A New State Transition Model for Forecasting-Aided State Estimation for the Grid of the Future

Mohammad Hassanzadeh

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Cansin Yaman Evrenosoglu, Co-Chair
Jaime De La Ree, Co-Chair
Lamine Mili
Virgilio Centeno
William Bauman
Eric de Sturler

May 2014
Blacksburg, VA

Keywords: State transition model, forecasting-aided state estimation, time-series analysis, vector autoregression

A New State Transition Model for Forecasting-Aided State Estimation for the Grid of the Future

Mohammad Hassanzadeh

ABSTRACT

The grid of the future will be more decentralized due to the significant increase in distributed generation, and microgrids. In addition, due to the proliferation of large-scale intermittent wind power, the randomness in power system state will increase to unprecedented levels. This dissertation proposes a new state transition model for power system forecasting-aided state estimation, which aims at capturing the increasing stochastic nature in the states of the grid of the future. The proposed state forecasting model is based on time-series modeling of filtered system states and it takes spatial correlation among the states into account. Once the states with high spatial correlation are identified, the time-series models are developed to capture the dependency of voltages and angles in time and among each other. The temporal correlation in power system states (i.e. voltage angles and magnitudes) is modeled by using autoregression, while the spatial correlation among the system states (i.e. voltage angles) is modeled using vector autoregression. Simulation results show significant improvement in power system state forecasting accuracy especially in presence of distributed generation and microgrids.

To my parents Ghasem Hassanazdeh & Jamalie Hashemi and my wife Parisa Nikkhoo

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Cansın Yaman Evrenosoğlu, for his patience, invaluable friendship, motivating energy, great supervising, and financially supporting me throughout my Ph.D. studies. He has been influential in my professional development.

I thank Dr. Lamine Mili for his help, patience, encouragement, and invaluable discussions about my research. I learned a lot from him and in his courses.

I thank my other committee members, Dr. Virgilio Centeno, Dr. Jaime De La Ree, Dr. William Baumann, and Dr. Eric de Sturler for their support and invaluable academic advices.

I would like to thank the Bradley Department of Electrical and Computer Engineering at Virginia Tech, and the Electrical & Biomedical Engineering Department at University of Nevada, Reno for the financial support during my Ph.D. studies. I also thank Dr. Mehdi Etezadi who provided me with the opportunity to continue my studies in the U.S.

My wife Parisa Nikkhoo and my parents always provided great support throughout my life. They never stopped encouraging me and they dedicated their life to supporting me. Thanks all of you for being there. I also thank all of my close friends in Reno and Blacksburg for being around.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

A. Motivation

Electric power system operators in the control centers are responsible for the secure and reliable operation of the grid. This requires continuous monitoring of system conditions using telemetered data, determination of the operating states and appropriate decision making for the required preventive and corrective actions using Energy Management System (EMS) functions such as power flow, state estimation, security constrained unit commitment, security constrained optimal power flow etc. In order to monitor the system conditions and determine the basis for necessary control actions, use of static state estimators for electric power systems is first proposed by Schweppe *et al*. in early 1970s [1-3]. Static state estimation filters a single scan of redundant telemetry providing a time snapshot of the network. Since the telemetered measurements are not time-stamped, it doesn't necessarily mean that measurements are obtained precisely at the same time. Thus, a static state estimator provides the best estimate of the system state (i.e. complex voltage at each node) by assuming that the difference between the arrival times of each measurement is negligible. This assumption is due to the fact that in a traditional network, demand and generation (consequently the voltage at each node) do not change drastically within the considered timeframes.

The importance of the static state estimators in Energy Management Systems (EMS) first came to prominence in late 1990s during the deregulation of electric energy markets. The report issued after the Northeastern blackout in August 2003 [4] in the U.S., established the fact that much could have been avoided with a functional topology processor and a state estimator.

As noted, the system state does not change drastically in the timeframe of a static state estimator in a traditional grid, and therefore, the operators have not been concerned with the fact that the static state estimation output actually represents a 'past state'. However, the importance of forecasting ability is recognized recently due to the proliferation of smart grids, which contains distributed small- and large-scale renewable-based power as well as microgrids. Wind and solar-based electric power is not only 'free' but also doesn't contribute to CO_2 emissions. Almost each state in the U.S. adopted legislation for increased renewable penetration, changing from 20% to

33% [5]. Due to the availability of high wind speeds and solar irradiation; wind and sun-based electric power dominates the ongoing and proposed 'renewable infrastructure'. Nevertheless, there are two main disadvantages of wind- and solar-based power generation from power system operator perspective:

1. Since they are not continuously available (intermittent), and cannot be accurately forecasted, they cannot be dispatched ahead of time similar to conventional thermal generation.
2. The highest speed of wind and solar irradiation are available in the areas far from the demand centers. Integration of these will require building new transmission lines (overhead and underground).

In addition, the proliferation of the microgrids will impact the classical load patterns and load forecasting traditions in grid operation. As microgrids are small networks with conventional and/or renewable generation, they are self-sufficient as long as they are equipped with effective and feasible storage systems. These systems are also designed such that when necessary they can be operated as 'grid connected'. It is obvious that the behavior of the microgrids is a 'new type of a demand' which is highly correlated with the presence of renewable generation.

The main motivation of this dissertation stems from the challenges of increased renewable penetration, which will significantly increase the possibility of sudden changes in the grid states within small timeframes to unprecedented levels. The contributions of this dissertation aim at mitigating the impacts of the proposed renewable penetration levels, and microgrids on the grid operation.

In 1970, Debs and Larson proposed dynamic state estimators (DSE) using Extended Kalman Filter (EKF) [6] to address the lack of forecasting in static state estimators. It is noted here that 'dynamic' refers to changes in system states due to variations in load and generation. The timeframe of interest is in the order of several seconds to few minutes. In order to avoid confusion with dynamic state estimators using the models of machine dynamics [7-15], DSE is also referred as forecasting-aided state estimation (FASE) [16]. This dissertation does not focus on dynamic estimators of system stability but concentrates on FASE to improve the real-time operations by providing a 'predictive' database of power system states. The predictive database increases the measurement redundancy by providing pseudo-measurements and helps re-establishing

observability due to loss of measurements, bad data detection and elimination, and topological error identification [16-19]. Moreover, a predictive database improves the real-time power system operation by providing short-term trajectory of the system state for short-term scheduling and congestion management [16, 20].

The analysis of the innovation vector along with normalized residuals in FASEs (and in early DSEs) is used to detect sudden load and topology changes [21]. The innovation vector can also be used to identify presence of bad data in critical measurements which is not possible by analyzing residual vector. FASE can provide optimal estimates of interconnected network states for systems in which some or all measurements are delayed due to a problem in communication channels [22].

FASE is implemented successfully in two utilities [21, 23]. However, the most important challenge in FASE still remains as the lack of a deterministic transition model to forecast the state [16, 20]. The represented models suffer from three drawbacks:

1. Time correlation is not addressed and exploited effectively. The impact of high percentage integration of intermittent generation on the system states can be exploited to capture the short-term behavior of states. The proper exploitation of the increased randomness in system states by using statistical models can help improving state transition model under certain assumptions.
2. The power system states are assumed to vary independently from each other, which leads to a diagonal state transition matrix. This assumption is not realistic in power systems in the presence of microgrids, distributed and intermittent power generation, which sometimes are connected to the grid through radial buses. The grid of the future is becoming more decentralized resulting in the decrease of number of buses with large amount of power injections (so called ‘electrical hubs’.) Therefore, these electrical hubs lose their significance and dominance in the grid, and the spatial correlation among the states throughout the grid become more observable. Spatial correlation defines the level of interaction of a state with the states of the neighboring buses. Taking spatial correlation into account enables the operator to restore observability when a state is unobservable due to loss of telemetry.
3. New measurement technologies such as phasor measurement units (PMUs) provide time-stamped, synchronized complex voltage and current information from various locations

throughout the grid. This information enables the system operators to be able to monitor the system with measurements at unprecedented frequencies (unlike the regular measurements, which arrive at every few seconds, the measurements provided by PMUs arrive at control centers every $1/30^{th}$ of a second.) In addition, PMUs provide the angles of bus voltages with respect to a reference unlike the conventional measurements, which are only capable of providing the voltage magnitudes. The existing state transition models are not effectively utilizing this valuable information provided by the PMUs.

This dissertation focuses on the above challenges and proposes a new state transition model by using time-series analysis techniques. The contributions of this dissertation are summarized in the following section.

