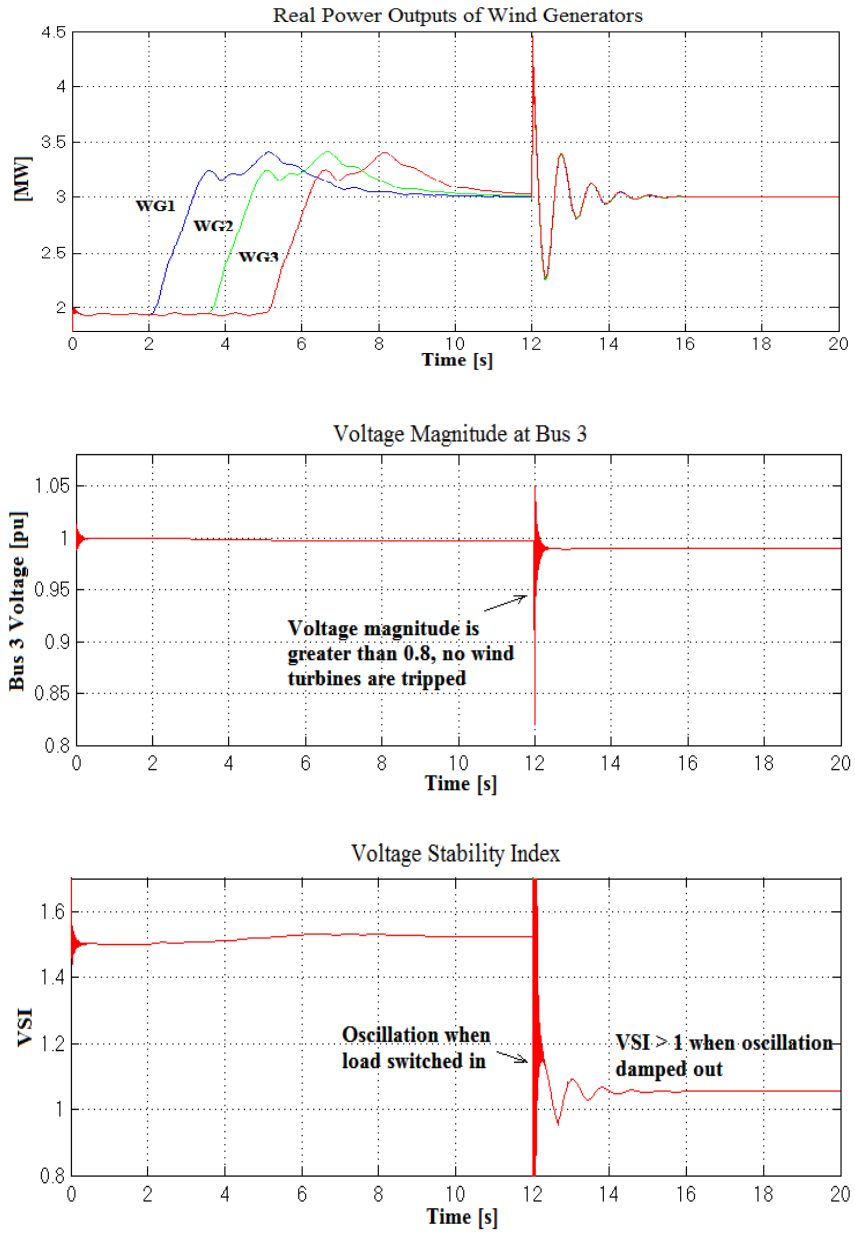


Figure 31 Simulation results of SCIG-based wind farm with load and STATCOM

In this case, a STATCOM is combined at the PCC (Bus 3) to provide reactive

power support. The converter rating of the STATCOM is set to be 30 MVA. The same load is switched in at t=12s. Figure 31 shows the simulation results.

As can be observed from Figure 31, although some oscillations of the power outputs occur, all of the wind turbines remain in service after the sudden load increase. This is because the STATCOM has provided sufficient reactive power to support the node voltage at Bus 3.

5.4.5 Case Study: DFIG-based Wind Farm

This case looks at the impact of DFIG-based variable speed wind turbines on voltage stability. The gust of wind occurs at t=3s and increases the wind speed from 8m/s to 14m/s in six seconds.

Since wind turbines using DFIG are equipped with built-in PWM-based voltage source converter (VSC), the STATCOM is no longer needed and is removed from the test system. By switching in the same load at t=12s, simulation results are presented in Figure 32.

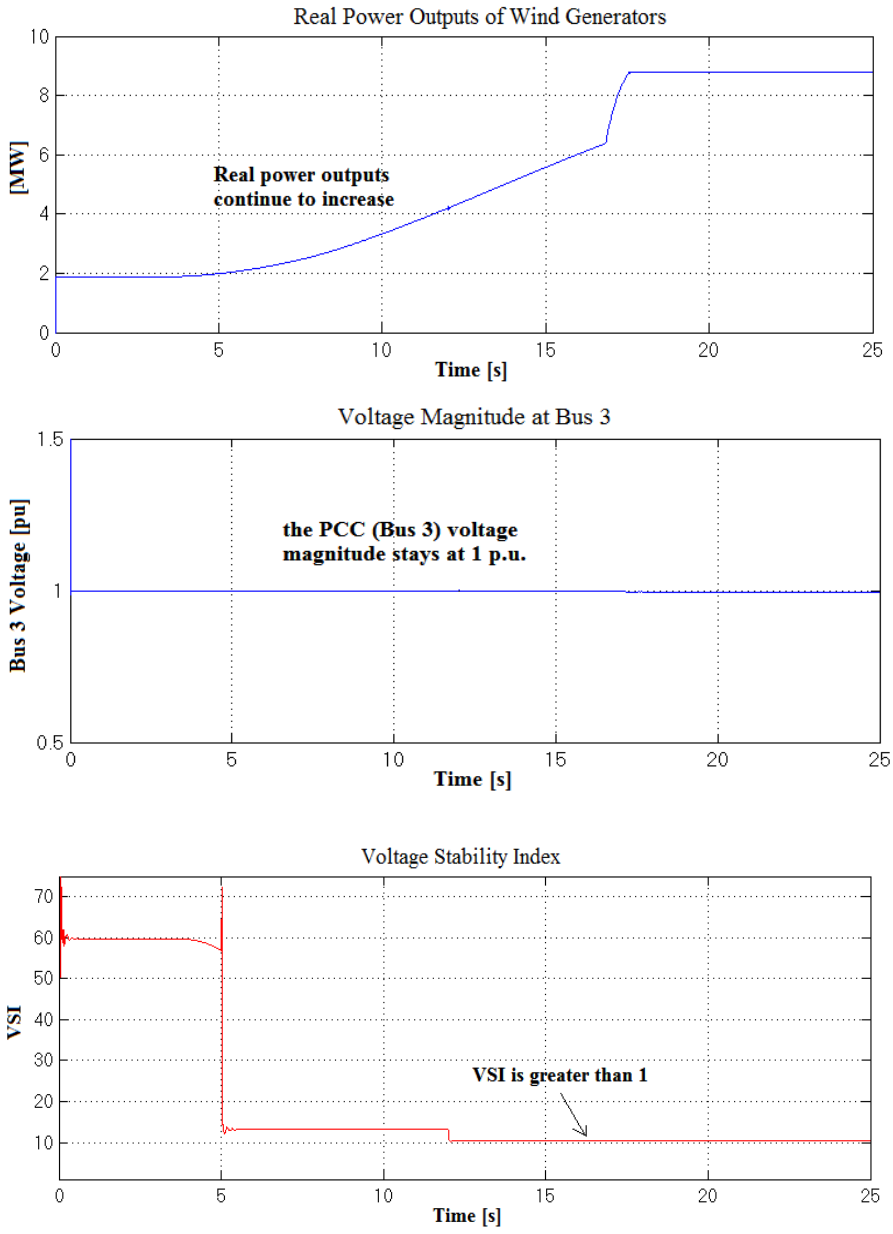


Figure 32 Simulation results of DFIG-based wind farm with load and without additional reactive power compensation

Comparing Figure 32 with Figure 29, it could be found that the built in VSC could dynamically regulate its grid side voltage and maintain normal operation of the

wind generators. Thus it performs better than the conventional SCIG-based wind turbines during gust of wind and sudden load increase.

5.5 Summary

The research conclusions reported in this section include:

- A voltage stability index (VSI) to quantify the voltage stability margin for a general two-bus transmission system using time-synchronized data of the line two ends is proposed and applied to field PMU measurements. Experimental results have demonstrated the efficacy of the proposed index;
- The VSI is modified to adapt to the situation of radial distribution system. Only local measurements are needed after the modification. No time synchronization is needed;
- The performance of both fixed-speed and variable speed wind turbine are examined in simulations. Load switching and wind gusts are considered in studying the corresponding node voltage responses. STATCOM is also included to demonstrate its capability for voltage stability improvement. Simulation results demonstrate that the fixed-speed wind turbine requires large amount of reactive power support, whereas the variable speed wind turbine could enhance node voltage stability without additional reactive power compensation devices;
- The proposed VSI is shown to have the capability of effectively quantifying node voltage stability margin in real time.

6. DATA MINING FOR SMALL DISTURBANCE VOLTAGE STABILITY ANALYSIS CONSIDERING WIND UNCERTAINTY*

6.1 Problem Formulation

As wind generation continues to expand in size and penetration level, a deeper understanding of its dynamic behavior and impact on system stability becomes necessary. Since the wind farm production is primarily determined by wind speed and thus fluctuating constantly, one of the most important studies is to investigate the dynamic phenomena induced by the variation of wind generation.

Historically power system stability has been associated with the generator rotor angle dynamics. Stimulated by several major voltage collapses, the framework of power system voltage stability as defined in [6] was originated around 1980s [66] [67] and had been extensively studied in 1990s [68] [69]. Voltage stability can be further classified into small disturbance and large disturbance categories. The small disturbance voltage stability refers to the system's ability to maintain steady voltage levels following small disturbances experienced through continuous changes in load [70] [71]. From this view the small disturbance voltage stability is predominantly load stability. With the recent rapid growth of wind generation, operational uncertainty will extend from demand side variability to a significant portion of the supply side variability as well, which will

---

* Part of the material in this section is reprinted from "Impact of wind generation uncertainty on power system small disturbance voltage stability: a PCM-based approach," by Ce Zheng and Mladen Kezunovic, *Electric Power Systems Research*, vol. 84, no. 1, pp. 10-19, March 2012, DOI: 10.1016/j.epsr.2011.10.001

impact system dynamic performance and cause voltage deviations. As a result, it is imperative to take into account the stochastic nature of wind farm output in voltage stability study.

From the end of the last century, the wind generators based on fixed-speed wind turbines have been included in stability studies [72]. Later, the grid stability enhancement of the new variable-speed wind generators, especially the Doubly-Fed Induction Generators (DFIG), has been reported in several publications [73]-[75]. Recently, research efforts to study how further large scale integration of variable-speed wind generators will influence the power system stability and electricity market have been proposed [76].

While such studies are important in their own right, the important issue of how the system small disturbance voltage stability may be influenced by the constantly changing wind generation has not been fully explored yet.

A review of literature reveals that several studies have been reported in the related area. References [77] and [78] conducted modal and eigenvalue sensitivity analysis on grid-connected DFIGs. The work reported in [79] explored the relationship between uncertain wind generation and system probabilistic small-signal stability analysis via Monte Carlo simulation method.

As very few tools are available to analyze the parameter uncertainties in time-domain simulations, Tatang et al. developed the Probabilistic Collocation Method (PCM) to mathematically describe the system response in terms of uncertain parameters [80]. This method has been successfully applied to time-domain simulation studies in the

area of global climate change [81]. Later in [82], Hockenberry et al. have introduced the Probabilistic Collocation Method in power system dynamic analysis and discussed its applicability in load uncertainty analysis. More studies have been reported to evaluate the advantage of PCM over other uncertainty analysis techniques in the time-domain analysis of power system load parameter uncertainties [83].

Normally the uncertainty analysis is performed under one of the following situations:

1) The relationship between the uncertain parameter and the output of interest is known analytically;
2) The above relationship is unknown. A model of the "black box" needs to be approximated first.

The PCM is designed to address the second situation. This approach allows the use of nonlinear models and evaluation of complicated output functions. It is particularly appealing because it requires a much smaller number of simulations to reach an accurate approximation which may take hours or days for the traditional techniques.

This research studies how wind generation variation will impact voltage stability under the new energy transfer scenario where traditional generators are supplanted by variable-speed wind generators. Since the relationship between uncertain wind generation and system small disturbance voltage stability is not analytically clear, the PCM is introduced to address the problem. It begins with the mathematical definitions of the collocation method. Then, the wind-connected power system dynamic modeling and small disturbance analysis procedures are provided in Section 3. Section 4 and Section 5

constitute the key aspects of the work. First, a simple system is established to explore the applicability of implementing PCM in the wind generation uncertainty analysis. Next, a more complicated 23-bus power system is presented. Deeper understanding is obtained by monitoring the eigenvalue movement and voltage instability induced by the variation of wind farm output. The computation efficiency of PCM has been demonstrated by comparing it with traditional simulation based approaches.

6.2 Probabilistic Collocation Method

The basic idea of PCM is to approximate the relationship between uncertain parameters of the system and the outputs of interest through polynomial models. Based on the probability density function of the uncertain parameters, the concepts of Orthogonal Polynomials and Gaussian Quadrature Integration are incorporated to solve for polynomial approximation functions. Once the polynomials are obtained, collocation methods are generated to solve for the model coefficients. One major advantage of PCM is that only a handful of simulations are needed to determine the approximation model.

6.2.1 PCM with Multiple Inputs

To be general, let $x_1, x_2, \ldots, x_n$ be the uncertain parameters. Suppose a system is represented by a complex, high-ordered, or even "black-box" model. Its response in terms of the uncertain parameters is expressed as:

$$U = P(x_1, x_2, \cdots, x_n) \tag{6.1}$$

where U is the output of interest (system response). The objective of PCM is to find the

following approximation of U:

$$\hat{U} = C_0 + \sum_{i=1}^{n}\left[C_{i1}p_{i1}(x_i) + C_{i2}p_{i2}(x_i) + \cdots + C_{im}p_{im}(x_i)\right]$$

$$+ \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\left[C_k\, p_{i1}(x_i)\, p_{j1}(x_j)\right] \tag{6.2}$$

where $\hat{U}$ is an approximation of U, m is the order of this polynomial model, C_0, C_{i1}, …, C_{im}, C_k are model coefficients, and $p_{i1}(x_i)$, $p_{i2}(x_i)$, …, $p_{im}(x_i)$ are polynomial functions in terms of each uncertain parameter x_i.

6.2.2 Solving for Polynomials

What we need then is to find the set of polynomials and coefficients listed in (6.2). The polynomials could be derived by deploying the concept of Orthogonal Polynomials [23]. The definition of orthogonal polynomials is:

$$\int_x P(x)H_i(x)H_j(x)dx = \begin{cases} 1, & if\ i = j \\ 0, & if\ i \neq j \end{cases} \tag{6.3}$$

where $P(x)$ is user-defined weighting function of x, $H_i(x)$ and $H_j(x)$ are orthogonal polynomials of x with the order of i and j ($i, j=0, 1, \ldots$). Equation (6.3) suggests that the inner product of any two orthogonal polynomials of different order is always zero.

Assume the probability density function of the i^{th} uncertain parameter is $f(x_i)$. By substituting the weight function $P(x)$ with $f(x_i)$, and using the following definition:

$$H_{-1}(x) = 0,\ H_0(x) = 1 \tag{6.4}$$

the remaining higher-order orthogonal polynomials can be derived one by one. Fitting the derived orthogonal polynomials to $p_{i1}(x_i)$, $p_{i2}(x_i)$, …, $p_{im}(x_i)$ of (6.2), only the

coefficients are left to be solved.