B. Contributions

In this dissertation, a non-diagonal state transition model is proposed based on the time-series modeling of voltages. The method takes the temporal and spatial correlations between bus voltages into account

- to accurately capture the impacts of each bus voltage on the voltages of its neighboring buses,
- to exploit the increased randomness in bus voltages due to the intermittent renewable-based generation, and microgrids by using stochastic modeling, and
- it exploits the information available from PMUs to increase the accuracy of the state transition model.

Relative degrees of temporal and spatial correlation are determined for each state by observing the results of three metrics (i.e. electrical connectivity, electrical centrality, and electrical node significance analyses). Vector auto-regression (VAR) based time-series is then used to model simultaneous presence of temporal and spatial correlation. The time-series models of those states that are not spatially correlated to others are determined using auto-regression (AR). The performance of the resulting non-diagonal state transition matrix is tested on IEEE 57-, 118- and 300-bus networks, and the comparative results demonstrate the efficacy of the proposed method.

C. Chapter Organization

Chapter II reviews the operating states and control actions in power systems. Security analysis and the importance of static state estimation are discussed. The modeling assumptions and network modes to perform state estimation using weighted least square technique is reviewed. Obsevablity analysis and common bad data processing techniques are also provided.

Chapter III gives an overview of power system tracking state estimation. Tracking state estimators are proposed in 1970s [24] for faster state estimation solutions. These estimators assume that the change in the system state between two successive instants of time is extremely small. Therefore, at each time the state vector is easily updated by an amount proportional to the measurement residual. However, this estimator is unable to provide any forecasted state. The literature review is provided and the estimator is compared with a conventional state estimator.

Chapter IV gives an overview of power system forecasting-aided state estimation (FASE). Lack of forecasting in static and tracking state estimators are addressed in FASE using simple state transition models. A comprehensive literature review is provided and the main challenges are emphasized. In addition, the reasons to take temporal and spatial correlation between bus voltages into account in state transition models are described.

Chapter V reviews three metrics to identify and assess the time and space correlation between system states (i.e. bus voltages) in power networks. Electrical centrality, electrical network connectivity and electrical node significance are used in this dissertation to determine the relative degree of the temporal and spatial correlation between bus voltages. The information obtained from these metrics is used to determine the structure of the proposed state transition model for FASE using time series analysis techniques.

Chapter VI provides a brief review of time-series analysis. Univariate and multivariate autoregressive models are reviewed and the techniques to identify and fit a proper model for a set of time-series data are briefly summarized.

Chapter VII presents a new time-series based state transition model for power system FASE. The model takes the temporal and spatial correlation in and among the power system states into account. The performance of the model is tested on IEEE testbeds in presence of large-scale distributed wind penetration and microgrids.

Finally, Chapter VIII summarizes the contributions of this dissertation.

CHAPTER II

POWER SYSTEM STATIC STATE ESTIMATION

A. Introduction

The operating state of a power system can be determined using the network model and complex phasor voltages of all of the buses. The operating state of a network with respect to various operating conditions may move into one of the six possible states; secure, correctively secure, alert, controllable emergency, non-controllable emergency (in extremis), and restorative [25].

A power system operates in a normal state if the equality and the inequality constraints of the network are within the limits. The equality constraints refer to the loads supplemented by the available generations. The inequality constraints state that transmission lines flows, bus voltages and generation output power must not exceed the limits [26]. Normal state technically can be classified as secure and correctively secure states. In secure state, all loads are supplied without operating limit violations. In the event of a contingency, no violations will occur and the network survives without any need for post contingency actions. Common contingencies include the outage of transmission lines or generators due to faults. If the system will need appropriate control actions to prevent loss of load due to the contingency violations, then the state is classified as 'correctively secure' [27]. In the alert state, some violations caused by a contingency cannot be corrected with any loss of load. Therefore, some preventive actions must be taken to avoid line overloading and unwanted tripping of the protection devices.

If no preventive control actions are taken, operating condition may change significantly and operating constraints would be violated. This state is classified as 'emergency state' and immediate corrective actions are needed by the operator to bring the system back to the alert and normal state. Lack of appropriate actions may lead the system evolve into a 'non-controllable emergency' (in-extremis) state. In this condition, cascading of the component outages result in partial or system-wide blackout.

In order to bring the system back to the normal state; various loads, lines, transformers or other equipment should be disconnected to eliminate the operating constraint violations. In this stage, the system would be recovered with reduced load and reconfigured topology. Such an operating state is classified as 'restorative state' and the required action to start supplying power to all the

loads are referred to as ‘restorative control’. The operating states and control actions in a power system are summarized in Figure 2.1.

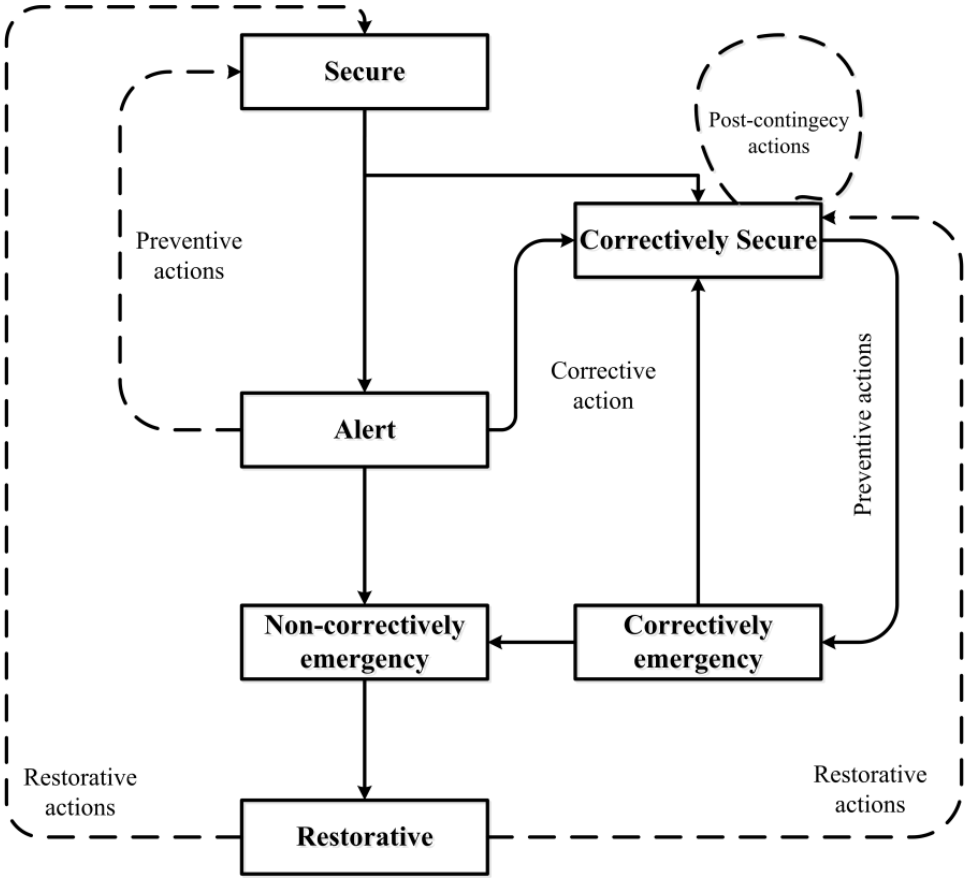

Figure 2.1. Operating states and control actions in a power system.

The main goal of system operators is to keep the system in normal state. Continuous monitoring of the system conditions is necessary in order to be able to keep the grid in normal state. Consequently the operators determine the operating states and they make appropriate decisions to take preventive or corrective actions. This procedure is called ‘security analysis’. In order to monitor and determine the current state of the system, measurement acquisition is required. Devices called remote terminal units (RTU), which are available in substations, perform collection and transmission of various types of measurements. Nowadays, new instruments called intelligent electric devices (IED) are also used to complement the duties of RTUs. The measurements are transferred from all monitored substations via one of many possible types of communication links such as fiber optics, satellites, and microwaves to the Supervisory Control and Data Acquisition (SCADA) host computer at the control center. The measurements includes line flows, bus voltage and line current magnitudes, generator outputs, loads, circuit breaker and

switch status, transformer tap positions, and switchable capacitor bank values. The raw data and measurements are processed by the state estimators to filter the measurement noise, and to detect/identify gross errors. The state estimation solution provides an optimal state of the system using the measurements and the network model. The determined states are sent to the EMS functions such as contingency analysis, automatic generation control, corrective real and reactive power dispatch, load forecasting and optimal power flow to provide initial conditions. Figure 2.2 shows the configuration of a typical EMS/SCADA.

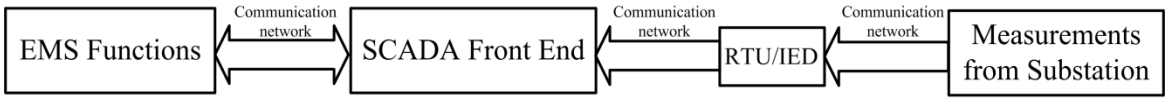

Figure 2.2. EMS/SCADA configuration [28]

Therefore, a power system state estimator is considered as the main component of the on-line security analysis. Various applications of the on-line security analysis procedure are shown in Figure 2.3 [28]. The solution of the state estimation is classified into one of the six categories mentioned above. In the case of normal state, contingency analysis is carried out to determine the system security against a set of predetermined contingencies. In the case of insecurity, EMS tools such as security constrained optimal power flow (SCOPF) are used to determine preventive control actions and move the system towards the desired normal and secure states.

The need for state estimators and the advent of the computers motivated Fred Schweppe to propose the use of state estimation (SE) in power systems [1-3]. On-line state estimators are embedded in the EMS to facilitate accurate and efficient monitoring of operational conditions. They help power operators with provision of a reliable real-time database of the system, including existing state. The database can be used to perform contingency analysis reliably and to determine any required corrective actions. The main functions of a state estimator are topology processor, observability analysis, bad data processing, and topology error identification [28]. The one-line diagram of the system is determined based on the status data of circuit breakers and switches provided by the topology processor. Observability analysis determines if the set of available measurements is sufficient to have a state estimation solution. It also identifies the unobservable branches and the observable islands in the system if any exists.

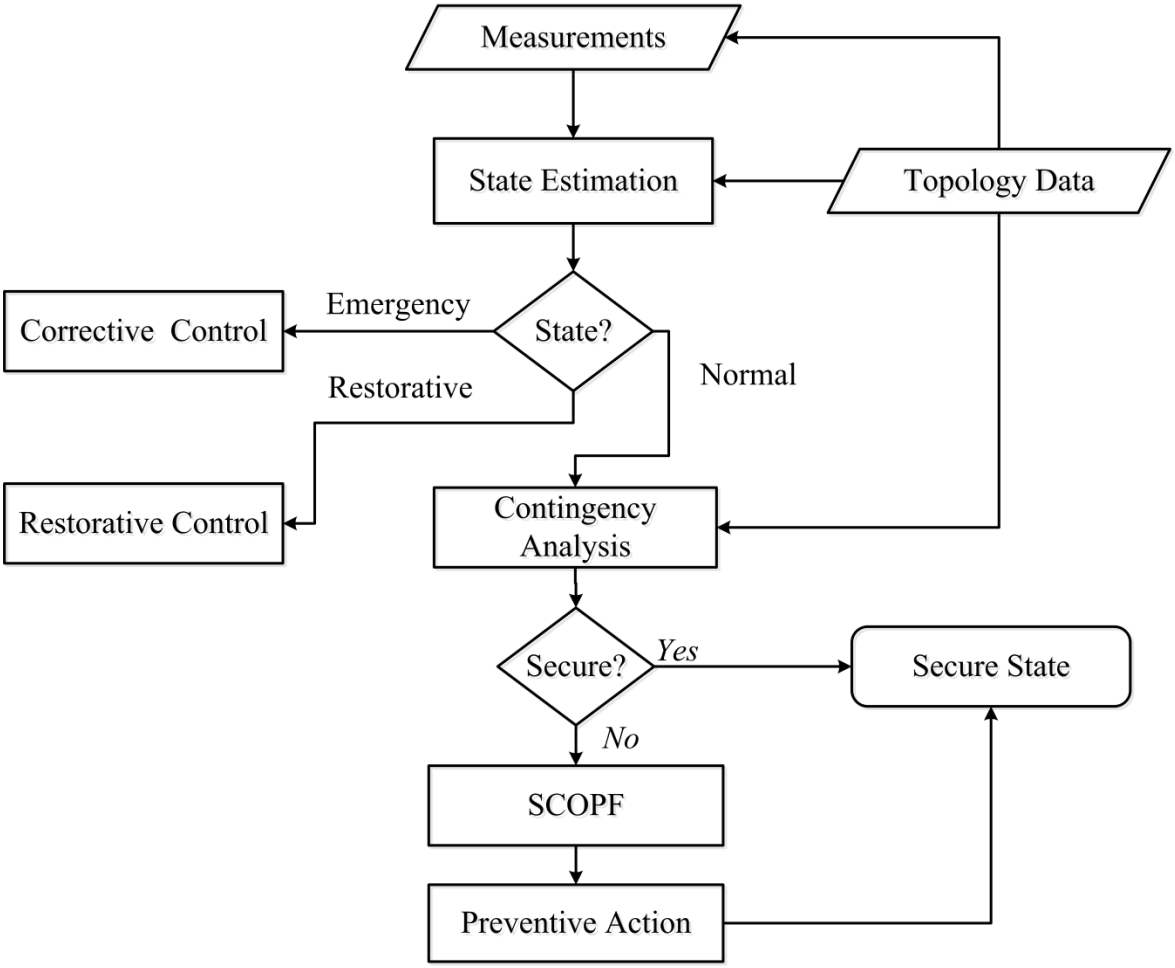

Figure 2.3. On-line Static Security Assessment: Functional Diagram [28]

State estimation solution determines the optimal estimate for the system state, i.e. complex bus voltages. A state estimator also functions as a filter against inaccurate or bad measurements, and other information received through the SCADA system. If there is enough redundancy in the measurement configuration, a state estimator can also identify and eliminate bad measurements or gross errors in the measurement set. This process is the result of bad data processing function. Those errors related to the topology information and/or network parameters can be identified by topology error processing. In the next section, models, assumptions, and the procedure for the state estimation are described.

B. Static State Estimation

In daily operation, the loads vary according to pattern cycles with small variations. Sudden load changes are infrequent and they happen due to the disconnection of industrial consumers,

broadcast of very popular television programs, scheduled outages of network components, adverse weather conditions etc. Therefore, considering the 'quasi-static' regime, the system operating condition is fully characterized at a given point in time using network parameters such as branch reactances, branch resistances, shunt capacitors, shunt inductors, and variables such as bus loads, generation, nodal injection power, line flows, and bus voltages (magnitudes and phase angles). These interdependent variables are related by Kirchhoff laws or so called power flow equations. Power flow equations are used to determine the nodal power injections, line flows, and bus voltages based on the load and generation. The bus voltages are defined as 'system states' rendering the power flow equations as nonlinear. The goal of static state estimation is to provide power system operators with complex voltages (phasors) of all of the system buses at a given point in time. Since, measurements are vulnerable to errors and telemetry failures, a set of redundant measurements is used in state estimation procedure in order to filter out such errors and find the best estimate. The measurements may include active and reactive power flows, voltage and current magnitudes, and synchronized voltage and current phasors. The assumptions behind the modeling as well as the component models are provided below.

1. Assumptions and Network Modeling

It is assumed that power system operates in a steady-state under balanced conditions. In balanced conditions, all branches are in three-phase and fully transposed, and the load at each phase is equal to each other. All series or shunt devices are in three-phase. Based on balanced and 3-phase assumptions, a single-phase positive sequence circuit is used as a 'one-line' diagram for the entire system. The obtained solution based on such a model is the positive sequence component of the system state during balanced steady state operation.

Transmission Lines

A two-port π-model is used to represent the branches, whose parameters correspond to the positive sequence equivalent circuit of transmission lines. A transmission line with a positive sequence series impedance of $R+jX$ and total line charging susceptance of jB, is modeled by the equivalent circuit shown in Figure 2.4.

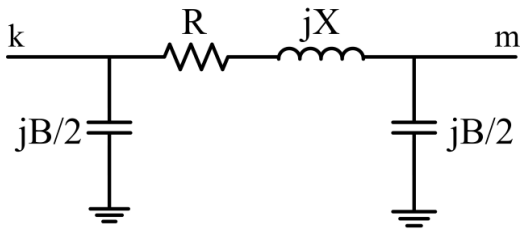

Figure 2.4. Equivalent circuit for a transmission line [28]

Shunt Capacitors or Reactors

Shunt capacitors or reactors are represented by their per-phase susceptance at the corresponding bus. The type of the shunt element is determined with respect to the sign of the susceptance value. For shunt capacitors, the sign is positive and for shunt reactors the sign is negative.

Tap-Changing and Phase-Shifting Transformers

A transformer with off-nominal but in-phase tap is modeled as a series impedance in series with an ideal transformer as shown in Figure 2.5[28].