6.2.3 Solving for Coefficients

As long as the polynomial functions in (6.2) are known, the model coefficients can be calculated by feeding different inputs into the system and recording corresponding system response. Suppose the system has *n* uncertain parameters, and we are using a PCM model with the order of *m*, the sets of inputs that are needed will be:

$$1 + n \times m + \binom{n}{2} \tag{6.5}$$

Take the linear PCM model with single uncertain parameter *x* as an example, Equation (6.2) could be rewritten as:

$$\hat{U} = C_0 + C_1 p_1(x) \tag{6.6}$$

What we need next is to feed the real system with two different values of parameter *x*, and substitute $\hat{U}$ with the real system response *U* of each run. Thus the coefficients C_0 and C_1 in (6.6) could be solved.

In the above linear model example, the two different input values are also called collocation points. It should be noted that the selection of collocation points has significant impact on the accuracy of model approximation. In order to find a good approximation for the PCM model with smallest number of model runs, the Gaussian Quadrature Integration [84] approach is deployed: while selecting the collocation points, the points for the model runs from the roots of the next higher order orthogonal polynomial will be selected for each uncertain parameter. This approach enables the

collocation points spanning the high probability regions of their distribution and capturing as much of the behavior of system response as possible.

The whole idea of applying PCM in power system uncertainty analysis is demonstrated in Figure 33. Selected set of uncertain parameters is fed into the full dynamic model of a power system and corresponding output of interest is recorded. The relationship between uncertain parameters and system response is then approximated through PCM once sufficient tests are conducted. In this work, our output of interest is the "voltage stability indicator", which essentially monitors the stability of system voltage following small disturbances experienced through continuous changes in wind farm production [85]. An eigenvalue-based indicator is defined in the follow-up sections.

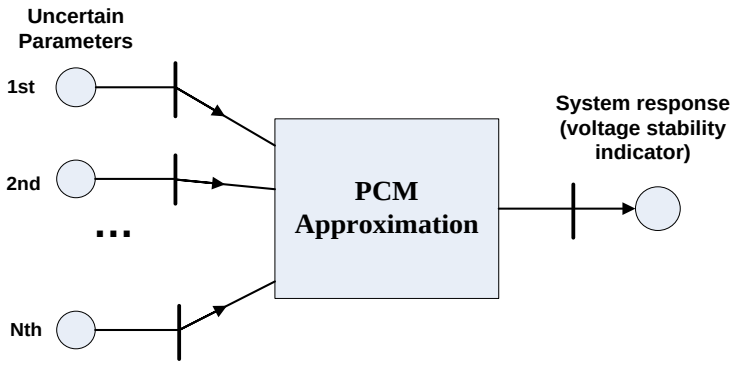


Figure 33 PCM approximation used for voltage stability analysis

6.3 Small Disturbance Voltage Stability

Voltage stability has long been evaluated by both static and dynamic approaches while it essentially has dynamic nature. This work investigates the small disturbance (SD) voltage stability problem caused by fluctuation in wind generation from the

viewpoint of its dynamical mechanism.

In a power system, on-load tap changing transformers and thermal units in charge of load frequency control are working slowly in a time frame of minutes, while generators, automatic voltage regulators, governors and induction motor loads respond much faster. Therefore, the power system can be approximated by two simplified systems: slow and fast response subsystems. Their dynamic behavior can be described by a set of differential equations:

$$\dot{x}_S = f_S(x_F, V, u) \tag{6.7}$$

$$\dot{x}_F = f_F(x_S, x_F, V, y, z, u, load) \tag{6.8}$$

where x_S and x_F represents system slow and fast response state variables respectively, u represents the controllable variables, *load* represents uncontrollable variable vector which governs load consumption, V is the bus voltage vector, y is the node specification vector in load flow equation, and z is the dependent variable vector. Equations (6.7) and (6.8) stand for all the dynamic characteristics existing in the whole system. The static relationships in loads and generating units, etc., are usually described by a set of algebraic equations:

$$0 = g_F(x_S, x_F, V, y, z, u, load) \tag{6.9}$$

$$y = g_N(x_S, x_F, V, u) \tag{6.10}$$

A common approach to evaluate the voltage stability of a fast subsystem is through linearization of its differential equations and eigenvalue analysis. Slow variables x_S are treated as constant. Fixing u and *load*, equations (6.8) (6.9) and (6.10) are

linearized around system equilibrium point by eliminating both *y* and *z*:

$$\begin{bmatrix} \Delta\dot{x}_F \\ 0 \end{bmatrix} = \begin{bmatrix} H_F & H_V \\ Y_F & J_N - Y_V \end{bmatrix} \begin{bmatrix} \Delta x_F \\ \Delta V \end{bmatrix} \tag{6.11}$$

The state-space matrices H_F, H_V, Y_F, J_N and Y_V are defined in [68]. By eliminating ΔV, the linearized fast response subsystem could be presented as:

$$\Delta\dot{x}_F = \left[H_F - H_V\left(J_N - Y_V\right)^{-1} Y_F\right]\Delta x_F = A_{FF}\,\Delta x_F \tag{6.12}$$

The matrix A_{FF} is formulated from system differential algebraic equations (DAE), which are detailed by dynamic modeling of each network component.

6.3.1 DAE of Doubly-Fed Induction Generator

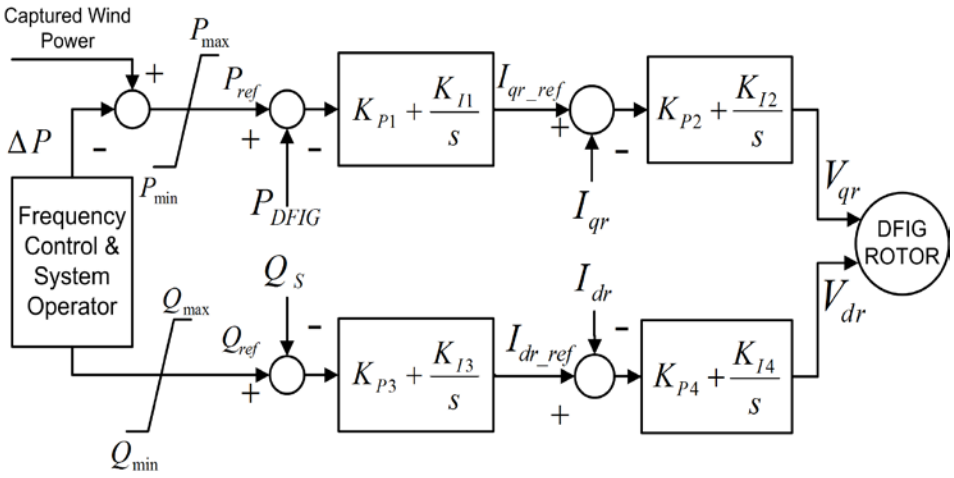


Figure 34 DFIG frequency (top) and reactive power (bottom) controllers

Dynamic modeling of the DFIG-based wind generator is realized by modeling of its several components: wind turbine, shaft, generator, converter, and control system [9].

While there is a general acceptance of models for wind turbine, drive train (shaft), and induction generator, a variety of modeling schemes for the frequency and VAR controllers are being used nowadays. The DFIG frequency and VAR controllers adopted in this work are represented in Figure 34.

The differential equations to describe the dynamic behavior of a DFIG are derived and listed in (6.13) and (6.14):

$$\begin{cases} \dot{E}_q^{'} = -\dfrac{1}{T_0^{'}}\left[E_q^{'} + (X_s - X_s^{'})I_{ds}\right] + \left[\omega_s \dfrac{X_m}{X_r} V_{dr} - (\omega_s - \omega_r)E_d^{'}\right] \\ \dot{E}_d^{'} = -\dfrac{1}{T_0^{'}}\left[E_d^{'} + (X_s - X_s^{'})I_{qs}\right] + \left[-\omega_s \dfrac{X_m}{X_r} V_{qr} + (\omega_s - \omega_r)E_q^{'}\right] \\ \dot{\omega}_r = \dfrac{\omega_s}{2H}\left(T_m - E_d^{'} I_{ds} - E_q^{'} I_{qs}\right) \end{cases} \tag{6.13}$$

$$\begin{cases} \dot{m}_1 = K_{I3}\left[Q_{ref} - (Q_r + Q_s)\right] \\ \dot{m}_2 = K_{I4}\left\{K_{P3}\left[Q_{ref} - (Q_r + Q_s)\right] + m_1 - I_{dr}\right\} \\ \dot{m}_3 = K_{I1}\left[P_{ref} - (P_r + P_s)\right] \\ \dot{m}_4 = K_{I2}\left\{K_{P1}\left[P_{ref} - (P_r + P_s)\right] + m_3 - I_{qr}\right\} \end{cases} \tag{6.14}$$

In (6.14), $m_1 \sim m_4$ are state variables associated with the reactive power and frequency PI controllers. The definition of the state variables and constants included in (6.13) and (6.14) could be found in [78]. The algebraic equations of DFIG are derived in the Appendix.

6.3.2 DAE of Electric Grid

Using the dynamic modeling process described in [4], the DAEs for the remaining components of the electric network, including synchronous machines, other

types of generators, induction motors, FACTs devices, load and tap-changing transformers are formulated. By combining the network DAE with the set of DFIG DAE, the overall set of state variables x_F, as well as matrices H_F, H_V, Y_F, J_N, Y_V and A_{FF} in (6.11) and (6.12) can be determined.

6.3.3 Small Disturbance Voltage Stability Evaluation

The SD voltage stability analysis essentially monitors the eigenvalue trajectory of the matrix A_{FF} in (6.12). For a given wind generation level, using the DAEs formulated in the previous steps, a system state matrix A_{FF} can be uniquely decided. Calculating the eigenvalues of A_{FF} at each operating point, the system SD voltage stability can be evaluated by checking if all the eigenvalues are located in the left hand half of the complex plane. The fast response subsystem is losing voltage stability when one of the real parts of the eigenvalues of A_{FF} vanishes. The critical voltage stability point is identified when one of the eigenvalues reaches imaginary axis.

Note that although the above treatment is generally accepted and widely used, there is still a lack of mathematical validation of the treatment. However, rigorous numerical simulation studies have demonstrated that the singularity or sign change of determinant of system Jacobian matrix always indicates the loss of voltage stability [3-5]. Hence the above eigenvalue monitoring approach seems valid when studying the voltage problems in a power system. To verify the above treatment, a linear voltage stability indicator detailed in [86] will be used to compare the results of the proposed eigenvalue-based indicator.

6.4 Implementation of PCM in a Simple System

In this section, the collocation method will be applied in the uncertainty analysis of a simple four-bus system to demonstrate its validity. The system one-line diagram is illustrated in Figure 35, in which the DFIG-based wind farm is connected at Bus 3.

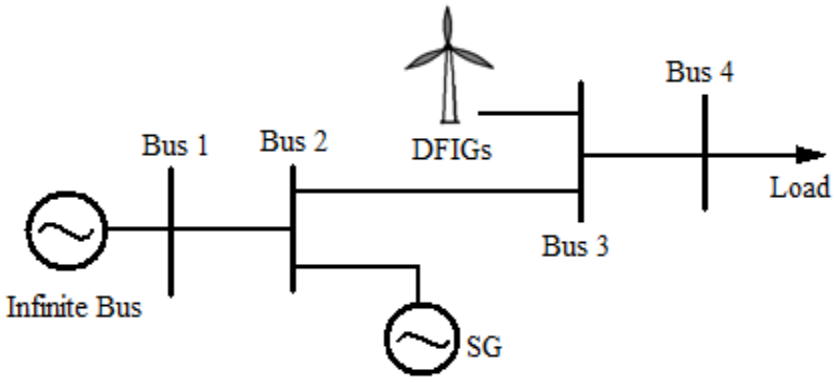


Figure 35 Four-bus test system with a local wind farm at Bus 3

6.4.1 Parameter Specification

It is necessary to determine the uncertain parameters and corresponding output of interest in Figure 33 before conducting uncertainty analysis. In the test system, the wind speed v_{wind} is selected as the single uncertain parameter. As discussed earlier, the real part of the critical eigenvalues Eig_{real} of matrix A_{FF} is able to indicate the SD voltage stability. Hence it is selected as the output of interest. The relationship between v_{wind} and Eig_{real} is analytically clear. The time-varying wind speed will constantly change the nodal injection of the wind farm at its Point of Common Coupling (PCC), which has significant impact on the system load flow and oscillation modes. Therefore the variation of wind speed will influence the distribution of critical eigenvalues on the complex

plane.

The DFIG dynamic model presented in the previous section is adopted, with the rated power, cut-in, cut-out and rated wind speeds specified as 3.6 MW, 4 m/s, 20.9 m/s and 12.9 m/s respectively.

Also, the Probability Density Function (PDF) of the uncertain parameter needs to be specified beforehand. In our research, the wind speed with a Gaussian distribution is assumed first. This is only for the purpose of illustrating the calculation process of PCM, although it may not be realistic. A more widely-used Weibull distribution will be adopted and examined in the next section. With the combination of energy storage, the wind farm injection instead of the wind speed will be taken as the input parameter for PCM. Under this situation, PDFs other than the Weibull distribution, such as the normal distribution, need to be considered.

The PDF of a Gaussian distribution is:

$$f(v_{wind}) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(v_{wind}-\mu)^2}{2\sigma^2}} & v_{wind} \geq 0 \\ 0 & v_{wind} < 0 \end{cases} \tag{6.15}$$

When $v_{wind} \geq 0$, a mean of μ=9 and a standard deviation of σ=2 are assumed.

6.4.2 Polynomial Approximation

What we need then is to approximate the relationship between the uncertain parameter v_{wind} and system response Eig_{real} based on the PDF of wind speed. A "variable transformation" is applied in PCM: the random variable X with any kind of Gaussian

distribution will be represented by the following transformation:

$$X = \mu + \sigma[p_1(\delta)] \tag{6.16}$$

where $p_1(\delta)$ is the first order orthogonal polynomial of the standard normal distribution δ, which has a mean value of $\mu_\delta=0$ and a standard deviation of $\sigma_\delta=1$. The advantage of this transformation lies in the convenience of using same orthogonal polynomials to stand for any Gaussian distribution [80] [81].

The Hermite Polynomials designed for standard normal distribution are deployed to derive orthogonal polynomials. The first to sixth order polynomial expressions and their respective roots are listed in Table 9. The roots of the involved polynomials are listed for the Gaussian Quadrature Integration method to generate collocation points.

Table 9 $1^{st} \sim 6^{th}$ Order Orthogonal Polynomials for Standard Gaussian Distribution δ

Order	Orthogonal Polynomials	Roots
1^{st}	δ	{0}
2^{nd}	$\delta^2 - 1$	{-1, 1}
3^{rd}	$\delta^3 - 3\delta$	{-1.7321, 0 , 1.7321}
4^{th}	$\delta^4 - 6\delta^2 + 3$	{-2.3344, -0.7420, 0.7420, 2.3344}
5^{th}	$\delta^5 - 10\delta^3 + 15\delta$	{-2.8570, -1.3556, 0, 1.3556, 2.8570}
6^{th}	$\delta^6 - 15\delta^4 + 45\delta^2 - 15$	{-3.9694, -3.4839, -j3.5908, j3.5908, 3.4839, 3.9694}

Equation (6.2) will be solved using the generated collocation points and

corresponding system responses. Note that for the case of a single uncertain parameter and m^{th} order PCM model, the number of required collocation points, according to (6.5), is $m+1$.

The algorithm steps and pseudocode to obtain the polynomial model between the uncertain parameter v_{wind} and system response Eig_{real} are summarized in Table 10 (description of how to calculate SSR for model error evaluation will be introduced later).