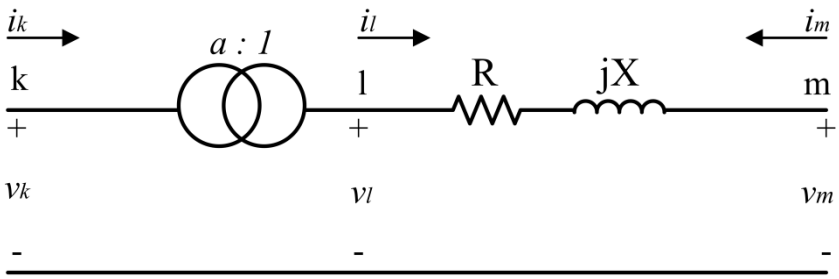

Figure 2.5. Equivalent circuit for an off-nominal tap transformer [28].

The terminal current injections of the two-port circuit in Figure 2.5 are given by:

$$\begin{bmatrix} i_k \\ i_m \end{bmatrix} = \begin{bmatrix} y/a^2 & -y/a \\ -y/a & y \end{bmatrix} \begin{bmatrix} v_k \\ v_m \end{bmatrix} \tag{2.1}$$

where *a* is the in-phase tap ratio. The equivalent circuit for the above set of nodal equations is shown in Figure 2.6 for a phase-shifting transformer, where the off-nominal tap value *a* is complex, the nodal equations change to [28]:

$$\begin{bmatrix} i_k \\ i_m \end{bmatrix} = \begin{bmatrix} y/|a|^2 & -y/a^* \\ -y/a & y \end{bmatrix} \begin{bmatrix} v_k \\ v_m \end{bmatrix} \tag{2.2}$$

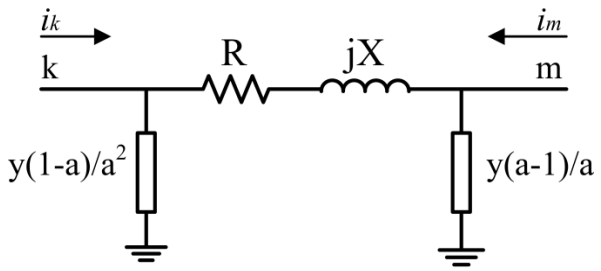

Figure 2.6. Equivalent circuit of an in-phase tap-changer [28]

Loads and Generators

Loads and generators, except constant-impedance type loads, are modeled as equivalent complex power injections and therefore have no effect on the network model [28]. Constant impedance type loads are modeled as shunt admittances at the corresponding buses.

Network Model

The network model is built based on the component models. A set of nodal equations are derived by applying Kirchhoff's current law at each bus as follows [28]:

$$\underline{I} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_N \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1N} \\ Y_{21} & Y_{22} & \cdots & Y_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{N1} & Y_{N2} & \cdots & Y_{NN} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = Y.\underline{V} \tag{2.3}$$

where,

$\underline{I}$ vector of net current injections

$\underline{V}$ vector of bus voltage phasors

i_k is the net current injection phasor at bus *k*.

v_k is the voltage phasor at bus *k*.

Y_{km} is the $(k,m)^{th}$ element of Y.

Diagonal and off-diagonal entries of bus admittance matrix, *Y*, are calculated as follows:

$$Y_{kk} = \sum_{m=1}^{N} y_{km}$$

$$Y_{km} = -y_{km}$$

where y_{km} is the admittance of the component connecting bus *k* to bus *m*. The shunt components are added to the relevant diagonal terms. The transformer model can be introduced by modifying the following four entries in Y as:

$$Y_{kk}^{new} = Y_{kk} + y/|a|^2 \qquad Y_{km}^{new} = Y_{km} - y/a^* \qquad Y_{mk}^{new} = Y_{mk} - y/a \qquad Y_{mm}^{new} = Y_{mm} + y$$

In the following section, the measurement models relating the measurements to system states are provided.

2. Measurement Models

The set of measurements are written as:

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} h_1(x_1, x_2, \cdots, x_n) \\ h_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ h_m(x_1, x_2, \cdots, x_n) \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \underline{h}(x) + \underline{v} \tag{2.4}$$

where:

$\underline{z}$ is the vector of measurements

$h_i(x)$ is the nonlinear function relating measurement *i* to the state vector x

$\underline{x}^T = [x_1, x_2, \cdots, x_n]$ is the system state vector

$\underline{h}^T = [h_1(x), h_2(x), \cdots, h_m(x)]$

$\underline{v}^T = [v_1, v_2, \cdots, v_n]$ is assumed to be a 'normal' error affecting the measurements and resulting from the limited accuracy of metering devices.

Regarding the statistical properties of the measurement errors, it is usually assumed that measurement errors are white noises with the following properties:

- $E[v_i] = 0\,, i = 1, \dots, m$
- Measurement errors are independent, i.e. $E[v_i v_j] = 0$

  Hence, $Cov(v) = E[v_i v_j] = R = diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2\}$

The standard deviation σ_i of each measurement *i* is calculated to reflect the expected accuracy of the corresponding meter and the communication devices. The following formula is used for the state estimator used by American Electrical Power (AEP) [29]:

$$\sigma_i = 0.0067\eta_i + 0.0016FS_i$$

where

$$\eta_i = \begin{cases} \sqrt{P_{km}^2 + Q_{km}^2} & for\ power\ flow\ k-m \\ \sqrt{P_k^2 + Q_k^2} & for\ power\ injection\ at\ k \\ |V_k| & for\ voltage\ magnitude\ at\ k \end{cases}$$

The Measurement Function, $\underline{\boldsymbol{h}}(\boldsymbol{x})$

The measurements consist of nodal voltage magnitudes, active and/or reactive line flows and active and reactive nodal injections. With the invention of PMU we can also have access to voltage phase angles. More detail about inclusion of PMU measurements in SE is provided in Section B.4 of this chapter. Typically, the total number of measurements ranges from two to six times of the number of state variables to ensure system observability. The ratio of the number of measurements to the number of network states is defined as 'redundancy'. In polar representation of a system with *n* buses, the state vector will have (*2n*-1) elements: *n* bus voltage magnitudes and (*n*-1) phase angles. The phase angle of reference (slack) bus is assumed to be zero. Assuming bus 1 as the reference bus, the state vector x will have the following form:

$$\underline{x}^T = [\theta_2, \theta_3, \cdots, \theta_n, V_1, V_2, \cdots, V_n]$$

Considering the general two-port π-model in Figure 2.7 for network branches, the expressions for each of the above types of measurements are given below.

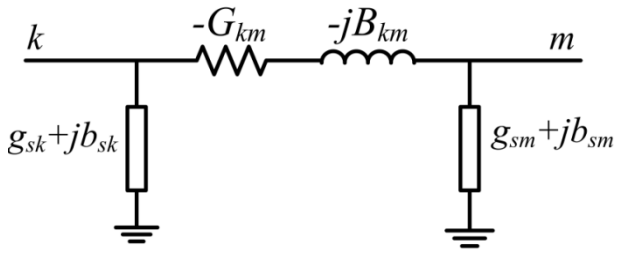

Figure 2.7. Two-port π-model of a network branch [28]

In Figure 2.7, $G_{km} + jB_{km}$ is the km^{th} element of the bus admittance matrix and they are calculated by using the parameters in Figure 2.6 as follows:

$$Y_{km} = G_{km} + jB_{km} = -y_{km} = -\frac{1}{R_{km} + jX_{km}}$$

$g_{sk} + jb_{sk}$ is the half of total line susceptance modeling the line-charging as also shown in Figure 2.4.

Real and reactive power injection at bus *k* [28]:

$$P_k = V_k \sum_{m \in N_k} V_m (G_{km} cos\theta_{km} + B_{km} sin\theta_{km})$$

$$Q_k = V_k \sum_{m \in N_k} V_m (G_{km} cos\theta_{km} - B_{km} sin\theta_{km})$$

Real and reactive power flow from bus *k* to bus *m*:

$$P_{km} = -{V_k}^2 (g_{sk} - G_{km}) + V_k V_m (G_{km} cos\theta_{km} + B_{km} sin\theta_{km})$$

$$Q_{km} = -{V_k}^2 (b_{sk} - B_{km}) + V_k V_m (G_{km} sin\theta_{km} - B_{km} sin\theta_{km})$$

Line current flow magnitude from bus *k* to bus *m*:

$$I_{km} = \frac{\sqrt{{P_{km}}^2 + {Q_{km}}^2}}{V_k}$$

Ignoring the shunt admittance $(g_{sk} + jb_{sk})$, we have:

$$I_{km} = \sqrt{\left({G_{km}}^2 + {B_{km}}^2\right)({V_k}^2 + {V_m}^2 - 2V_k V_m cos\theta_{km})}$$

where

V_k, θ_{ki} is the voltage magnitude and phase angle at bus *k*.

$$\theta_{km} = \theta_k - \theta_m$$

N_k is the set of bus numbers that are directly connected to bus *k*

The Measurement Jacobian, *H*

Ignoring the PMU measurements, the structure of the measurement Jacobian *H* is as follows [28]:

$$H = \begin{bmatrix} \frac{\partial P_{inj}}{\partial \theta} & \frac{\partial P_{inj}}{\partial V} \\ \frac{\partial P_{flow}}{\partial \theta} & \frac{\partial P_{flow}}{\partial V} \\ \frac{\partial Q_{inj}}{\partial \theta} & \frac{\partial Q_{inj}}{\partial V} \\ \frac{\partial Q_{flow}}{\partial \theta} & \frac{\partial Q_{flow}}{\partial V} \\ \frac{\partial I_{mag}}{\partial \theta} & \frac{\partial I_{mag}}{\partial V} \\ \frac{\partial V_{mag}}{\partial \theta} & \frac{\partial V_{mag}}{\partial V} \end{bmatrix}$$

The measurement Jacobian entries corresponding to the real power injection measurements are as follows [28]:

$$\frac{\partial P_k}{\partial \theta_k} = \sum_{m=1}^{N} V_k V_m(-G_{km} sin\theta_{km} + B_k cos\theta_{km}) - {V_k}^2 B_{kk}$$

$$\frac{\partial P_k}{\partial \theta_m} = V_k V_m(G_{km} sin\theta_{km} - B_{km} cos\theta_{km})$$

$$\frac{\partial P_k}{\partial V_k} = \sum_{m=1}^{N} V_m(G_{km} cos\theta_{km} + B_{km} cos\theta_{km}) + V_k G_{kk}$$

$$\frac{\partial P_k}{\partial V_m} = V_k(G_{km} cos\theta_{km} + B_{km} sin\theta_{km})$$

The measurement Jacobian entries corresponding to reactive power injection measurements are as follows:

$$\frac{\partial Q_k}{\partial \theta_k} = \sum_{m=1}^{N} V_k V_m(G_{km} cos\theta_{km} + B_{km} sin\theta_{km}) - {V_k}^2 G_{kk}$$

$$\frac{\partial Q_k}{\partial \theta_m} = V_k V_m(-G_{km} cos\theta_{km} - B_{km} sin\theta_{km})$$

$$\frac{\partial Q_k}{\partial V_k} = \sum_{m=1}^{N} V_m(G_{km} sin\theta_{km} - B_{km} cos\theta_{km}) - V_k B_{kk}$$

$$\frac{\partial Q_k}{\partial V_m} = V_k(G_{km} sin\theta_{km} - B_{km} cos\theta_{km})$$

The measurement Jacobian entries corresponding to real power flow measurements are as follows:

$$\frac{\partial P_{km}}{\partial \theta_k} = -V_k V_m(-G_{km} sin\theta_{km} + B_{km} cos\theta_{km})$$

$$\frac{\partial P_{km}}{\partial \theta_m} = V_k V_m(-G_{km} sin\theta_{km} + B_{km} cos\theta_{km})$$

$$\frac{\partial P_{km}}{\partial V_k} = -V_m(-G_{km} cos\theta_{km} - B_{km} sin\theta_{km}) + 2(-G_{km} + g_{sk})V_k$$

$$\frac{\partial P_{km}}{\partial V_m} = -V_k(-G_{km} cos\theta_{km} - B_{km} cos\theta_{km})$$

The measurement Jacobian entries corresponding to reactive power flow measurements are as follows:

$$\frac{\partial Q_{km}}{\partial \theta_k} = V_k V_m(-G_{km} cos\theta_{km} - B_{km} sin\theta_{km})$$

$$\frac{\partial Q_{km}}{\partial \theta_m} = V_k V_m(-G_{km} cos\theta_{km} - B_{km} sin\theta_{km})$$

$$\frac{\partial Q_{km}}{\partial V_k} = V_m(G_{km} sin\theta_{km} - B_{km} cos\theta_{km}) + 2(B_{km} - b_{sk})V_k$$

$$\frac{\partial Q_{km}}{\partial V_m} = V_k(G_{km} sin\theta_{km} - B_{km} cos\theta_{km})$$

The measurement Jacobian entries corresponding to voltage magnitude measurements are as follows:

$$\frac{\partial V_k}{\partial V_k} = 1, \frac{\partial V_m}{\partial V_m} = 0, \frac{\partial V_k}{\partial \theta_k} = 0, \frac{\partial V_k}{\partial \theta_k} = 0$$

The measurement Jacobian entries corresponding to current magnitude measurements (ignoring the shunt admittance of the branch) are as follows:

$$\frac{\partial I_{km}}{\partial \theta_k} = \frac{{G_{km}}^2 + {B_{km}}^2}{I_{km}} V_k V_m sin\theta_{km}$$

$$\frac{\partial I_{km}}{\partial \theta_m} = -\frac{{G_{km}}^2 + {B_{km}}^2}{I_{km}} V_k V_m sin\theta_{km}$$

$$\frac{\partial I_{km}}{\partial V_k} = \frac{{G_{km}}^2 + {B_{km}}^2}{I_{km}}(V_k - V_m cos\theta_{km})$$

$$\frac{\partial I_{km}}{\partial V_m} = \frac{{G_{km}}^2 + {B_{km}}^2}{I_{km}}(V_m - V_k cos\theta_{km})$$

In the following section, application of weighted least squares (WLS) technique to state estimation is briefly reviewed.

3. The WLS State Estimator

Maximum likelihood estimation (MLE) is used to apply WLS method to power system state estimation to find the best estimate of the bus voltages. MLE is widely used in statistics, and we find it necessary to briefly review the concept.

Consider a sample $\{z_1, z_2, \ldots, z_m\}$ of m one-dimensional observations that are independent and identically distributed (i.i.d). It is assumed that the observations are distributed as $f(z_1, z_2, \ldots, z_m;\ \theta) = f(\underline{z}; \theta)$, where $f(\underline{z}; \theta)$ is a function of $\underline{z}$ given θ. θ is the parameter of location of the joint density function to be estimated. θ belongs to the sample space and its quantity is assumed to be deterministic. The likelihood function is defined as [30]:

$$L(z_1, z_2, \ldots, z_m;\ \theta) = cf(\underline{z}; \theta)$$

If we consider c=1, the likelihood function reduces to the assumed joint probability density function (p.d.f.). It is a function of θ for fixed $\underline{z}$. Assuming that the m observations are independent, the joint p.d.f. is equal to the product of the marginal density functions:

$$f(\underline{z}; \theta) = f(z_1; \theta) f(z_2; \theta) \ldots f(z_m; \theta) \tag{2.5}$$

Therefore, the likelihood function becomes:

$$L(\underline{z}; \theta) = cf(\underline{z}; \theta) = \prod_{i=1}^{m} c_i\, f(z_i; \theta)$$

The product $c_i f(z_i; \theta)$ can be regarded as the probability that z_i falls in the interval [30]:

$$\left[z_i - \frac{c_i}{2}, z_i + \frac{c_i}{2}\right]$$

Hence, $L(\underline{z}; \theta)$ is considered as the likelihood function of the sample $\underline{z}$ for a given θ. The likelihood function is proportional to the probability distribution of the sample. The value of θ

that maximizes $L(\underline{z};\theta)$ makes the sample the most likely to occur. The estimator that maximizes $L(\underline{z};\theta)$ is defined as a *maximum likelihood estimator* (MLE). And, it is given as [30]:

$$\hat{\theta}_{ML} = argmax\, L(z_1, z_2, \ldots, z_m;\, \theta) = argmax\, f(z_1, z_2, \ldots, z_m; \theta)$$