Table 10 Algorithm Steps: Polynomial Approximation Between v_{wind} and Eig_{real}

1. Choose threshold ε
2. Specify the PDF of v_{wind}: $f(v_{wind})$
   Perform variable transformation (e.g. Eq. (6.16)) if needed
3. Substitute into Equation (6.3) and Equation (6.4):
   $\int_x f(v_{wind})H_1(x)(1)dx = 0$, $\int_x f(v_{wind})H_2(x)H_1(x)dx = 0$, $\cdots$
4. Solve for the orthogonal polynomials: $H_1(x)$, $H_2(x)$, $\cdots$, $H_N(x)$
5. Repeat:
   for k=1→N do
      find the roots of $H_{k+1}(x)$: $R_1, R_2, \cdots, R_{k+1}$
      run model on collocation points $R_1, R_2, \cdots, R_{k+1}$
      substitute $\widehat{U}$ in Equation (6.2)with the model output Eig_{real}
      solve the coefficients of Equation (6.2): $C_0, C_1, \cdots, C_{k+1}$
   end for
      SSR ≤ ε (SSR calculated from Equation (6.19))
6. Use approximation for uncertainty analysis

6.4.3 Discussion

Although the nodal injection power of a wind farm is primarily determined by the on-site wind speed, several other factors may influence its value in normal practice. For example, wind generation may vary along with the real time electricity prices. The setting of wind turbine controllers may be switched between terminal voltage control mode and power factor control mode. System operators may request the wind farm to adjust its output level according to the frequency control requirements. In addition, the system voltage stability is dependent on a number of stochastic system parameters such as load fluctuations and equipment outages. Under these situations, the problem is more complicated and system response is harder to predict. However, it is still possible for PCM to approximate the relationship by taking into account multiple uncertain parameters. The polynomial model will be expanded according to equations (6.1) and (6.2) described in Section 6.2.1.

Suppose we are considering two uncertain parameters, that is, load total active power demand P_L and wind power injection P_{wind}. The new PCM expansion will be:

$$\widehat{U} = C_0 + C_1 H_1(P_L) + C_2 H_1(P_{wind}) + C_3 H_2(P_L) + \\ + C_4 H_2(P_{wind}) + \cdots + C_n H_1(P_L) H_1(P_{wind}) \tag{6.17}$$

As shown in (6.17), for a model which is second-order or higher, a cross-product term is added to the end of the expansion for higher approximation accuracy. In the above example, the collocation points (roots) will be derived for P_L and P_{wind} separately. Different pairs of roots will be formed using different combinations. The ranking of the combinations is arranged according to the probability of the involved roots. The first pair

will always be the highest probability root for each parameter, which is the so called "anchor point". Also, one more input pair is needed corresponding to the cross-product term. For this we use the second-highest probability root of P_L and P_{wind} respectively.

Nevertheless, the wind speed is chosen as the single uncertain parameter in this research for the purpose of not adding complexity to the problem. The merit of doing this is that the relationship between v_{wind} and Eig_{real} could be clearly identified. Further, the power factor control for all DFIGs in a wind farm is assumed, which complies with the grid code of most power systems in the United States.

The incorporation of energy storage units help address uncertainty in wind generation and modern forecasting technique is sufficiently mature that wind speed could be predicted with acceptable accuracy. This helps to relieve the stochastic nature of wind farm output, yet it does not contradict with the proposed method:

- The supply side variability introduced by renewable generations shall not be overlooked. Energy storage system compensates uncertainty; however it does not eliminate the variation. Operational constraints such as limited capacity of storage units limit the maximum generation commitment that can be met reliably by a wind farm;
- Knowing exactly how much power will be generated can actually help to take advantage of the PCM technique, that is, the PDF of wind farm output will be less dependent on the on-site historical data but can be extracted from the forecasted values;
- The proposed method is able to establish a computationally less expensive link

between the wind farm output and system SD voltage stability. As engineers try to capture more detail and complexity of the power system dynamic processes, the models become more complex to run. This simplification makes the proposed method particularly appealing.

We will next discuss the difference of the PCM model when wind speed falls in different zones divided by the DFIG cut-in speed v_{in}, cut-out speed v_{out} and rated wind speed v_{rated}. The mechanical power extracted from the wind turbine is given as:

$$P_m = \begin{cases} 0 & v_{wind} < v_{in}; v_{out} \le v_{wind} \\ g(v_{wind}) & v_{in} \le v_{wind} < v_{rated} \\ P_{rated} & v_{rated} \le v_{wind} < v_{out} \end{cases} \tag{6.18}$$

where $g(v_{wind})=1/2\rho A_{wt}C_p(\lambda,\theta)v^3_{wind}$, in which ρ is the air density, A_{wt} is the wind turbine swept area, and C_p is function of tip speed ratio λ and pitch angle θ [72]. An observation of (6.18) is that the system response should be discussed differently as the way the mechanical power is divided. This means although the PDF of wind speed is continuous, the system may respond in a discrete manner. So the relationship between v_{wind} and Eig_{real} should be treated carefully in PCM. The system response Eig_{real} should be modeled as:

$$Eig_{real} = \begin{cases} P_1(v_{wind}) & S1: v_{wind} < v_{in} \ \& \ v_{out} < v_{wind} \\ P_2(v_{wind}) & S2: v_{in} < v_{wind} < v_{rated} \\ P_3(v_{wind}) & S3: v_{rated} < v_{wind} < v_{out} \end{cases} \tag{6.19}$$

where P_1, P_2 and P_3 are PCM models under the situations of *S1*, *S2* and *S3* respectively.

The formulation for first to third order models under the situation of *S2* is given in Table 11. The dominant state variables of the critical eigenvalues are speed deviation

ω_{SG} and angle deviation θ_{SG} of the synchronous generator at Bus 2. It can be observed that in 3rd and 4th order models, sometimes we need to obtain Eig_{real} even when v_{wind} is out of the interval of *S2*. In such cases, we still assume that the wind generation keeps increasing/decreasing even if the wind speed goes beyond the rated speed or drop below the cut-out speed. This makes sure the model approximation could accurately reflect the system dynamic behavior within *S2*.

6.4.4 Error Evaluation

By feeding the generated collocation points into the test system, the corresponding critical eigenvalues are obtained. The coefficients of the PCM model could then be solved by substituting into (6.2). The calculated coefficients are also listed in Table 11.

Last but not least, the approximation error should be evaluated to guarantee the model accuracy. In this work, the Sum-Square-Root (SSR) error is calculated:

$$SSR = \sqrt{\frac{\sum_{i=1}^{j}\left[(\hat{Y}_i - Y_i)^2 \times f(\delta_i)\right]}{f(\hat{\delta}) \times j}} \tag{6.20}$$

where Y is the real system response and $\hat{Y}$ is the PCM approximation, j is the number of collocation points generated for error evaluation, $f(\delta_i)$ is the PDF at the value of δ_i, and $\hat{\delta}$ is the collocation point with highest probability.

To check the error, the collocation points derived from the next higher order orthogonal polynomials are used to run the model several more times. The error evaluation results for 1st ~ 3rd order models are shown in Table 12.

Table 11 Collocation Points, Critical Eigenvalues, and Calculated Coefficients for 1^{st} ~ 4^{th} Order Model Under the Second Situation S2

Order	Collocation Points	Corresponding Wind Speed (m/s)	Real Part of Critical Eigenvalues $Re\{\lambda(A_{FF})\}$	Calculated Coefficients
1^{st}	-1	7	0.08200	$C_0 = -0.0697$ $C_1 = -0.1517$
	1	11	-0.22130	
2^{nd}	-1.7321	5.5358	0.39060	$C_0 = -0.0521$ $C_1 = -0.1920$ $C_2 = 0.0551$
	0	9	-0.10702	
	1.7321	12.4642	-0.27450	
3^{rd}	-2.3344	4.3312	0.60692	$C_0 = -0.0694$ $C_1 = -0.1715$ $C_2 = 0.0492$ $C_3 = -0.0100$
	-0.7420	7.5160	0.01757	
	0.7420	10.4840	-0.20061	
	2.3344	13.6688	-0.30839	
4^{th}	-2.8570	3.2860	0.82942	$C_0 = -0.0544$ $C_1 = -0.1835$ $C_2 = 0.0473$ $C_3 = -0.0040$ $C_4 = -0.0018$
	-1.3556	6.2888	0.18819	
	0	9	-0.10702	
	1.3556	11.7112	-0.24875	
	2.8570	14.7140	-0.33729	

Table 12 Error Evaluation for 1st ~ 3rd Order PCM Approximation Under the Situation of S2

Order	Expected $\hat{Y}$	SSR
1^{st}	-0.0697	0.0602
2^{nd}	-0.0521	0.0274
3^{rd}	-0.0694	0.0075

A larger SSR indicates a larger deviation of PCM model approximation from real system behavior and vice versa. In this case an error criterion of 1×10^{-2} is assumed. From Table 12 it can be concluded that the linear and quadratic models have poor approximations and should not be deployed. The third order model, with an acceptably small SSR, is able to represent the relationship between v_{wind} and Eig_{real} more accurately.

6.4.5 Comparison with Linear Voltage Collapse Indicator

The above obtained PCM model will be compared with a linear Voltage Collapse Indicator (VCI) first reported in [86]. The test system has been simplified to a two-bus system: the Thevenin equivalent is derived for the network to the left of Bus 3; the wind farm, treated as negative load, is combined with the load at Bus 4. Preliminary results are shown in Figure 36. Note that while the abscissa for PCM model is wind speed, we use the wind farm MW injection as input for VCI.

It can be concluded from Figure 36 that with the decrease of wind speed (hence the wind farm injection), there is an agreement of the two indicators to their voltage instability point (0 for the eigenvalue indicator and 1 for VCI). We can also observe that compared with the linear VCI, the third-order PCM model exhibits a non-linear characteristic.

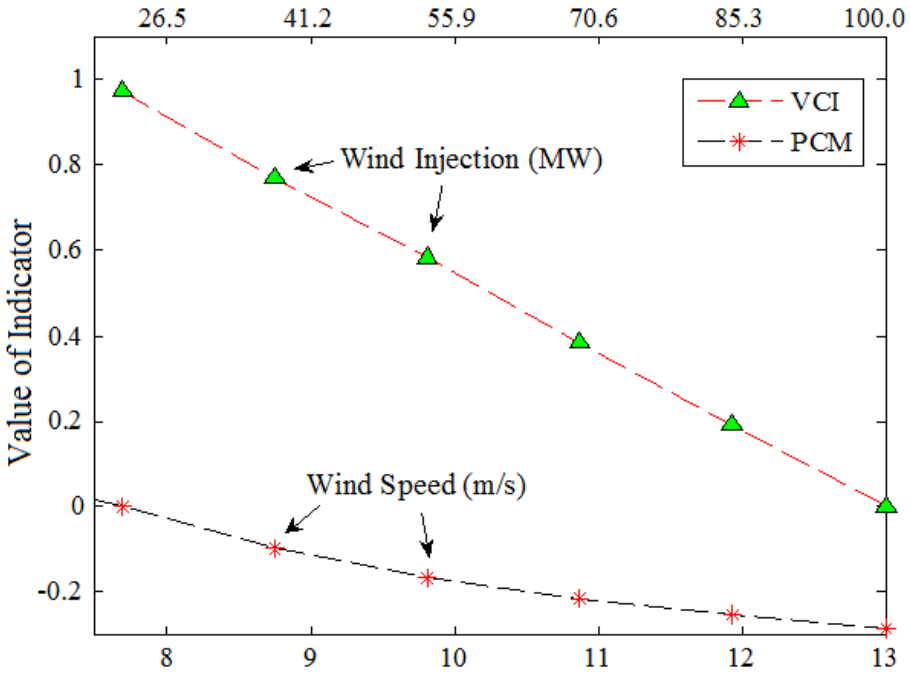


Figure 36 Comparison of the linear VCI and the PCM model

For situations *S1* and *S3*, it is easy to obtain their PCM models because however the wind speed varies under these two scenarios, the wind generator output will be kept constant. Thus the system structure will stay unchanged and so will the system critical eigenvalues. The corresponding results are shown in Table 13.

Table 13 PCM Approximation for Situations S1 and S3

Situation	Critical Eigenvalues	PCM Model
S1 (Cut-out)	$1.1376 \pm j5.0616$	$\hat{U} = 1.1376$
S3 (Rated)	$-0.2916 \pm j7.6980$	$\hat{U} = -0.2916$

Until now, the system response in terms of an uncertain input has been modeled by means of PCM approximation. The obtained third order model is in a discontinuity form induced by the three different wind speed conditions.

6.5 Case Study Using a Larger System

Now that its applicability has been verified in a simple system, the PCM-based uncertainty analysis will be conducted in this section for a larger system. A deeper understanding will be obtained by applying it to the SD voltage stability analysis.

6.5.1 System Description

A 6-machine 23-bus system to be examined has the single-line diagram shown in Figure 37. This network contains six generator buses (Bus 1 to Bus 6), and six load buses (Bus 7 to Bus 12). The system raw data is originally provided by the commercial software PSS/E.

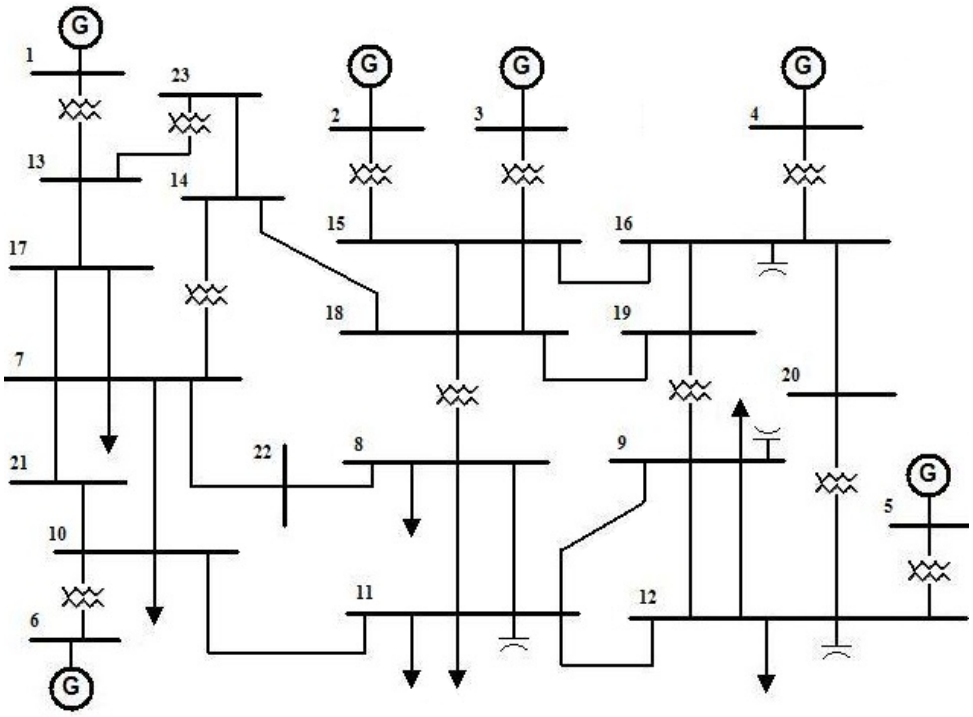


Figure 37 Single-line diagram of the 6-machine 23-bus system

The SD voltage stability of this system is explored when the generator buses are replaced by the collector of wind generators one by one. The wind generator model used is the GE 3.6 MW DFIG-based WTG. The technical specifications of this model can be found in [87]. In the simulation, to replace the original Bus *i* generators with the capacity of P_i, a wind farm equipped with $N\ (= P_i / 3.6)$ DFIGs is assumed.

To make the application more realistic and representative, a Weibull distribution shown in Figure 38 is applied to describe the wind speed probability distribution. (Note that if the proposed method is applied to an actual system, the local historical data should be used to specify the PDF of wind speed.) The PDF of a Weibull distribution is:

$$f(x;\lambda,k)=\begin{cases}\dfrac{k}{\lambda}\left(\dfrac{x}{\lambda}\right)^{k-1}e^{-(x/\lambda)^k} & x\geq 0\\ 0 & x<0\end{cases} \tag{6.21}$$

Weibull curves with different λ and k are shown in Figure 38. The distribution with k=2.2, λ=10.5 is adopted.