$\hat{\theta}_{ML}$ is the value of θ that makes the measurements the most likely. Given (2.5), we have:

$$\hat{\theta}_{ML} = argmax \prod_{i=1}^{m} f(z_i; \theta)$$

Hence, an optimization problem can be set up in order to maximize the likelihood function as a function of θ. In state estimation, the measurement errors are commonly assumed to have a Gaussian (Normal) distribution and the parameters for such a distribution are its mean and its variance. The problem of maximum likelihood estimation is then solved for these two parameters [28]. The normal p.d.f. for a random variable z is defined as:

$$f(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left\{\frac{z-\mu}{\sigma}\right\}^2} \tag{2.6}$$

where

z : random variable,

μ : mean (or expected value) of z

σ : standard deviation of z

Referring to (2.5) the joint p.d.f. can simply be expressed as the product of individual p.d.f. 's if each measurement is assumed to be independent of the rest:

$$f_m(z) = f(z_1) f(z_2) \ldots f(z_m)$$

where z_i : *i*th measurement and $\underline{z}^T = [z_1, z_2, \ldots, z_m]$

Hence, MLE can be obtained by solving the maximization problem of log $f_m(z)$:

$$Maximize\ \log f_m(\mathrm{z}) = Maximize \sum_{\mathrm{i}=1}^{\mathrm{m}} \log f(\mathrm{z}_i)$$

$$= Maximize \left\{ -\frac{1}{2}\sum_{i=1}^{m}\left(\frac{z_i-\mu_i}{\sigma_i}\right)^2 - \frac{m}{2}\, log\, 2\pi - \sum_{i=1}^{m} log\, \sigma_i \right\}$$

The maximization problem can be rewritten as minimization problem as follows:

$$Minimize \sum_{i=1}^{m} \left(\frac{z_i - \mu_i}{\sigma_i}\right)^2 \tag{2.7}$$

Equation 2.7 can also be expressed in terms of the measurement residuals, which are defined as:

$$r_i = z_i - \mu_i \tag{2.8}$$

where μ_i or $E(z_i)$ can be expressed as $h_i(x)$. Square of each residual r_i is weighted by W_{ii}, which is inversely related to the error variance for that measurement, R_{ii}. Thus, the minimization problem in (2.7) can be re-written as:

$$Minimize \sum_{i=1}^{m} W_{ii} {r_i}^2 \tag{2.9}$$

$$s.t.\ z_i = h_i(x) + r_i$$

The solution of the above optimization problem is called the WLS estimation for *x*. For power system state estimation, WLS problem can be written as:

$$minimize\ J(x) = \ minimize\ \ \sum_{i=1}^{m} \left(z_i - h_i(x)\right)^2 / R_{ii} = [\underline{z} - \underline{h}(x)]^T W [\underline{z} - \underline{h}(x)] \tag{2.10}$$

The first-order optimality conditions to find the minimum leads to the following expression:

$$g(x) = \frac{\partial J(x)}{\partial x} = -H^T(x) W \left[\underline{z} - \underline{h}(x)\right] = 0 \tag{2.11}$$

where, $H(x) = [\frac{\partial h(x)}{\partial x}]$

Using the iterative Newton method to find the minimum leads to:

$$\underline{x}^{k+1} = \underline{x}^k - \left[G\left(\underline{x}^k\right)\right]^{-1} g\left(\underline{x}^k\right) \tag{2.12}$$

where

k is the iteration index,

$\underline{x}^k$ is the solution vector at iteration *k*,

$$G\left(\underline{x}^k\right) = \frac{\partial g\left(\underline{x}^k\right)}{\partial x} = H^T\left(\underline{x}^k\right) W H\left(\underline{x}^k\right) \tag{2.13}$$

$$g\left(\underline{x}^k\right) = -H^T\left(\underline{x}^k\right) W \left(\underline{z} - \underline{h}\left(\underline{x}^k\right)\right) \tag{2.14}$$

$G\left(\underline{x}\right)$ is a sparse and symmetric matrix and is called the 'gain' matrix. All the eigenvalues of the gain matrix are non-negative for a fully observable network. The gain matrix is typically decomposed into triangular factors instead of inversion in order to reduce the computational

complexity. Therefore, the following sparse linear set of equations is solved using forward/back substitutions at each iteration *k*:

$$\left[G\left(\underline{x}^{k}\right)\right]\Delta\underline{x}^{k+1} = H^{T}\left(\underline{x}^{k}\right)W\left(\underline{z} - \underline{h}\left(\underline{x}^{k}\right)\right) \tag{2.15}$$

where $\Delta\underline{x}^{k+1} = \underline{x}^{k+1} - \underline{x}^{k}$.

The iterative solution algorithm for WLS state estimation problem is shown in Figure 2.8. As in the case of the power flow solution, the initial guess corresponds to the flat voltage profile (flat start), where all bus voltages magnitude are assumed to be 1 per unit and voltage phase are assumed to be 0. In each iteration *k*, measurement function $\underline{h}\left(\underline{x}^{k}\right)$ and measurement Jacobian $H\left(\underline{x}^{k}\right)$ are calculated. Also, the gain matrix is built, and then decomposed into triangular factors, and forward/backward substitutions are performed to solve for $\Delta\underline{x}^{k}$[28]. Incorporation of PMU measurements to a conventional SE is described in the following section.

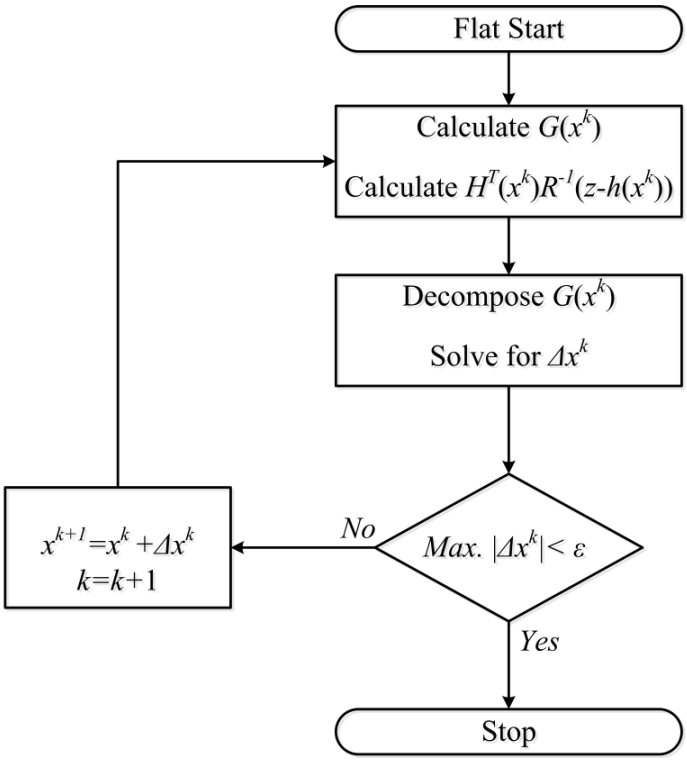

Figure 2.8. Iterative solution algorithm for WLS state estimation

4. Incorporation of PMU Measurements in SE

PMUs are used in different applications to improve dynamic- and steady-state power system monitoring. PMUs significantly help the power system state estimators by providing time-

stamped and synchronized measurements with high accuracies. The time difference of two consecutive scans of PMUs is 1/30th of a second which is on average 10–15 times better than the arrival frequency of conventional measurements. PMUs also enable the operators to identify the unobservable islands and restore the observability within a second. [31] reports that when sufficient numbers of phasor measurements are used along with other conventional measurements, the accuracy of the state estimation is significantly improved. However, using current and voltage phasor measurements in a traditional state estimator requires significant modifications to the existing EMS software. We name the traditional measurement as $\underline{z}_1$ with measurement error covariance matrix of R_1 and a set of positive sequence voltage and current phasors from PMU as $\underline{z}_2$ with measurement error covariance matrix of R_2. The appended measurement vector is obtained by adding the current and voltage phasors to the traditional measurement vector as follows:

$$\underline{z} = \begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \end{bmatrix} = \begin{bmatrix} \underline{z}_1 \\ \underline{V}_r \\ \underline{V}_i \\ \underline{I}_r \\ \underline{I}_r \end{bmatrix}$$

The subscripts *r* and *i* represent both the real and imaginary parts of the phasor measurements. The voltage and current phasors are nonlinear functions of the state vector. The nonlinear function relates the state vector in polar form to phasor measurement vector in rectangular form. Therefore, the new measurement function and their error covariance matrix are as follows:

$$\begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \end{bmatrix} = \begin{bmatrix} \underline{h}_1(x) \\ \underline{h}_2(x) \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \end{bmatrix} \tag{2.16}$$

$$R = \begin{bmatrix} R_1 & 0 \\ 0 & {R_2}^a \end{bmatrix}$$

where ${R_2}^a$ is the measurement error covariances from the phasor measurements which reflects the transformation of the state vector from rectangular form to polar form. The measurement Jacobian matrix is similarly constructed as:

$$H(x) = \begin{bmatrix} H_1(x) \\ H_2(x) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial H_1(x)}{\partial x} \\ \dfrac{\partial H_2(x)}{\partial x} \end{bmatrix} \tag{2.17}$$