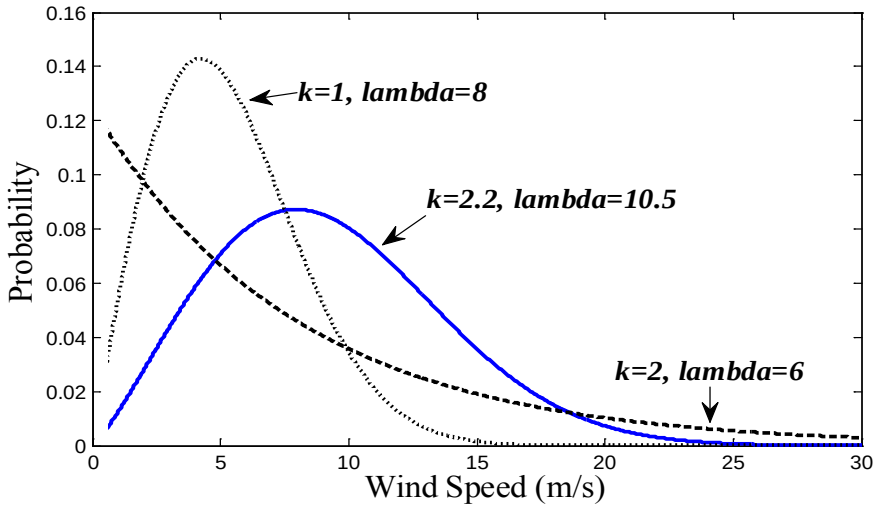


Figure 38 Wind speed represented by Weibull distribution

6.5.2 PSDVS Calculation

In [79] and [88], the system probabilistic small-signal stability is calculated as a stability index. Similarly, the system SD voltage stability could be evaluated in a probabilistic manner, and the probabilistic collocation method is an efficient tool to conduct this task.

In our research, the PCM is implemented to obtain the probability of small disturbance voltage stability (PSDVS). Computer programs are developed to calculate

the PSDVS, which is illustrated in the flow chart of Figure 39.

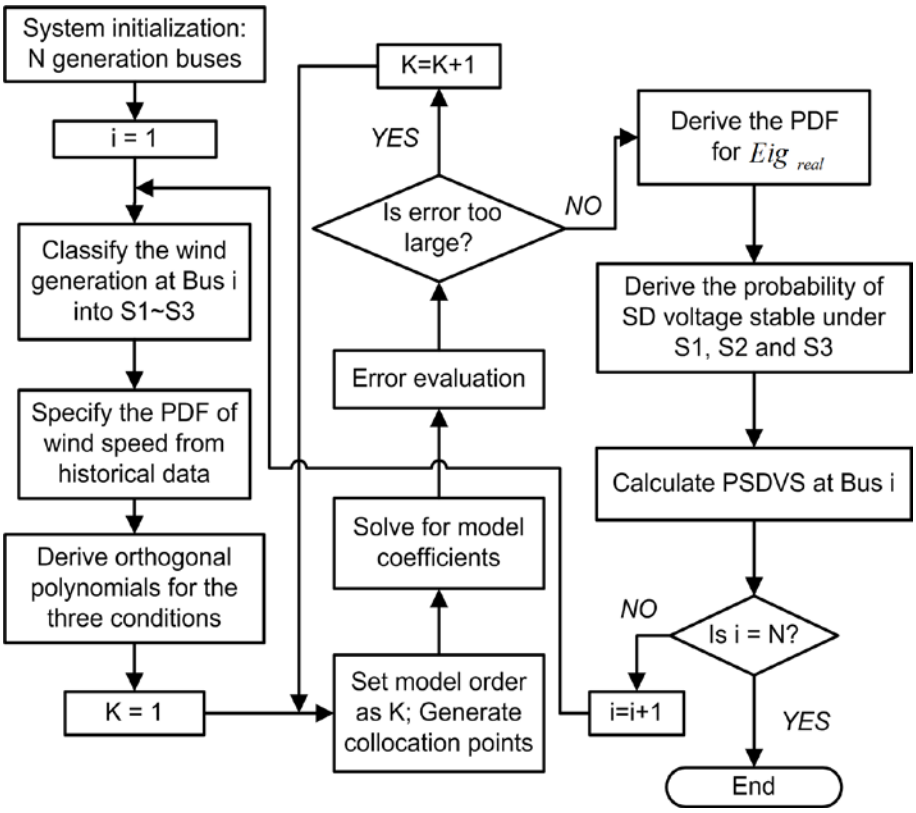


Figure 39 Flow chart of the PSDVS calculation process

We will replace the Bus 1 generator with DFIG-based wind generators to demonstrate the calculation of PSDVS. First, the orthogonal polynomials need to be derived. For v_{wind} with Weibull distribution, the Associated Laguerre Polynomials [89] are deployed to derive the orthogonal polynomials. The derivation of orthogonal polynomials for Weibull distribution is detailed in Appendix B.

The polynomial approximation for Eig_{real} is then obtained by means of the previously illustrated algorithm steps:

$$Eig_{real} = \begin{cases} S1: \ 0.4936 \\ S2: C_0 = -0.454, C_1 = -0.179, C_2 = 0.051, C_3 = -0.011 \\ S3: -0.7150 \end{cases}$$

According to the above PCM model, the trajectory of Eig_{real} with the variation of v_{wind} under the situation of *S2* is plotted in Figure 40. From the discussion in Section 6.3, the voltage instability is possible depending whether or not the above eigenvalue passes the origin, which is also described as the dynamic bifurcation point.

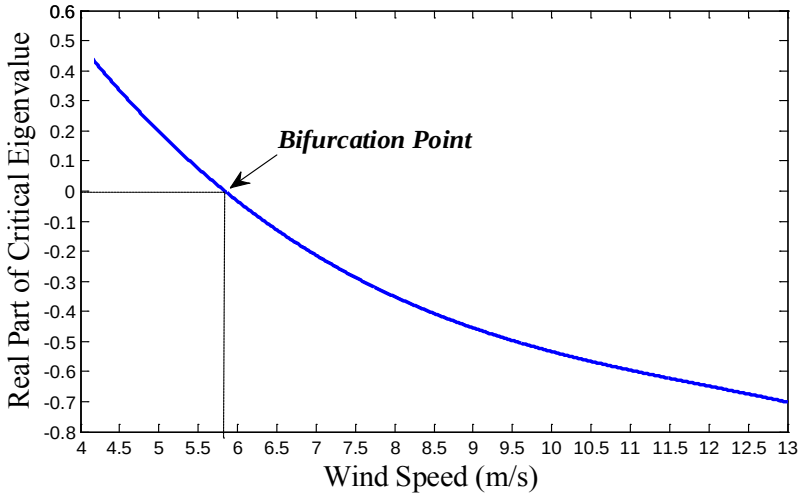


Figure 40 Trajectory of Eigreal with the variation of vwind under the situation of S2 using third-order PCM model

It could be observed that Eig_{real} has been represented by a monotonic function of v_{wind} within *S2*. Given the PDF of random variable v_{wind} as $f_X(v_{wind})$, and the PCM monotonic model $Eig_{real}=P_2(v_{wind})$, the PDF of Eig_{real} could be calculated using the following equation [90]:

$$f_Y(y) = \left| \frac{1}{P_2'\left(P_2^{-1}(y)\right)} \right| \times f_X\left(P_2^{-1}(y)\right) \tag{6.22}$$

where $f_Y(y)$ is the PDF of Eig_{real}, P_2^{-1} denotes the inverse function, and P_2' denotes the derivative. The meaning of this equation could be interpreted as follows: If the PDF of a random variable X is known as $f_X(x)$, it is possible to calculate the PDF of some variable $Y = g(X)$. This is also called a "change of variable" and is in practice used to generate a random variable of arbitrary shape $f_{g(X)} = f_Y$ using a known random number generator.

Implementing the above equation, the probability distribution of Eig_{real} is calculated and shown in Figure 41.

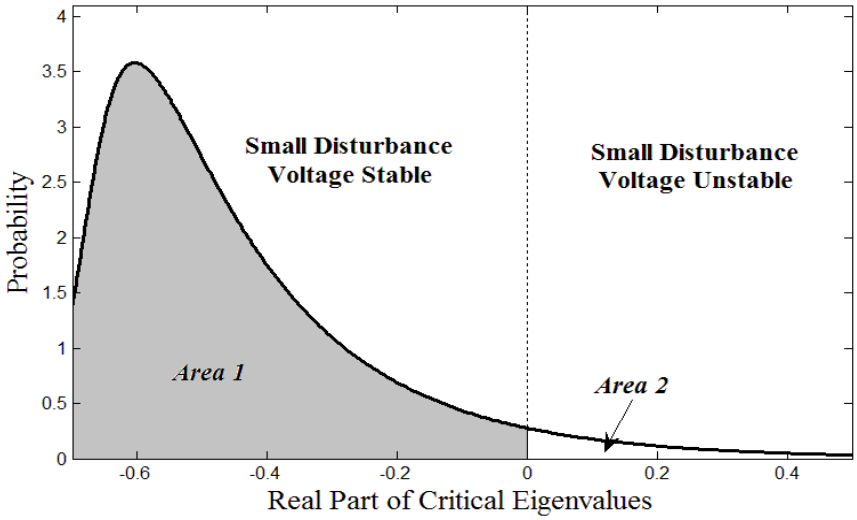


Figure 41 PDF of Eigreal under the situation of S2 using third-order PCM model

Area 1 in Figure 41 indicates the voltage is SD stable. The integral of Area 1 gives PSDVS under the situation of *S2*. From the analysis of previous section as well as

simulation results, all Eig_{real} under *S3* naturally falls on the stable side of the plane and all Eig_{real} under *S1* are located on the unstable side. Two approaches could be used then to obtain the total PSDVS. One is the analytical method. Define *f(U)* as the probability of Eig_{real} at the value of U:

$$PSDVS = \int_{S1}^{U<0} f(U)dU + \int_{S2}^{U<0} f(U)dU + \int_{S3}^{U<0} f(U)dU$$

$$= \left(\int_{S2}^{U<0} f(U)dU + \int_{S3} f(U)dU \right) \times 100\% \qquad (6.23)$$

The other approach approximates the PDF of Eig_{real} via a sufficient number of model simulations. Assuming that m out of n sample points are located on the stable side, then the PSDVS is represented as the ratio of m over n.

6.5.3 Numerical Results

The PSDVS could be viewed as a voltage security index to reflect how the system would be SD stable if the traditional generators are replaced by wind generators at one particular bus. Table 14 gives the calculated PSDVS of each of the six generator buses in the 23-bus system. Meanwhile, using participation factors, the dominant state variables associated with each mode are obtained and listed in Table 14. E'_d and E'_q are the direct and quadrature axis components of the voltage behind the transient reactance, δ is the rotor angular displacement, and ω is the rotor angular velocity. The subscript numbers stand for the most relevant generators in each mode.

The results shown in Table 14 provide useful information to be considered by the system designers while determining the optimum wind farm locations. Further, based on

the calculation results, VAR arrangement strategy could be made for wind farm to adjust the reactive power output of DFIGs as well as the other reactive power compensators (SVC, STATCOM, etc.), and thus improve the voltage stability of the overall system.

Table 14 Comparative Results using PCM and Monte Carlo Method

Bus No.	$E(\hat{Y})$	PSDVS			Dominant States
		PCM	MC 5000	MC 10000	
1	-0.4544	87.5%	85.8%	87.9%	$E'_{q4}, \omega_5, \theta_5$
2	0.0347	42.3%	43.4%	42.6%	$E'_{q3}, \omega_3, \theta_3$
3	-0.0354	57.5%	56.2%	57.7%	$E'_{q2}, \theta_2, E'_{q5}$
4	-0.3219	81.6%	76.9%	80.1%	E'_{q2}, E'_{q3}
5	1.3822	0	0	0	E'_{d2}, E'_{d3}
6	0.0101	49.0%	47.8%	48.1%	E'_{q4}

Also listed in Table 14 are the numerical results obtained from traditional Monte Carlo method. This simulation-based approach is applied using 5000 samples and 10000 samples respectively. Simulation results are shown in Figure 42 to compare with the results obtained by third order PCM model.

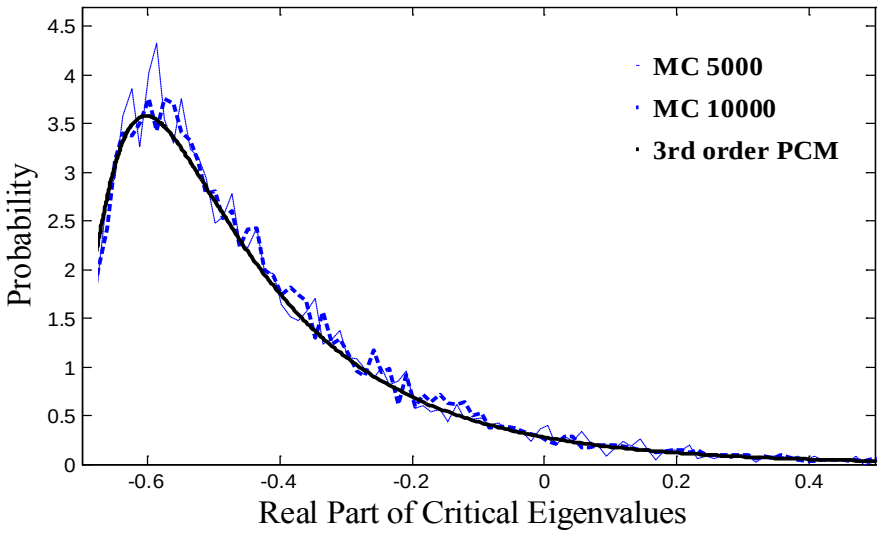


Figure 42 Calculated probabilistic distribution of Eigreal under the situation of S2 using PCM and Monte Carlo method

If the uncertainty analysis is conducted using the Monte Carlo approach from the beginning, thousands of simulations are needed to get the PDF of Eig_{real}. In comparison, with the ability to obtain the probabilistic distribution of Eig_{real} with similar accuracy, the third-order polynomial approximation method needs only four simulations for model derivation and a few more simulations for error evaluation. This could save a huge amount of time and computational resources. The PCM technique has exhibited great potential in the uncertainty analysis for wind generation.

6.6 Summary

In this section, the issue of small disturbance voltage stability considering the nodal injection uncertainty of grid-connected DFIGs has been investigated:

- The Probabilistic Collocation Method is introduced to perform the uncertainty analysis for wind generation. A classification of situations derived from the wind turbine cut-in, cut-out and rated wind speed is required. With the three situations specified, the PCM is applicable for small disturbance voltage stability analysis;
- Error evaluation is necessary for PCM to accurately represent the relationship between uncertain parameters and the outputs of interest. The PCM model with lower deviation from the real system enables a better approximation;
- The PCM is applied to a larger system. Simulation results indicate that the variation of wind farm output has considerable impact on the distribution of system critical eigenvalues. A deeper understanding could be obtained through the calculation of voltage security index PSDVS;
- As compared with the traditional simulation based approaches, the collocation method could provide quite an accurate approximation for the probabilistic distribution of system response with a much smaller number of model runs.

7. ONLINE ESTIMATION OF OSCILLATORY STABILITY USING A MEASUREMENT-BASED APPROACH

7.1 Introduction

Power system oscillatory stability assessment is the task of monitoring the rotor angle synchronism of generators at different locations. The recent trend in the electric power industry is to interconnect transmission lines linking small autonomous systems into large integrated systems, some of which span the entire continent. For example, in the United States and Canada generators which are located thousands of miles apart are operated simultaneously and synchronously. As a consequence inter-area electromechanical oscillations are becoming a more common occurrence. Since modern systems are optimally run near their stability threshold, the estimation of the distance of an operating point from instability region is critical for stable operation.

Traditional oscillatory stability assessment methods may not satisfy the online monitoring requirements because: 1) they are based on time-domain model simulations which are computationally intensive and time-consuming; 2) they use data collected from Supervisory Control and Data Acquisition (SCADA) systems, or state estimation functions, both of which are updated relatively infrequently.

With improved data acquisition technology, such as temporal synchronization of measurements at different locations, it may be possible to detect the onset of instability more accurately. The ability of synchrophasors to capture system-wide dynamics shows their potential in real-time system stability monitoring applications.

The advantages of a measurement-based approach include lower computational complexity, reduced knowledge requirements about system model parameters, and the potential to provide system stability assessment in real time. Most measurement-based approaches use appropriate signal processing or spectral analysis techniques to extract information from periodically collected power systems data. One such method is Prony analysis, which has been investigated by Kumaresan et al. in exponentially damped signal analysis [91]-[92], and later applied to power systems by Hauer et al. in oscillatory stability assessment [93]-[94]. Prony analysis is a powerful tool for mode parameter identification of electromechanical oscillations. However, if noise is present in measurements it performs poorly [92]. Another shortcoming of Prony's method is that it is only suitable for transient, or ringdown, data analysis, and cannot be applied to ambient data such that the system is excited by random load variations [95]. Therefore it is termed a ringdown analyzer that operates specifically on transient portion of a measured signal.

Alternatively, several mode meters, such as the Yule-Walker method [95], autoregressive moving average (AR/ARMA) model [96], and subspace estimation method [97]-[100], have been extensively studied in the past two decades in order to estimate mode parameters from both ambient data and transient data. While in previous efforts accurate estimation has been achieved for oscillation mode frequency, the problem of identifying mode damping, a more important task in terms of stability assessment, has not been satisfactorily resolved, although encouraging results were reported under certain test scenarios [93]-[101].

In this work, a data mining approach is used to estimate oscillatory stability in real time [98]-[99]. The decision tree (DT) method proposed by Breiman et al. [7] is deployed to map system operating point at each instant to one of several pre-defined stability states. Compared to previous research, the proposed approach casts the task as a multi-class classification problem, as detailed in Section 7.3. In Section 7.4 we show the results of the proposed method using the IEEE 39-bus test system. Finally, the data mining approach is evaluated on field PMU measurements from Salt River Project (SRP), a public electrical utility in Phoenix, Arizona, U.S.A.

7.2 Theoretical Background

7.2.1 Oscillatory Stability Assessment

As described in Section 2.1, oscillatory stability is related to Hopf Bifurcation. An instability event occurs when, following a small disturbance, the damping torques are insufficient to bring the system back to a steady-state operating condition, identical or close to the pre-disturbance condition.

Power system oscillations may be classified into four categories in terms of frequency: 1) speed governor band, from 0.01 to 0.15 Hz; 2) inter-area electromechanical band, from 0.15 to 1.2 Hz; 3) local electromechanical band, from 1.2 to 5 Hz; and 4) torsional dynamics band, from 5 to 15 Hz. This work focuses on the second category: the low-frequency inter-area oscillations.

7.2.2 Mode Identification without System Model

Traditionally, the stability of inter-area oscillations is evaluated through modal analysis of the system's non-linear differential algebraic equations (DAE) using detailed system model parameters, as detailed in Section 2.1. The inter-area oscillation modes that carry significant amount of energy but with insufficient DR are critical among all modes and need to be closely monitored.

In contrast to the model-based approach, the measurement-based approach does not require detailed system model information. Recent efforts take measurements from different locations during the same period of time, and identify oscillation mode parameters through signal processing techniques. The mode parameters that can be estimated include frequency, f, damping, σ, amplitude, A, and phase, θ, as shown in Figure 43.

There are three types of relevant power system measurements: ambient data, transient (ringdown) data, and probing data. Figure 44 shows the ambient and ringdown measurements. The probing data is beyond the scope of this work and will not be discussed further. For ambient data an AR/ARMA model is used to derive mode parameters while Prony analysis is used for ringdown data.

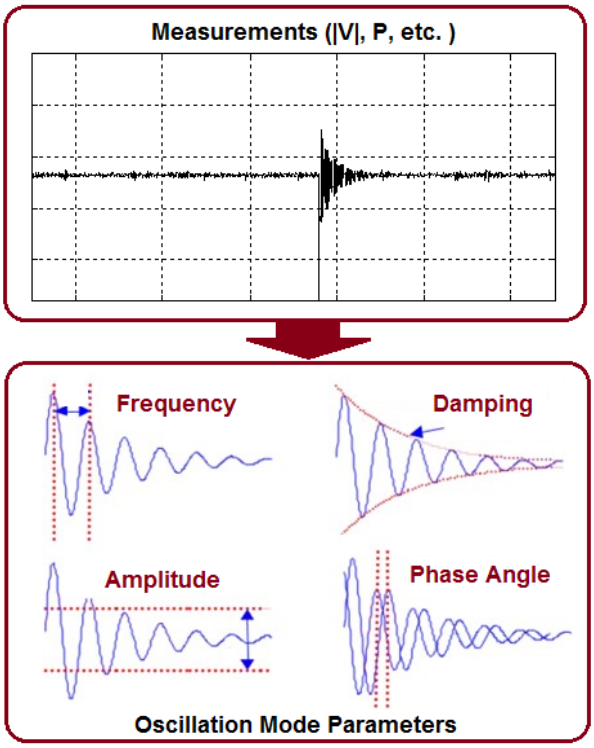


Figure 43 Mode parameters identified from power system measurements

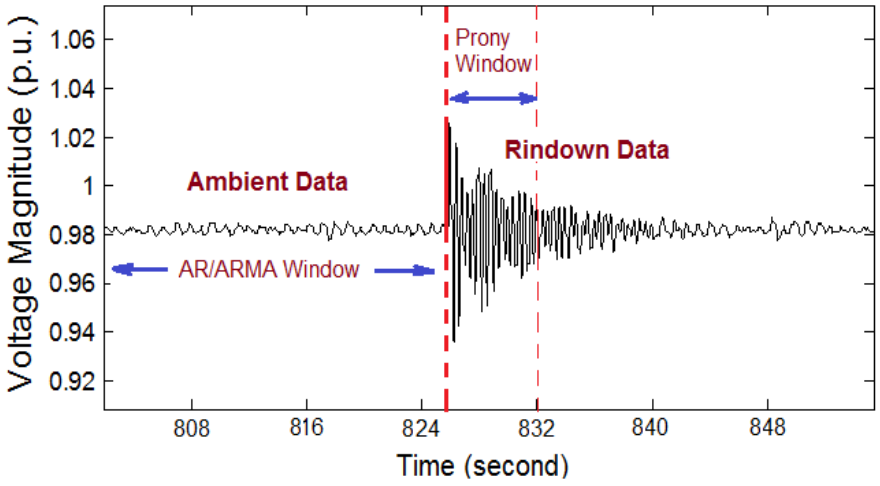


Figure 44 Ambient/ringdown signals and corresponding analysis windows

7.2.3 Data Mining Approach

The DT algorithm has been used as a classification tool for online oscillatory stability estimation. The DT is created by sequentially splitting the training data set at each tree node, starting from the root. The node splitting rule is determined by searching all candidate attributes, and finding the split which gives the largest decrease in class impurity. A terminal node is reached when maximum purity has been achieved.

In the experimental section we compared results obtained using DTs with those obtained using feed-forward neural networks and support vector machines. Both techniques are well known for their powerful modeling and generalization capabilities in classification analysis.

In this work the commercial software MATLAB [43] is used to implement the neural networks and support vector machines. Synchrophasors collected from PMUs are used as the input attributes to data mining tools.

7.3 Proposed Approach

7.3.1 Framework

A framework of the proposed measurement-based scheme is shown in Figure 45. The model-based approach, which was investigated by the authors in [38] and [45], is also shown in the figure for comparison purposes.

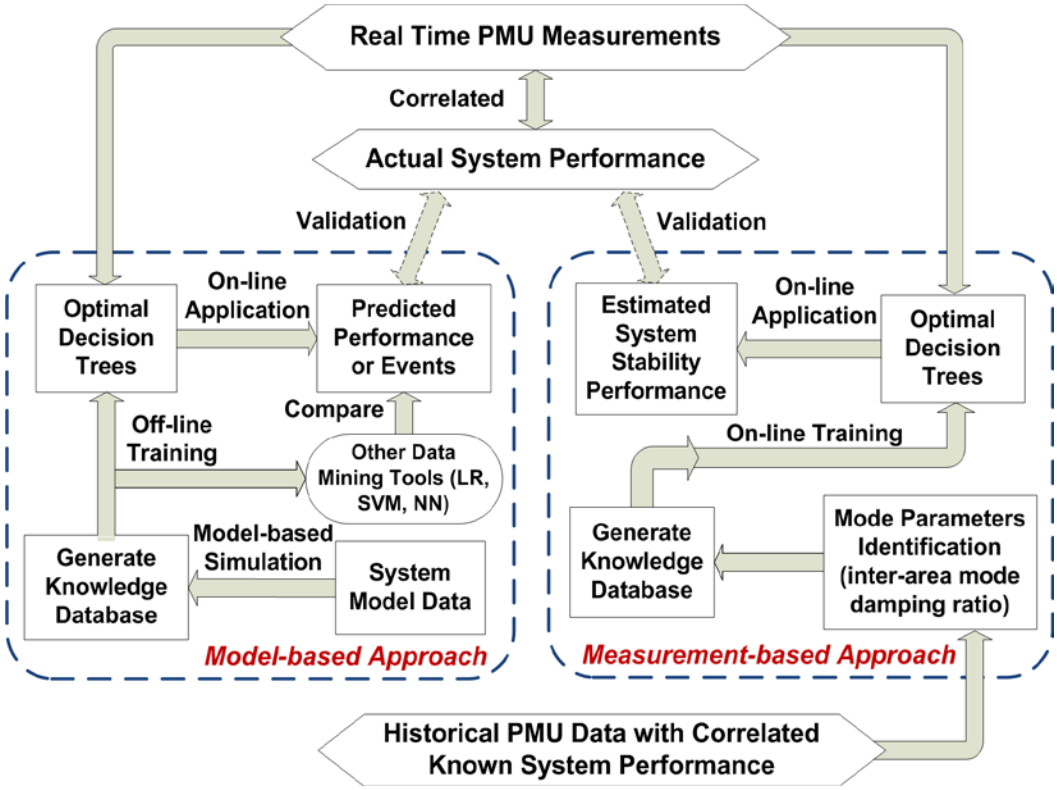


Figure 45 Model-based (left) and measurement-based (right) methods

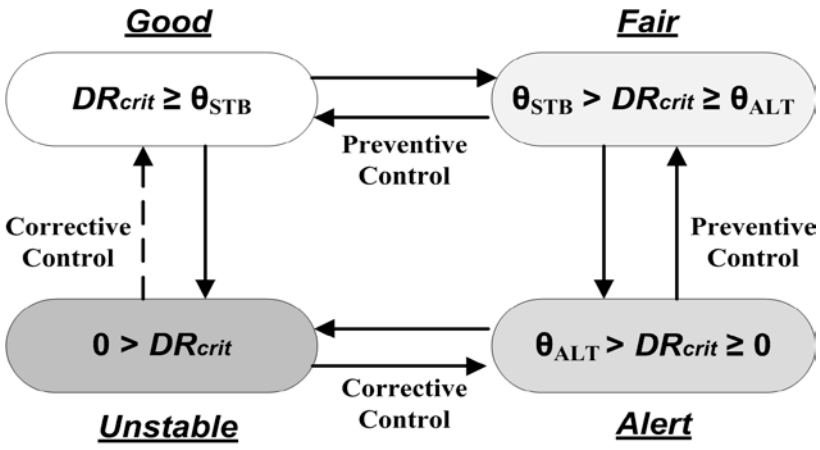


Figure 46 Classification of oscillatory stability states

For each power system, several stability thresholds are specified with respect to the typical damping ratio of the critical oscillation mode (DR_{crit}), and a set of stability states is defined accordingly. As shown in Figure 46, for the given oscillatory stability thresholds $\boldsymbol{\theta_{STB}}$ and $\boldsymbol{\theta_{ALT}}$ ($\boldsymbol{\theta_{STB}} > \boldsymbol{\theta_{ALT}}$), operating points (OPs) will be labeled as 'Good' if they satisfy $DR_{crit} \geq \boldsymbol{\theta_{STB}}$ $VS_{\text{margin}} > 40\%$; "Fair" if they satisfy $\boldsymbol{\theta_{STB}} > DR_{crit} \geq \boldsymbol{\theta_{ALT}}$; "Alert" if they satisfy $\boldsymbol{\theta_{ALT}} > DR_{crit} \geq 0$; and 'Unstable' when $0 > DR_{crit}$. In practice, the values of $\boldsymbol{\theta_{STB}}$ and $\boldsymbol{\theta_{ALT}}$ are usually around 10% and 5% respectively.

7.3.2 Mode Parameter Identification

Figure 47 illustrates the online application procedures of the proposed scheme. As the first step, a knowledge base needs to be created in order to train the classification tree. Included in the knowledge base are the input PMU measurements at each system operating point (OP), as well as the oscillatory stability state corresponding to each OP.

The procedure is initialized with a window scanning of the historical PMU measurements. An Oscillation Detector (OD) is designed to detect whether a transient event occurs by monitoring the presence of a sudden deviation in recorded measurements. If there are no abnormal changes, the OD suggests that the system is operated under a steady state, and an AR/ARMA model is employed to estimate the mode parameters in a sliding window manner. The required window length for ambient data analysis varies from 5 minutes to half an hour, depending on the variation level of system loads. If a sudden deviation is detected, but only limited to fewer than 5 data points, the corresponding measurements are considered outliers caused by sensor or

communication error, and are discarded from consideration. If a continued deviation has been observed, the OD will report that a transient process is potentially occurring, and Prony analysis is applied to scan the transient data using a sliding window with a length of 5 to 10 seconds, depending on the critical mode frequency of the inter-area electromechanical oscillation.