According to (2.10), we have:

$$\underline{x}^{k+1} = \underline{x}^{k} - \left[G\left(\underline{x}^{k}\right)\right]^{-1}\left[H_1^T\left(\underline{x}^{k}\right)R_1^{-1}\left(\underline{z}_1 - \underline{h}_1\left(\underline{x}^{k}\right)\right) + H_2^T\left(\underline{x}^{k}\right)R_2^{-1}\left(\underline{z}_2 - \underline{h}_2\left(\underline{x}^{k}\right)\right)\right] \tag{2.18}$$

where k is the iteration index, $\underline{x}^{k}$ is the solution vector at iteration k.

$$G\left(\underline{x}^{k}\right) = H_1^T\left(\underline{x}^{k}\right)R_1^{-1}H_1\left(\underline{x}^{k}\right) + H_2^T\left(\underline{x}^{k}\right)R_2^{-1}H_2\left(\underline{x}^{k}\right)$$

Therefore, the algorithm can be computed in a similar way to the one in traditional state estimator. As the downfall of including phasor measurements in this way is that it requires significant changes to the existing code, an alternative way to include phasor measurements with traditional state estimator softwares is presented in [32]. The alternative approach uses the results of a traditional state estimator using conventional measurements and combines this solution with the phasor measurements as a post-processing linear estimator. If the phasor measurements are obtained in rectangular form, the results of traditional state estimator in polar form should also be converted to rectangular form. The combined measurement equation is given as:

$$\underline{z}^{new} = \begin{bmatrix} \underline{\hat{V}}_r \\ \underline{\hat{V}}_j \\ \underline{V}_r \\ \underline{V}_j \\ \underline{I}_r \\ \underline{I}_j \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \\ I^a & 0 \\ 0 & I^a \\ C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \begin{bmatrix} \underline{V}_r \\ \underline{V}_j \end{bmatrix} \equiv A \begin{bmatrix} \underline{V}_r \\ \underline{V}_j \end{bmatrix} \tag{2.19}$$

where I in the above equation represents a unity matrix, whereas the I^a represents a unity matrix with zeros on the diagonal where no voltage phasor has been measured. The matrices C_1 through C_4 are composed of line conductances and susceptances for those lines with current phasor measurements. C matrices are similar to the bus admittance matrix, with those nonzero entries corresponding to branches where current phasors are measured. Equation (2.19) is linear and leads to a weighted least-squares solution for the system state:

$$\underline{\hat{x}} = \left[A^T R^{a\,-1}\, A\right]^{-1}\left[R^a\right]^{-1} A\, \underline{z}^{new} \tag{2.20}$$

where A is defined in (2.19), and R^a is the covariance matrix as follows:

$$R^a = \begin{bmatrix} {R_1}^a & 0 \\ 0 & {R_2}^a \end{bmatrix}$$

where both error covariance sub-matrices have been transformed through rotation to account for the use of rectangular coordinates. [32] demonstrates that the linear state estimator as a post-

processing tool produces the same results with the traditional state estimator using simultaneous processing of phasor measurements and conventional measurements. It is also recognized that PMU data can be used to improve the network observability, to aid in bad data processing, and to accurately and frequently update the network topology. In the next section, fundamentals of observability analysis are briefly reviewed.

5. Observability Analysis

The network observability analysis determines whether the network is observable with the measurement configuration. It has the following main tasks:

1. It determines whether a state estimation solution can be obtained using the set of measurements received at the last measurement scan,
2. It determines the unobservable and/or observable islands if the network is not fully observable.
3. It recommends appropriate locations to place pseudo-measurements to make the network fully observable.

This analysis can be carried out off-line or on-line. Telecommunication errors, topology changes or meter failures may occasionally lead to cases rendering the parts or the entire network unobservable. If unobservable branches are identified, the system will contain several observable islands, each one having its own angle reference. Observability of a given network is determined by the type and location of the available measurements as well as by the topology of the network. Thus, graph theory is utilized in the analysis of network observability. In the next section, numerical and topological methods for observability analysis are briefly reviewed.

a) Numerical observability analysis method

Network observability method assumes that the power measurements (flows and injections) come in *PQ* pairs. Therefore, analyzing the observability of θ based on P using DC power flow is sufficient. Note that, unlike θ, the voltage solution requires a measured reference bus. Hence, it should be further checked to ensure that at least one voltage measurement exists in each observable island.

Numerical observability analysis uses linearized measurement model and measurement errors are ignored due to their irrelevance in the observability analysis:

$$\Delta \underline{z} = H \Delta \underline{x} \tag{2.21}$$

where:

$\Delta \underline{z} = \underline{z} - \underline{h}(\underline{x}_0)$, is the mismatch between the measurement vector and its calculated value at an estimate $\underline{x}_0$, $\Delta \underline{x} = \underline{x} - \underline{x}_0$ and H is the measurement Jacobian matrix. The WLS estimate $\Delta \underline{\hat{x}}$ is given by [28]:

$$\Delta \underline{\hat{x}} = (H^T W H)^{-1} H^T W \Delta \underline{z} \tag{2.22}$$

A unique solution for $\Delta \underline{\hat{x}}$ can be calculated if $(H^T W H)$ is nonsingular or when H has full column rank. The system observability is independent of the branch parameters and operating state of the system. In this method, all the branches are assumed to have 1 pu reactance and all the voltages are set to 1. The DC power flow along the system branches can be written as [28]:

$$\underline{P_f} = A^T \underline{\theta} \tag{2.23}$$

where,

A: branch-bus incident matrix

$\underline{\theta}$: vector of the voltage phase angle

$\underline{P_f}$: vector of branch flows

If the estimated state $\hat{\theta}$ is zero, then all branch flows will be zero according to (2.23). Using the DC measurement model:

$$\underline{z_A} = H_{AA} \underline{\theta} \tag{2.24}$$

where $\underline{z_A}$ is the real power measurement vector and H_{AA} is the decoupled Jacobian for the real power measurements. The WLS estimate for $\underline{\hat{\theta}}$ is given by:

$$\underline{\hat{\theta}} = (H_{AA}^T H_{AA})^{-1} H_{AA}^T \underline{z_A} \tag{2.25}$$

A null estimate for $\underline{\hat{\theta}}$ is obtained for an observable system when all system measurements, $\underline{z_A}$, i.e. flows and injections, are zero. If there exists an estimate of $\underline{\hat{\theta}}$, which satisfies the measurement equation.

$$H_{AA} \underline{\hat{\theta}} = 0 \tag{2.26}$$

and if a nonzero branch flow is calculated by:

$$\underline{\hat{P}_f} = A^T \underline{\hat{\theta}} \neq 0 \quad (2.27)$$

then, $\underline{\hat{\theta}}$ is called an unobservable state. Furthermore, those branches with nonzero flows are referred to as unobservable branches. The algorithm is as follows:

1. Remove all irrelevant branches. These are branches that have no incident measurements.
2. Form the decoupled linearized gain matrix for the *P*-θ estimation problem:

$$G_{AA} = H_{AA}^T R^{-1} H_{AA} \quad (2.28)$$

3. Factorize G_{AA}, modify the zero pivots and the right hand side vector. Since G_{AA} is singular, at least it has one zero pivot during the factorization. Because, even for fully observable systems the reference bus phase angle is included in the state vector $\underline{\theta}$. The zero pivot is replaced by 1.0 and an arbitrary value of choice is assigned to the corresponding entry of the right hand side vector. The assigned arbitrary values should be distinct from each other and therefore integer numbers in increasing order, such as 0, 1, 2, etc. are assigned.
4. Solve for $\underline{\hat{\theta}}$ using $G_{AA}\underline{\hat{\theta}} = H_{AA}^T R^{-1} H_{AA} \underline{z}_A$
5. Compute $\underline{\hat{P}_f} = A^T \underline{\hat{\theta}}$ and identify all unobservable branches where $\underline{\hat{P}_f}(i) \neq 0$. Remove all unobservable branches and all injections that are incident to these unobservable branches.
6. If no more unobservable branches are found, then determine the observable islands separated by the unobservable branches and stop. Else, go to step 2.

In the next section, topological method for observability analysis is briefly described.

b) Topological observability analysis method

Observability analysis can also be performed using the topological method based on graph theory. In this method, no calculation is used and decision is based on the information about the network topology, measurement type and their location. The actual parameters of the network elements are not used in any part of this analysis. Also, it is assumed that the measurements come in real and reactive pairs. Therefore, the real part of the decoupled (or DC) measurement model is used for the observability analysis. In this model, the error-free real power flow and injection measurements are linearly related to the bus voltage phase angles (excluding the slack bus) based on Equation (2.24). The system is fully observable if a tree can be formed and each branch of this tree contains a power flow measurement, then the phase angles at all buses can be determined. Therefore, power flow measurements are assigned to their respective branches in the first step and it is tried to form a spanning tree, i.e. a tree that reaches each and every bus in the

system. If the spanning tree cannot be found, a forest with several smaller size trees is provided. In that case, injection type measurements are used to reduce the size of the forest. If the procedure results in a single tree, the system is observable [28].