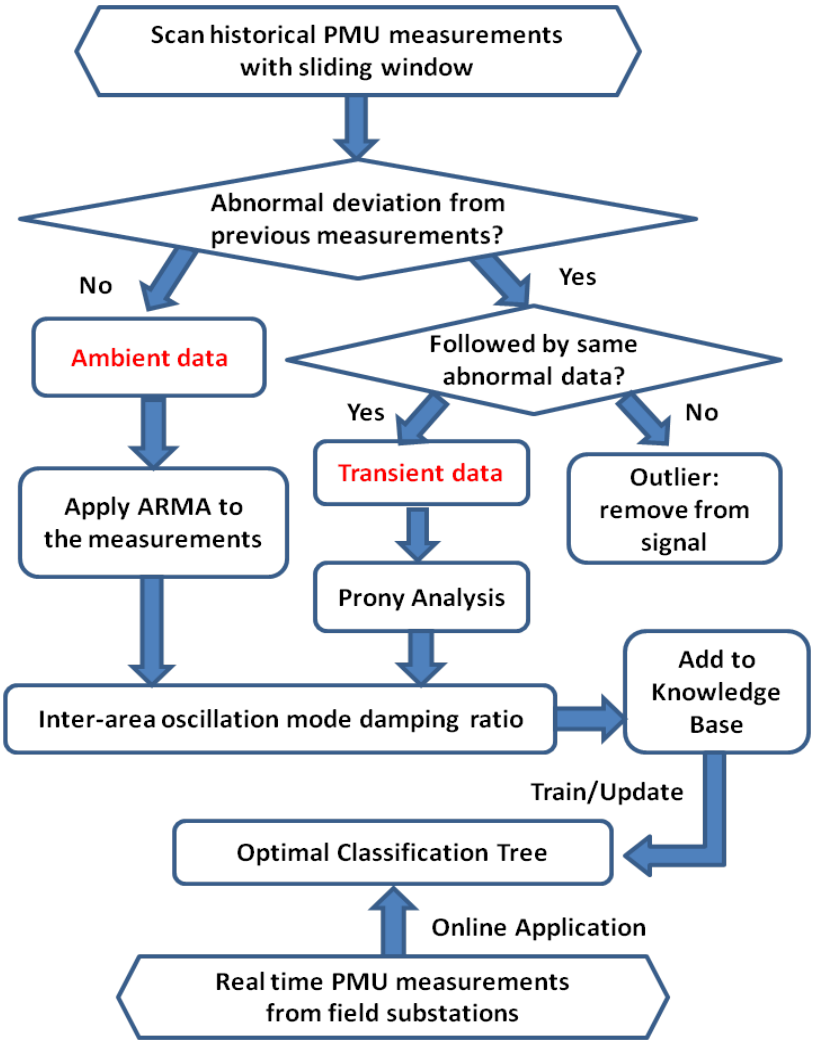


Figure 47 Online application of the proposed scheme

7.3.3 Classification Tree for Stability Assessment

In order to overcome the limitations of Prony and ARMA methods, the ringdown data is pre-processed using a low-pass filter, and the window length of AR/ARMA model is sufficiently large to assure accurate estimation. Once a sufficient number of cases have been accumulated, the knowledge base is used to train the classification trees. The derived optimal DT is then applied online. As shown in Figure 47, new PMU measurements are dropped down through the tree to predict the oscillatory stability status of each OP in real time.

One of the key challenges of embedding DTs in online applications is the problem of evolving system operating conditions. Due to variations in system generation and loading patterns, and changes in system topology, the DR_{crit} of inter-area electromechanical oscillations may also change. To deal with this eventuality, the classification tree derived in CART needs to be periodically refreshed in order to reflect the most current system operating conditions. This is done by updating the knowledge base using the most recent PMU measurements, and re-training the DT.

7.4 Case Study

The IEEE 10-machine 39-bus test system (New England system) is used to implement the proposed scheme. Its one-line diagram is shown in Figure 7. Firstly the oscillation mode parameters are estimated through model-based eigenvalue analysis. They will be used later to validate the results of the measurement-based approach.

The 39-bus system is modeled in MATLAB/SIMULINK. As shown in Figure 48, the Network Solution Module initializes the time-domain simulation, calculates power flow, and provides real time network solutions using dynamic model parameters.

The low-frequency oscillation modes with insufficient DRs are listed in Table 15. They are obtained from model-based eigenvalue analysis of the IEEE 39-bus system. Also listed in this table are the dominant generators that participate in the correlated oscillation modes.

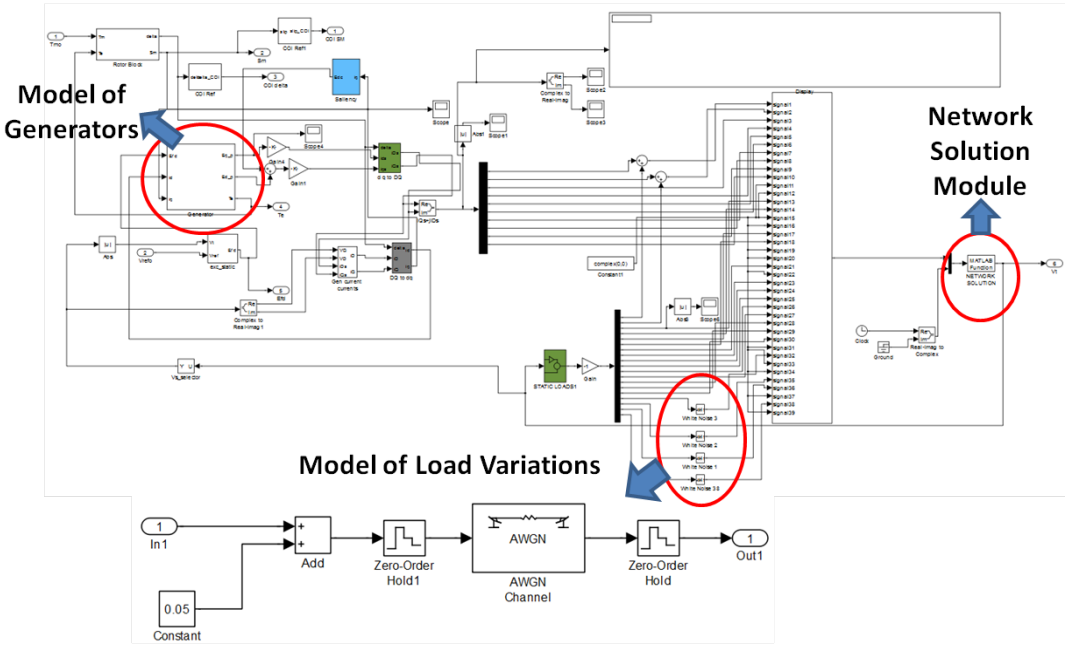


Figure 48 Simulink model of the IEEE 39-bus test system

In this work the Mode #5 with a frequency of 0.58 Hz is targeted for monitoring. To simulate the load variations, Gaussian noise with Mean = 0.05 and Signal to Noise Ratio (SNR) = 20 dB has been introduced to four system loads. The time-domain

simulation has been performed for 15 minutes. To create transient signal, a fault that caused the line between Bus 26 and Bus 28 to trip has been simulated. The fault occurred at t = 700s, and lasted for 0.02s. The resulting measurements from all system buses are recorded. In particular, the voltage magnitudes and phase angles at Bus 7 and Bus 39 are shown in Figure 49 and Figure 50.

Table 15 Low-Frequency Oscillation Modes Obtained from Model Initialization

	Mode #1	Mode #2	Mode #3	Mode #4	Mode #5
Frequency (Hz)	**1.21**	**1.13**	**1.03**	**0.96**	**0.58**
Damping Ratio (%)	**1.06**	**4.62**	**1.87**	**8.81**	**6.35**
Dominant Generator	**G1, G3**	**G4, G6**	**G3**	**G10**	**G2**

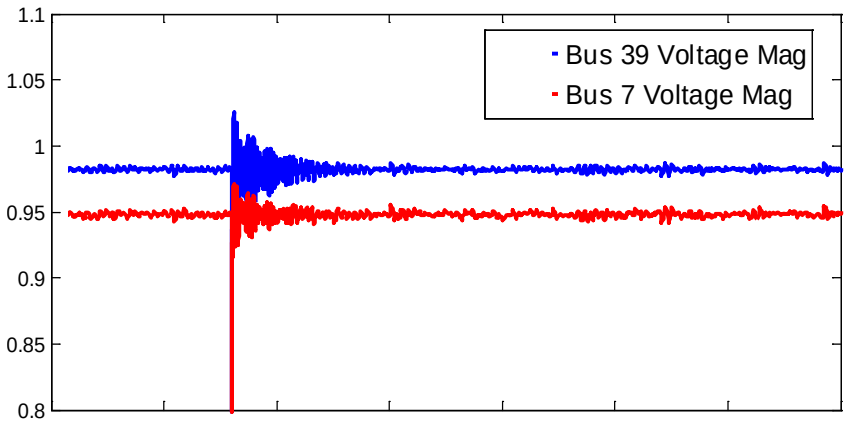


Figure 49 Voltage magnitude signals

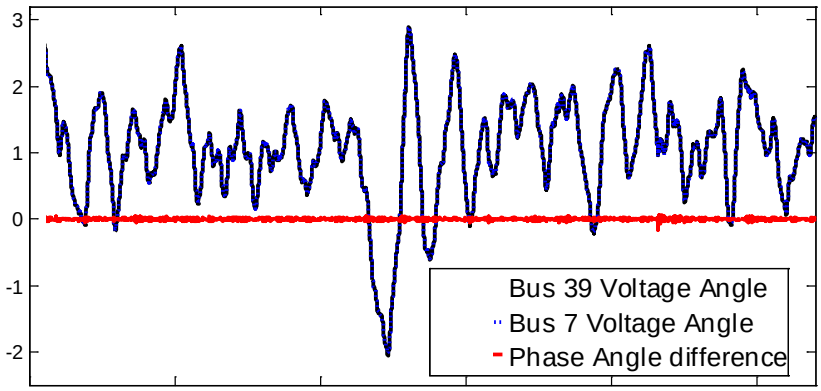


Figure 50 Phase angles and their difference

Prony analysis has been applied to the Bus 39 voltage magnitude signal during the transient process. The sliding window has a length of 5 seconds and the Prony model order is set to be N=30.

The AR model has been applied to the phase angle difference between Bus 7 and Bus 39, which is shown in Figure 50. The ambient data before the fault are treated using a sliding window with a length of 10 minutes. Different model orders have been deployed to compare the results. The mode damping ratios estimated by AR of order N=60 are drawn in Figure 51. The Mean of the damping ratios estimated with different model orders have been summarized in Table 16. Table 16 shows that the mode frequency estimated from AR and Prony are very close to the eigen-analysis results in Table 15. The damping ratio estimated by AR is approaching the actual value when increasing the model order. The DR estimated by Prony analysis is different due to the change in system topology.

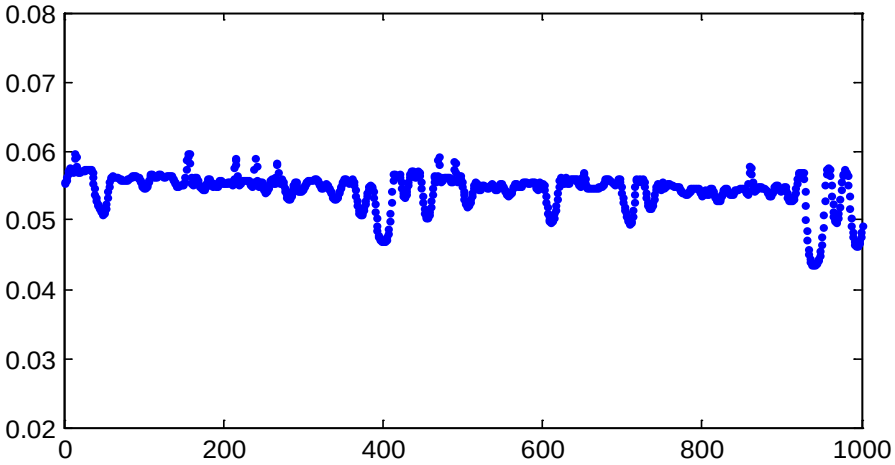


Figure 51 Damping ratios estimated from ambient measurements

Table 16 Estimate Mode #5 by Applying AR to Ambient Data

	Order	Frequency (Hz)	Damping Ratio (%)
AR	*N*=30	0.5622	4.391
	N=60	0.5819	5.637
	N=90	0.5753	6.224
Prony	*N*=30	0.5787	5.185

By varying the load disturbance level and fault scenario, the time-domain simulations have been replicated and a total of 4938 OPs with their corresponding stability states are included in the knowledge base. A classification tree has been developed in CART using 80% of the cases, and the rest 20% has been used in new case testing. The classification accuracy is evaluated as follows,

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Total\ Number\ of\ Prediction}. \tag{7.1}$$

The DT accuracy is summarized in Table 17. It is observed that an overall prediction accuracy as high as 98.38% has been achieved.

Table 17 Classification Tree Performance

	Good	Fair	Alert	Accuracy
Good	610	8	2	0.9839
Fair	3	349	1	0.9887
Alert	0	2	13	0.8667
Accuracy	0.9951	0.9721	0.8125	0.9838

7.5 Application to Field PMU Measurements

The filed PMU measurements received from a public electrical utility in Phoenix, Arizona, U.S.A., the Salt River Project (SRP), have been used to evaluate the proposed scheme. The data include synchronized voltage and current phasor measurements, under both ambient and transient conditions. The transient data recorded two consecutive brake insertion applications at a major transmission substation. The voltage magnitude measured at another substation has been divided into two 5-minute signals as shown in Figure 52. Each of the signals includes one transient process.

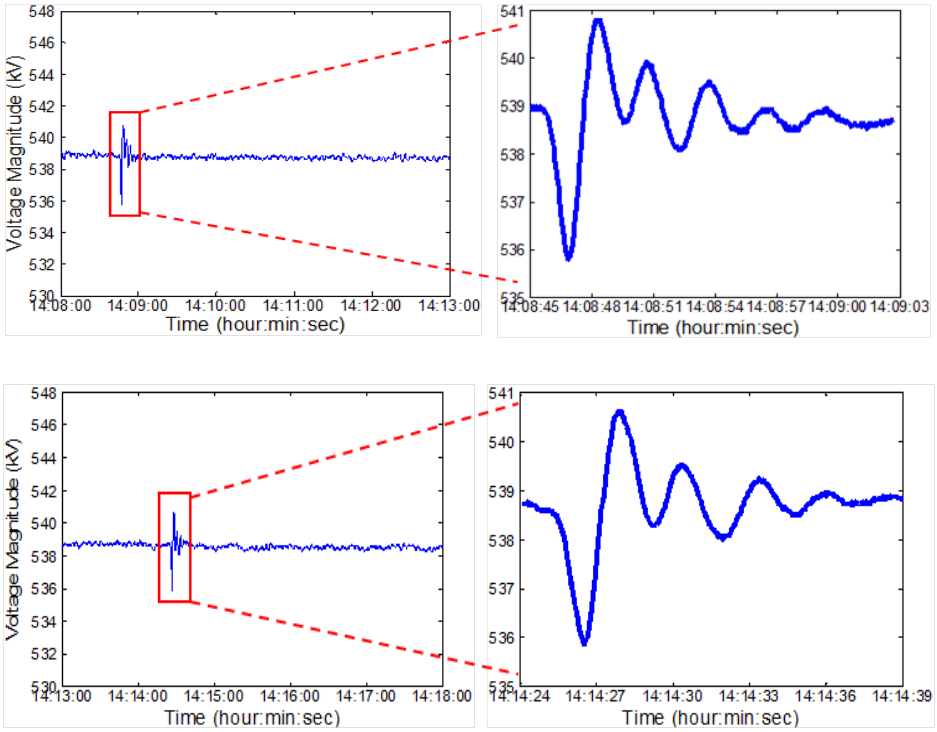


Figure 52 Field voltage magnitude measurements from PMUs

A knowledge base has been created by applying the same procedure introduced in Section 7.4 to the field measurements from PMUs. The resulting DT performance has been summarized in Table 18. Two other data mining tools, the artificial neural network (ANN) and support vector machine (SVM), have also been used to compare the results.

From Table 18, the DT-based prediction model achieved similar accuracy to other data mining tools. Compared to some "black-box" models, however, the DT provides a more transparent structure with a clearer cause-effect relationship. Its piece-wise structure and node splitting rules enable the identification of the critical variables and thresholds that should be analyzed to gain insight into the oscillatory stability of a system.

Table 18 Results Comparison

Data Mining Tools	Misclassification Rate			Overall Accuracy
	Good	Fair	Alert	
DT	0.0219	0.0667	0.0737	0.9739
ANN	0.0034	0.0902	0.1852	0.9873
SVM	0.0008	0.0738	0.0602	0.9940

7.6 Summary

The use of Decision Trees for online stability assessment without the knowledge of system model parameters has been investigated in this work:

- The proposed scheme is a measurement-based method that complements the traditional model-based approach. It is particularly useful when system model parameters are not readily available;
- The proposed approach is able to provide control center operators with real time support by making use of the quickly updated PMU measurements;
- Once trained using the knowledge base, the DT-based predictor can achieve high accuracy in online oscillatory stability estimation;
- The data mining tools are capable of reflecting the evolving system operating conditions when the most recent PMU measurements and corresponding knowledge base are used;

- When the results are compared with other data mining tools such as ANN and SVM, it is observed that almost identical prediction accuracy can be achieved.

8. CONCLUSIONS

8.1 Research Achievements and Contributions

Synchrophasors can help improving the quality of present-day SCADA/EMS solutions so that better initial operating points are available for real-time stability assessment applications. There are two complementary lines of defense against an instability event: preventive and corrective. This dissertation focuses on the preventive aspects, i.e. how synchrophasor measurements could help early detection of an impending instability event.

The main achievements and contributions of this research are summarized as follows:

- *Stability metrics based on synchrophasor measurements.*

  Two reliable metrics for monitoring system oscillatory stability and voltage stability have been deployed. Several stability states were defined according to the value of corresponding stability indicator. The two problems of stability state classification and stability margin prediction have been tackled.

- *Knowledge Base Generation.*

  A novel systematic approach of generating knowledge base for offline training of the data mining tools has been proposed and implemented. Stopping criteria were elaborated to assure a sufficient dataset for DT training. Synchronized voltage and current phasors have been used as DT input feature.

- *Data mining for stability assessment.*

A data mining methodology is proposed that takes use of the PMU collected synchrophasor measurements for online stability estimation and early detection of impending system instability events. In particular, the classification and regression trees were first utilized to evaluate system real time oscillatory and voltage stability status.

- *Performance examination of DTs.*

  The DT prediction accuracy for online assessment of system oscillatory stability and voltage stability has been tested. The accuracy and efficiency between the DT and other data mining tools such as Support Vector Machine (SVM) and Neural Networks (NN) has been compared.

- *Robustness of DTs.*

  The DT robustness with respect to PMU measurement errors and changes in system topology has been explored. The important issue of when and how to update the DTs in real time has been discussed and a feasible solution has been proposed.

- *Optimal PMU placement.*

  A novel methodology for optimal placement of PMUs in a power network is proposed. The variable importance derived from CART was utilized to rank the importance of network substations for stability assessment applications. The performance of DTs using synchrophasor measurements from a limited number of PMUs was checked.

- *Measurement-based voltage stability assessment.*

A measurement-based voltage stability index (VSI) to quantify the voltage stability margin for a general two-bus transmission system using time-synchronized data of the line two ends is proposed and applied to field PMU measurements. For distribution system, it has been modified to adapt to the situation where only measurements at feeder root are available. The proposed VSI was used to analyze the impact of wind farms connected to distribution system.

- *Data mining for uncertainty analysis.*

  The impact of grid integration of wind generation on system voltage stability has been explored using a new data mining tool, the Probabilistic Collocation Method (PCM). The issue of how data mining can help deal with the supply side uncertainty introduced by the stochastic nature of wind generation has been explored.

- *Measurement-based oscillatory stability assessment.*

  The measurement-based techniques such as ringdown analyzers and mode meters have limitations when noise is present in measurements or short windows of time-series data are analyzed, respectively. In this work we show how to overcome these disadvantages by using decision trees to directly map synchrophasor measurements to one of four predefined stability states. Results indicate that the proposed measurement-based approach complements the traditional model-based approach, enhancing situational awareness of control center operators in real time stability monitoring and control.

8.2 Conclusions

The following conclusions were reached in this research:

- The DT-based data mining model provides an accurate assessment of the stability status of each system operating point. Compared with some other data mining tools, using DT it is possible to identify the critical variables and thresholds that need to be analyzed to gain insight into the stability margin of a power system;
- Encouraging results were obtained through performance examination using the proposed knowledge base generation methodology. With a sufficiently captured system stability behavior, the DT model can predict the system oscillatory and voltage stability status with high accuracy;
- The CT classification accuracy is related to how the tree is developed, and the setting for minimum parent node cases can alter the shape of the resulted tree as well as its performance;
- According to the test results, the RT model is fast enough to process PMU measurements, and it is robust to handle measurement errors that are within 1% TVE;
- The RT sensitivity to system topology variation becomes less distinct in large sized network and under mild changes in topology. The proposed DT update methodology enables seamless online stability monitoring;
- The combined bus ranking derived from RT variable importance is used to suggest optimal PMU locations. Test results show that the measurements from reduced locations can still lead to satisfactory RT prediction accuracy;

- The proposed measurement-based voltage stability index can accurately quantify the voltage stability margin for transmission system. For distribution network only local measurements are needed. No time synchronization is needed;
- Simulation results demonstrate that the fixed-speed wind turbine requires large amount of reactive power support, whereas the variable speed wind turbine could enhance node voltage stability without additional reactive power compensation devices;
- The proposed data mining model based on PCM is applicable for small disturbance voltage stability analysis;
- Simulation results indicate that the variation of wind farm output has considerable impact on the distribution of system critical eigenvalues. A deeper understanding was obtained through the calculation of voltage security index PSDVS;
- As compared with the traditional simulation based approaches, the collocation method provides quite accurate approximation for the probabilistic distribution of system response with much fewer model runs;
- The proposed measurement-based oscillatory stability assessment method complements the traditional model-based approach. It is particularly useful when system model parameters are not readily available;
- The data mining tools are capable of reflecting the evolving system operating conditions when the most recent PMU measurements and corresponding knowledge base are used;
- When the results are compared with other data mining tools such as ANN and

SVM, it is observed that almost identical prediction accuracy can be achieved by using DT. In addition, the DT model structure provides a more transparent structure with a clearer cause-effect relationship.

8.3 Suggestions for Future Work

The following aspects need to be considered in future work:

- *Optimal PMU placement for other applications.*

  In this work the problem of finding the best power network locations to install PMUs has been investigated. The proposed methodology of using DT variable importance can be further implemented in other applications. For example, the sensor placement for optimal fault location accuracy may be facilitated by the proposed approach.

- *From statistical estimation to exact solution.*

  In this dissertation the impact of wind generation uncertainty on system stability behavior is studied in a probabilistic manner. In the future, the method can be extended to quantify the exact stability margin for each operating point with a certain amount of power produced by wind generators.

REFERENCES

[1] M. Kezunovic and A. Abur, "Merging the temporal and spatial aspects of data and information for improved power system monitoring applications," *IEEE Proceedings*, Vol. 9, Issue 11, pp. 1909-1919, 2005.

[2] C. Zheng, Y. Dong, O. Gonen, and M. Kezunovic "Data integration used in new applications and control center visualization tools," IEEE Power and Energy Society General Meeting, Minneapolis, USA, July 2010.

[3] M. Kezunovic, C. Zheng, and C. Pang, "Merging PMU, operational, and non-operational data for interpreting alarms, locating faults and preventing cascades," 43rd Hawaii International Conference on System Sciences (HICSS), Jan. 2010.

[4] P. Kundur, *Power System Stability and Control*. New York: McGraw-Hill, 1994.

[5] G. Rogers, *Power System Oscillations*. Boston: Kluwer Academic Publishers, 2000.

[6] T. V. Cutsem and C. Vournas, *Voltage Stability of Electric Power Systems*. Boston: Kluwer Academic Publishers, 1998.

[7] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Pacific Grove: Wadsworth, 1984.

[8] P. M. Mahadev and R. D. Christie, "Envisioning power system data: concepts and a prototype system state representation," *IEEE Trans. Power Syst.*, Vol. 8, No. 3, pp. 1084-1090, August 1993.

[9] T. J. Overby, "Effective power system control center visualization". PSerc Project S-25 Final Report (08-12), [Online]. Available: http://www.pserc.org.

[10] T. J. Overby, E. M. Rantanen, and S. Judd, "Electric power control center visualization using geographic data views," 2007 iREP Symposium - Bulk Power System Dynamics and Control, Charleston, SC, Aug. 2007.

[11] Y. Kobayashi, G. Karady, et al, "Satellite imagery for the identification of interference with overhead power lines," Project Final Report, PSERC, Jan. 2008. [Online]. Available: http://www.pserc.org.

[12] S. Kocaman, L. Zhang, A. Gruen, and D. Poli, "3D city modeling from high-resolution satellite images," ISPRS Ankara Workshop 2006, Ankara, Turkey, Feb. 2006.

[13] W. Prince, B. Wollenberg and D. Bertagnolli, "Survey on excessive alarms," *IEEE Trans. Power Syst.*, Vol. 4, No. 3, pp. 188-194, August 1989.

[14] C. Zheng and M. Kezunovic, "Integrated solutions for ubiquitous use of electricity and cyber services," IEEE PES 40th North American Power Symposium (NAPS), Calgary, Canada, September, 2008.

[15] X. Luo and M. Kezunovic, "Implementing fuzzy reasoning petri-nets for fault section estimation," *IEEE Trans. Power Delivery*, Vol. 23, No. 2, pp. 676-685, April 2008.

[16] M. Kezunovic, "Optimized fault location," PSerc Project T-32 Final Report (08-07), [Online]. Available: http://www.pserc.org.

[17] C. S. Chen, C. W. Liu, and J. A. Jiang, "A new adaptive PMU-based protection scheme for transposed/untransposed parallel transmission lines," *IEEE Trans. Power Delivery*, Vol. 17, Issue 2, pp. 395-404, April 2002.

[18] M. Kezunovic and B. Perunicic, "Automated transmission line fault analysis using synchronized sampling at two ends," *IEEE Trans. Power Syst.*, Vol. 11, No. 1, February 1996.

[19] H. Song and M. Kezunovic, "A new analysis method for early detection and prevention of cascading events," *Electric Power Systems Research*, Vol. 77, Issue 8, Pages 1132-1142, June 2007.

[20] C. Zheng and M. Kezunovic, "Synchronized sampling uses for real-time monitoring and control," IEEE PES 41st North America Power Symposium (NAPS), Starkville, Mississippi, USA, October, 2009.

[21] M. Kezunovic and C. Zheng, "Monitoring power system dynamic performance using synchronized sampling," CIGRE Conference on Monitoring of Power System Dynamics Performance, St. Petersburg, Russia, April 2008.

[22] M. Glavic and T. Van Cutsem, "Wide-area detection of voltage instability from synchronized phasor measurements. Part I: Principle," *IEEE Trans. Power Syst.*, Vol. 24, No. 3, pp. 1408-1416, June 2009.

[23] M. Glavic and T. Van Cutsem, "Detecting with PMUs the onset of voltage instability caused by large disturbances," IEEE PES General Meeting, 2008.

[24] B. Venkatesh, R. Ranjan, and H. B. Gooi, "Optimal reconfiguration of radial distribution systems to maximize loadability," *IEEE Trans. Power Syst.*, Vol. 19, No. 1, pp. 260-266, Feb. 2004.

[25] L. Wehenkel, T. V. Cutsem, and M. Ribbens-Pavella, "An artificial intelligence framework for on-line transient stability assessment of power systems," *IEEE Trans. Power Syst.*, Vol. 4 No. 2, pp. 789-800, May 1989.

[26] L. Wehenkel, M. Pavella, E. Euxibie, and B. Heilronn, "Decision tree based transient stability method a case study," *IEEE Trans. Power Syst.*, Vol. 9, No. 1, pp. 459-469, Feb. 1994.

[27] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *IEEE Trans. Power Syst.*, Vol. 9, No. 3, pp. 1417-1426, Aug. 1994.

[28] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Trans. Power Syst.*, Vol. 9, No. 1, pp. 459-469, Nov. 2007.

[29] T. V. Cutsem, L. Wehenkel, M. Pavella, B. Heilbronn, and M. Goubin, "Decision tree approaches to voltage security assessment," *Proc. Inst. Elect. Eng.*, Vol. 140, No. 3, pp. 189-198, May 1993.

[30] R. Diao, K. Sun, V. Vittal, et al., "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. Power Syst.*, Vol. 24, No. 2, pp. 832-839, May 2009.

[31] I. Kamwa, S. R. Samantaray, and G. Joos, "Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices," *IEEE Trans. Smart Grid*, Vol. 1, No. 2, pp. 144-158, Sep. 2010.

[32] R. Diao, V. Vittal, and N. Logic, “Design of a real-time security assessment tool for situational awareness enhancement in modern power systems,” *IEEE Trans. Power Syst.*, Vol. 25, No. 2, pp. 957-965, May 2010.

[33] S. P. Teeuwsen, I. Erlich, M. A. El-Sharkawi, and U. Bachmann, “Genetic algorithm and decision tree-based oscillatory stability assessment,” *IEEE Trans. Power Syst.*, Vol. 21, No. 2, pp. 746-753, May 2006.

[34] I. Kamwa, S. R. Samantaray, and G. Joos, “Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records,” *IEEE Trans. Power Syst.*, Vol. 24, No. 1, pp. 258-270, Feb. 2009.

[35] I. Kamwa, S. R. Samantaray, and G. Joos, “On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors,” *IEEE Trans. Smart Grid*, Vol. 3, No. 1, pp. 152-161, 2012.

[36] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, “Online monitoring of voltage stability margin using an artificial neural network,” *IEEE Trans. Power Syst.*, Vol. 25, No. 3, pp. 1566-1574, August 2010.

[37] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, “Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements,” *IEEE Trans. Power Syst.*, Vol. 26, No. 3, pp. 1474-1483, August 2011.

[38] C. Zheng, V. Malbasa, and M. Kezunovic, "A fast stability assessment scheme based on classification and regression tree," IEEE Conference on Power System Technology (POWERCON), Auckland, New Zealand, October 2012.

[39] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*. The Iowa State University Press, Ames, Iowa, 1977.

[40] Y. Zhou and V. Ajjarapu, "A fast algorithm for identification and tracing of voltage and oscillatory stability margin boundaries," *Proceedings of IEEE*, Vol. 93, No. 5, pp. 934-946, 2005.

[41] Carson W. Taylor, *Power System Voltage Stability*, New York: McGraw-Hill, 1994.

[42] V. Ajjarapu and C. Christy, "The continuation power flow: a tool for steady state voltage stability analysis," *IEEE Trans. Power Syst*., Vol. 7, No. 1, pp. 416–423, Feb. 1992.

[43] Mathworks Inc. MATLAB R2012b User's Guide. [Online]. Available: http://www.mathworks.com.

[44] Dan Steinberg and Mikhail Golovnya, *CART 6.0 User's Manual*. San Diego, CA: Salford Systems, 2006.

[45] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans. Power Syst*., Vol. 28, No. 2, pp. 1978-1987, May 2013.

[46] M. A. Pai, *Energy Function Analysis for Power System Stability*. Boston, MA: Kluwer, 1989.

[47] R. G. D. Steel and J. H. Torrie, *Principles and Procedures of Statistics*. New York: McGraw-Hill, 1960.

[48] R. F. Nau. (2005, February). Forecasting [Online]. Available: http://www.duke.edu/~rnau/rsquared.htm.

[49] Electric Power Research Institute, "DC multi-infeed study," EPRI TR-104586s, Projects 2675-04-05, Final Report 1994.

[50] *IEEE Standard for Synchrophasors for Power Systems*, IEEE Std. C37.118-2005, 2005.

[51] Y. Yang, X. Wu, and X. Zhu, "Mining in anticipation for concept change: proactive-reactive prediction in data streams," *Data Mining and Knowledge Discovery*, no. 13, pp. 261-289, 2006.

[52] Y. Dong, C. Zheng, and M. Kezunovic, "Enhancing accuracy while reducing computation complexity for voltage-sag-based distribution fault location," *IEEE Trans. Power Delivery*, Vol. 28, No. 2, pp. 1202-1212, April 2013.

[53] Thomas Ackermann, *Wind Power in Power Systems*. John Wiley & Sons Ltd, 2005

[54] Department of Energy, "20% energy by 2030: increasing wind energy's contribution to U.S. electricity supply," Washington, DC, July 2008.

[55] A. M. Chebbo, M.R. Irving, and M.J.H. Sterling, "Voltage collapse proximity indicator: behavior and implications," *IEE Proceedings - Generation, Transmission and Distribution*, Vol. 139, Issue 3, 1992.

[56] B. Venkatesh, R. Ranjan, and H. B. Gooi, "Optimal reconfiguration of radial distribution systems to maximize loadability," *IEEE Trans. Power Syst.*, Vol. 19, No. 1, pp. 260-266, Feb. 2004.

[57] B. Venkatesh, Alex Rost, and Liuchen Chang, "Dynamic voltage collapse index – wind generator application," *IEEE Trans. Power Delivery*, Vol. 22, No. 1, January 2007.

[58] E. Vittal, M. O'Malley, and A. Keane, "A steady-state voltage stability analysis of power systems with high penetrations of wind," *IEEE Trans. Power Syst.*, Vol. 25, No. 1, 2010.

[59] V. Akhmatov and P. B. Eriksen, "A large wind power system in almost island operation – a Danish case study," *IEEE Trans. Power Syst.*, Vol. 22, No. 3, 2007.

[60] C. Han, A. Q. Huang, M. E. Baran, et al, "STATCOM impact study on the integration of a large wind farm into a weak loop power system," *IEEE Trans. Energy Conversion*, Vol. 23, No. 1, March 2008.

[61] T. Petru and T. Thiringer, "Modeling of Wind Turbines for Power System Studies," *IEEE Trans. Power Syst.*, Vol. 17, No. 4, 2002.

[62] N. R. Ullah and T. Thiringer, "Variable speed wind turbines for power system stability enhancement," *IEEE Trans. Energy Conversion*, Vol. 22, No. 1, 2007.

[63] S. M. Muyeen, R. Takahashi, T. Murata, and J. Tamura, "A variable speed wind turbine control strategy to meet wind farm grid code requirements," *IEEE Trans. Power Syst.*, Vol. 25, No. 1, 2010.

[64] E. Muljadi, C. P. Butterfield, B. Parsons, and A. Ellis, "Effect of variable speed wind turbine generator on stability of a weak grid," *IEEE Trans. Energy Conversion*, Vol. 22, No. 1, March 2007.

[65] L. Xu, L. Yao, and C. Sasse, "Grid integration of large DFIG-based wind farms using VSC transmission," *IEEE Trans. Power Syst.*, Vol. 22, No. 3, 2007.

[66] P. Kessel and H. Glavitsch, "Estimating the voltage stability of a power system," *IEEE Trans. Power Delivery*, Vol. 1, No. 3, pp 346-354, 1986.

[67] M. Brucoli, F. Rossi, F. Torelli, and M. Trovato, "A generalized approach to the analysis of voltage stability in electric power systems", *Electric Power Systems Research,* Vol. 9, pp. 49-62, 1985.

[68] N. Yorino, H. Sasaki, Y. Masuda, et al, "An investigation of voltage instability problems", *IEEE Trans. Power Syst*., Vol. 7, pp. 600-611, 1992.

[69] C. D. Vournas, P. W. Sauer, and M. A. Pai, "Relationships between voltage and angle stability of power systems," *Electr. Power & Energy Syst.*, Vol. 18, pp. 493-500, 1996.

[70] N. Amjady and M. R. Ansari, "Small disturbance voltage stability evaluation of power systems", IEEE Transmission and Distribution Conference and Exposition, 2008.

[71] W. Freitas, L. C. P. DaSilva, and A. Morelato, "Small-disturbance voltage stability of distribution systems with induction generators," *IEEE Trans. Power Syst*., Vol. 20, No. 3, pp. 1653-1654, Aug. 2005.

[72] Z. Lubosny, *Wind Turbine Operation in Electric Power Systems*. New York: Springer, 2003.

[73] F. Zhou, G. Joos, and C. Abbey, "Voltage stability in weak connection wind farms," Proceedings of IEEE PES General Meeting, 2005.

[74] E. Muljadi, C. P. Butterfield, B. Parsons, and A. Ellis, "Effect of variable speed wind turbine generator on stability of a weak grid," *IEEE Trans. Energy Conversion*, Vol. 22, No. 1, pp. 29-36, March 2007.

[75] C. Zheng and M. Kezunovic, "Distribution system voltage stability analysis with wind farms integration," IEEE PES 42nd North America Power Symposium, Arlington, Texas, USA, 2010.

[76] V. Vittal, J. McCalley, V. Ajjarapu, and U. V. Shanbhag, "Impact of increased DFIG wind penetration on power systems and markets," PSerc Project Report, Publication 09-10.

[77] F. Mei and B. Pal, "Modal analysis of grid-connected doubly fed induction generators," *IEEE Trans. Energy Conversion*, Vol. 22, No. 3, pp. 728-736, September 2007.

[78] H. A. Pulgar-Painemal and P. W. Sauer, "Power system modal analysis considering doubly-fed induction generators," IREP Symposium on Bulk Power System Dynamics and Control, 2010.

[79] C. Wang, L. Shi, L. Wang, Y. Ni, and M. Bazargan, "Modeling analysis in power system small signal stability considering uncertainty of wind generation," Proceedings of IEEE PES General Meeting, 2010.

[80] M. A. Tatang, "Direct incorporation of uncertainty in chemical and environmental systems," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, 1995.

[81] M. Webster, M. A. Tatang, and G. J. Mcrae, "Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model," Joint Program on the Science and Policy of Global Change, MIT, Cambridge, MA, Tech. Rep. 4, 1996.

[82] J. R. Hockenberry and B. C. Lesieutre, “Evaluation of uncertainty in dynamic simulations of power system models: the probabilistic collocation method,” *IEEE Trans. Power Syst.*, Vol. 19, No. 3, pp. 1483-1491, Aug. 2004.

[83] D. Han, J. Ma, and R. He, “Effect of uncertainties in parameters of load model on dynamic stability based on probabilistic collocation method,” Proceedings of the IEEE 2007 Power Tech, 2007.

[84] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. New York: Academic, 1975.

[85] C. Zheng and M. Kezunovic, "Impact of wind generation uncertainty on power system small disturbance voltage stability: a PCM-based approach," *Electric Power Systems Research*, Vol. 84, No. 1, pp 10-19, March 2012.

[86] M. El-Kateb, S. Abdelkader, and M. S. Kandil, “Linear indicator for voltage collapse in power systems,” IEE Proceedings on Generation, Transmission and Distribution, Vol. 144, No. 2, pp. 139-146, March 1997.

[87] GE Energy, 3.6 MW Wind Turbine Technical Specifications.

[88] Z. Xu, Z. Y. Dong, and P. Zhang, “Probabilistic small signal analysis using Monte Carlo simulation,” Proceedings of IEEE PES General Meeting, 2005.

[89] M. Abramowitz and I. A. Stegun, *Orthogonal Polynomials*. Ch. 22 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, 1972.

[90] N. G. Ushakov, *Density of a Probability Distribution*. In Hazewinkel, Michiel, Encyclopedia of Mathematics, SpringerLink, 2001.

[91] R. Kumaresan and D.W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, pp. 833-840, Dec. 1982.

[92] R. Kumaresan, D.W. Tufts, and L.L. Scharf, "A Prony method for noisy data: choosing the signal components and selecting the order in exponential signal models," *Proc. IEEE*, pp. 230-233, February 1984.

[93] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Trans. Power Syst.*, Vol. 5, No. 1, pp. 80-89, Feb. 1990.

[94] J. F. Hauer, "Applications of Prony analysis to the determination of modal content and equivalent models for measured power system response," *IEEE Trans. Power Syst.*, Vol. 6, No. 3, pp. 1062–1068, Aug. 1991.

[95] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, "Initial results in electromechanical mode identification from ambient data," *IEEE Trans. Power Syst.*, Vol. 12, No. 3, pp. 1245–1251, Aug. 1997.

[96] R. W. Wies, J. W. Pierre, and D. J. Trudnowski, "Use of ARMA block processing for estimating stationary low-frequency electromechanical modes of power systems," *IEEE Trans. Power Syst.*, Vol. 18, No. 1, pp. 167–173, Feb. 2003.

[97] I. Kamwa, G. Trudel, and L. Gerin-Lajoie, "Low-order black-box models for control system design in large power systems," *IEEE Trans. Power Syst.*, Vol. 11, No. 1, pp. 303–311, Feb. 1996.

[98] C. Zheng, V. Malbasa, and M. Kezunovic, "Online estimation of oscillatory stability using synchrophasors and a measurement-based approach," 17th International Conference on Intelligent System Applications to Power Systems (ISAP 2013), to be published.

[99] V. Malbasa, C. Zheng, M. Kezunovic, "Power system online stability margin estimation using active learning and synchrophasor data," IEEE PES PowerTech 2013 Conference, to be published.

[100] N. Zhou, J. W. Pierre, and J. Hauer, "Initial results in power system identification from injected probing signals using a subspace method," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1296–1302, Aug. 2006.

[101] D. J. Trudnowski, J. M. Johnson, and J. F. Hauer, "Making Prony analysis more accurate using multiple signals," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 226–231, Feb. 1999.

APPENDIX A

1. RT Growing and Splitting

Suppose a knowledge base K consisting of N sample cases (x_1, y_1), (x_2, y_2), …, (x_N, y_N) is used to construct a RT.

Using the *Least Squares Regression*:

$$R(d) = \frac{1}{N}\sum_n \left(y_n - d(x_n)\right)^2$$

The value of $y(t)$ that minimizes $R(d)$ is the average of y_n for all cases (x_n, y_n) falling into node t, that is:

$$\overline{y}(t) = \frac{1}{N(t)}\sum_{x_n \in t} y_n$$

Given the set of candidate splits S, for any $s \in S$ that splits node t into t_L and t_R, let

$$\Delta R(s, t) = R(t) - R(t_L) - R(t_R)$$

The best split s^* of node t is that split in S which decreases $R(t)$ the most:

$$\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t)$$

A RT with T_{max} nodes is built by iteratively splitting nodes so as to maximize the decrease in $R(T)$. Splitting stops when for every $t \in T_{max}$, $N(t) \le N_{min}$. $N(t)$ is the number of samples falling into node t and N_{min} is a pre-defined threshold.

2. RT Pruning and Testing

For any subtree $T \le T_{max}$, define its complexity as $\widetilde{T}$: the number of terminal nodes in T. Then its cost-complexity measure $R_\alpha(T)$ is:

$$R_\alpha(T) = R(T) + \alpha\left|\widetilde{T}\right|$$

where $\alpha \geq 0$ and is called the complexity penalty.

For each value of α, find the subtree $T(\alpha) \leq T_{max}$ such that the cost-complexity $R_\alpha(T)$ is minimized:

$$R_\alpha(T(\alpha)) = \min_{T \leq T_{\max}} R_\alpha(T)$$

The result is a decreasing sequence of pruned trees, with an increasing sequence of α values:

$$T_1 > T_2 > T_3 > \ldots > \{t_1\}$$

$$0 = \alpha_1 < \alpha_2 < \alpha_3 < \ldots$$

where $T_1 \leq T_{max}$, t_1 is the tree contains the root node only.

To select the right sized tree from the sequence $\{T_1, T_2, \ldots\}$, a proportion of N is randomly selected and used as test samples TS. The cost of subtree T_k is:

$$R^{TS}(T_k) = \frac{1}{N_2} \sum_{(x_n, y_n) \in TS} \left(y_n - d_k(x_n)\right)^2$$

Another test method is the *V-fold cross-validation* (*CV*). Dividing N in V subsets $\{N_1, N_1, \ldots, N_V\}$, let:

$$R^{CV}(T_k) = \frac{1}{N} \sum_{V=1}^{V} \sum_{(x_n, y_n) \in N_V} \left(y_n - d_k(x_n)\right)^2$$

The relative error $RE^{CV}(T_k)$ of subtree T_k is given by:

$$RE^{CV}(T_k) = R^{CV}(T_k) / R(\bar{y})$$

3. Selection of the Best Pruned Tree

The *Standard Error* (SE) estimate is used to select the best pruned subtree commensurate with accuracy.

Take the cross-validation testing for example, the subtree with T_k nodes is selected as the best pruned tree if:

$$R^{CV}(T_k) \leq R^{CV}(T_{k0}) + SE$$

where

$$R^{CV}(T_{k0}) = \min_k R^{CV}(T_k)$$

APPENDIX B

1. Derivation of DFIG Algebraic Equations

Stator algebraic equations:

$$\begin{cases} V_{qs} = E_q^{'} - R_s I_{qs} - X_s^{'} I_{ds} \\ V_{ds} = E_d^{'} - R_s I_{ds} - X_s^{'} I_{qs} \\ P_s = V_{ds} I_{ds} + V_{qs} I_{qs} \\ Q_s = V_{qs} I_{ds} - V_{ds} I_{qs} \end{cases}$$

Rotor algebraic equations:

$$\begin{cases} V_{qr} = K_{P2}[K_{P1}(P_{ref} - P) + m_3 - I_{qr}] + m_4 \\ V_{dr} = K_{P4}[K_{P3}(Q_{ref} - Q) + m_1 - I_{dr}] + m_2 \\ P_r = V_{dr} I_{dr} + V_{qr} I_{qr} \\ Q_r = V_{qr} I_{dr} - V_{dr} I_{qr} \end{cases}$$

Other algebraic equations:

$$\begin{cases} I_{dr} = \dfrac{X_m}{X_r} I_{ds} + \dfrac{E_q^{'}}{X_m} \\ I_{qr} = \dfrac{X_m}{X_r} I_{qs} - \dfrac{E_q^{'}}{X_m} \\ P = P_r + P_s \\ Q = Q_r + Q_s \end{cases}$$

2. Derivation of Orthogonal Polynomials for Weibull Distribution

The Associated Laguerre Polynomials are orthogonal over [0, ∞) with respect to the measure with weighting function $x^a e^{-x}$:

$$\int_0^\infty x^a e^{-x} L_n^{(a)}(x) L_m^{(a)}(x) dx = 0 \qquad (n \neq m)$$

Compare with the orthogonal polynomials with "Weibull" weighting function:

$$\int_0^\infty \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} H_n(x) H_m(x) dx = 0 \quad (n \neq m)$$

It can be seen that variable transformation needs to be performed. Similar to that of Gaussian distribution, we assume:

$$Y = (\frac{x}{\lambda})^k$$

Substitute into the equation above, we have

$$\int_0^\infty Y^{\frac{k-1}{k}} e^{-Y} H_n(Y) H_m(Y) dY = 0 \quad (n \neq m)$$

Hence the associated Laguerre polynomials could be used considering the following:

$$\begin{cases} x = \lambda Y^{\frac{1}{k}} \\ a = \dfrac{k-1}{k} \end{cases}$$

The associated Laguerre polynomial of degree *n* is:

$$L_n^{(a)}(x) = \sum_{i=0}^{n} (-1)^i \binom{n+a}{n-i} \frac{x^i}{i!}$$

In our case:

$$a = \frac{k-1}{k} = 0.545$$

The first few orthogonal polynomials and collocation points are derived as follows:

$$H_0(Y) = 1$$
$$H_1(Y) = -Y + 1.545$$
$$H_2(Y) = \frac{1}{2}Y^2 - 2.545Y + 1.966$$
$$H_3(Y) = -\frac{1}{6}Y^3 + 1.773Y^2 - 4.511Y + 2.323$$

Note that Y is an intermediate variable. Once the collocation points are determined, they will be transformed back to corresponding wind speeds. The 1st to 5th order collocation points and corresponding wind speeds are provided below:

Order	Collocation Points	Corresponding Wind Speed (m/s)
1st	{1.5450}	{12.7958}
2nd	{0.9497, 4.1403}	{10.2566, 20.0290}
3rd	{0.6899, 2.8437, 7.1044}	{8.8697, 16.885, 25.6004}
4th	{0.5426, 2.1938, 5.1911, 10.2525}	{7.9524, 15.0065, 22.1977, 30.2453}
5th	{0.4474, 1.7913, 4.1501, 7.8057, 13.5305}	{7.2848, 13.6857, 20.0505, 26.7197, 34.3102}