Topological Observability Algorithm

There are several ways to implement topological observability analysis. In the followings, the essential steps of the algorithm are summarized [33]:

1. Assign each flow measurement to its respective branch.
2. Check if there is a spanning tree that covers every bus without creating a tree. The spanning tree can have branches. If a spanning tree is found stop. The system is observable, if not there are disconnected forest and go to 3.
3. Assign each injection measurement to an incident branch such that the forests are connected. Make sure there is no loop with/within the forest.
4. Discard the flow measurements that form a loop with the forest or do not help connecting the forests. Those measurements are redundant.
5. Discard those injection measurements that have at least one incident branch which does not form a loop with the branches of a forest (or a sub-forest) AND which is not part of any of forest.
6. Update the forest and repeat step 5 for each injection measurement until no injections are needed to remove.

In the following section, bad data detection and identification, is briefly reviewed.

6. Bad Data Detection and Identification

Bad data processing is one of the most important functions of a state estimator. The existence of bad data in a measurement set is determined by bad data detection procedure. After detection of the bad data, bad data identification procedure is used to determine the measurements that contain bad data. This function detects measurements errors if the error is not in a 'critical measurement', and then, identifies and eliminates them unless the bad data is multiple interacting and conforming, which will be addressed later in this section. If the removal of a measurement renders the system unobservable, then this measurement is called 'critical'. The error in a critical measurement is not detectable and using this measurement will result in a biased state estimation solution. The measurement system in a power system has to be configured such that critical

measurements are avoided. However, due to loss of communication channels or elimination of bad measurements, critical measurements can occur in the measurement set and identification of these measurements is a challenge addressed in [34-36].

Measurement errors can be due to various reasons. Finite accuracy of the meters and the telecommunication medium, failures/unexpected noise caused by interference in telecommunication system can result in random errors and large deviations in measurements. Large measurement errors can also occur when the meters have biases, drifts or wrong connections. Also, incorrect topology information can deceive the state estimator and relevant measurements connected to the components with wrong topological information will be interpreted as bad data by the state estimator. In presence of redundant measurements, such errors can be filtered by the state estimator [28]. Some of the bad data such as negative voltage magnitudes, measurements with several orders of magnitude larger or smaller than expected values, or large differences between incoming and leaving currents at a connection node, are obvious and can be easily detected and eliminated by simple plausibility checks. There are other types of bad data that cannot be easily detected by such means and measurement residuals are processed to detect and identify these during in WLS estimation process for non-critical (i.e. redundant) measurements. In the following section, bad data detection procedure is reviewed.

a) Bad data detection

Bad data detection refers to the determination of the existence of any bad data, while bad data identification tries to find out the specific measurement in error. Configuration of the measurement set in a network plays an important role in detection and identification of bad data. In the following section, two methods that are used to detect bad data are reviewed.

Chi-squares test

In WLS state estimation, the objective function $J(x)$ in (2.10) is approximated with function $f(x)$ as follows:

$$f(x) = \sum_{i=1}^{m} R_{ii}^{-1}\, e_i^2 = \sum_{i=1}^{m} \left(\frac{e_i}{\sqrt{R_{ii}}}\right)^2 = \sum_{i=1}^{m} (e_i^N)^2 \tag{2.29}$$

where e_i is the *i*th measurement error and approximates the measurement residual, r_i in Equation (2.9), R_{ii} is the diagonal entry of the measurement error covariance matrix and m is the total number of measurements. Assuming that e_i are all normally distributed random variables with

zero mean and R_{ii} variance, e_i^N's will have a standard normal distribution. Therefore, $f(x)$ will have a χ^2 distribution with at most (*m* - *n*) degrees of freedom. In a power system, at least *n* measurements is required to satisfy the power balance equations; thus, at most (*m*-*n*) of the measurement errors are linearly independent. Hence, the largest degree of freedom can be (*m*-*n*), (i.e. the difference between the total number of measurements and the system states). The steps of the Chi-squares test are given as follows [28]:

- Solve the WLS estimation problem, look up the value from the Chi-squares distribution table corresponding to a detection confidence with probability of $p = Pr(J(\hat{x}) \leq \chi^2_{(m-n),p}$ (e.g. 95%) and (*m*-*n*) degrees of freedom. Let this value be $\chi^2_{(m-n),p}$,
- if $J(\hat{x}) \geq \chi^2_{(m-n),p}$, then bad data is suspected,
- Else, it is assumed that there is no bad data in the measurements.

Normalized Residuals

Approximation of residual errors by Equation (2.29) using Chi square-test may result in failure in bad data detection [28]. Therefore, using the normalized residuals can provide us with a more accurate test. Normalized residual for measurement *i* can be obtained by simply dividing its absolute value by the corresponding diagonal entry in the residual covariance matrix:

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}} = \frac{|r_i|}{\sqrt{R_{ii}S_{ii}}} \tag{2.30}$$

where Ω_{ii} is residual covariance matrix and S is called the residual sensitivity matrix. This matrix, represents the sensitivity of the measurement residuals to the measurement errors and it calculated as follows:

$$S = I - K = HG^{-1}H^TW \tag{2.31}$$

The normalized residual vector r^N has a standard normal distribution. The largest element in r^N can be compared to a statistical threshold to determine the existence of the bad data. This threshold depends on the desired level of detection sensitivity. After detecting bad data in a measurement set, future processing can be performed to identify them. Normalized residuals can also be used to identify the bad data. In the next section, bad data identification using normalized residuals is reviewed.

b) Bad data identification

Once the bad data is identified, the measurements in error are either deleted from the measurement set or replaced by pseudo-measurements. This modified measurement set is then used by the WLS estimator to obtain the state estimate. Among the existing methods, *Largest Normalized Residual* test is described here.

Largest Normalized Residual Test

Largest normalized residual test is based on the properties of normalized residuals for a single non-critical bad data existing in the measurement set. This test can be used to identify and subsequently to eliminate the bad data. The test procedure is shown in Figure 2.10 [28]:

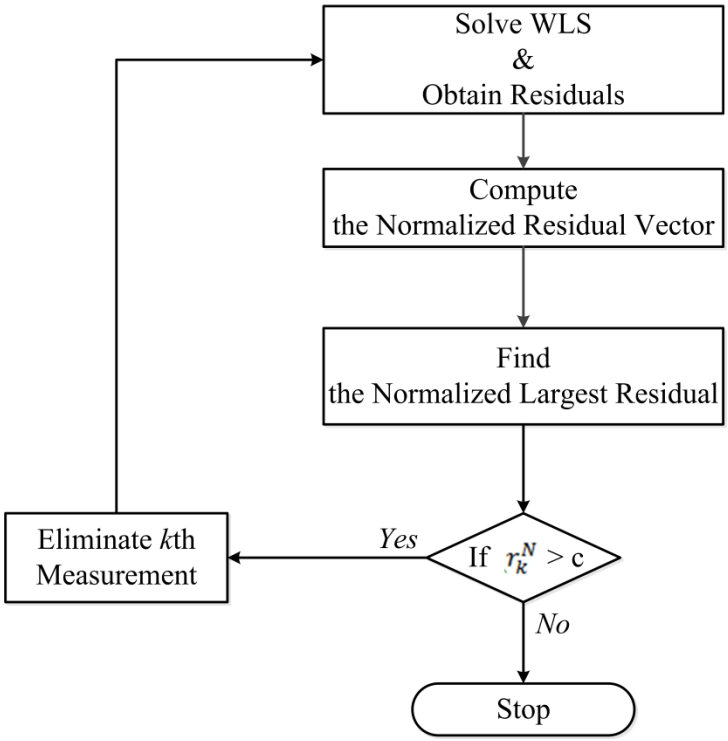

Figure 2.9. Bad data identification with r_{max}^N test method

The performance of this technique for multiple bad data depends on type and configuration of the measurements. In order to analyze the identification of multiple bad data, *residual sensitivity matrix, S* in Equation (2.31) is observed. Multiple bad data are categorized as follows:

- Non-interacting: if $S_{ik} \approx 0$, measurements *i* and *k* are classified as non-interacting. In this case, if both measurements *i* and *k* contain bad data, normalized residual test is able to identify them.

- Interacting, non-conforming: if $S_{ik} \gg 0$ measurements i and k are called interacting. In this case, if the errors in measurements i and k are not consistent with each other, the method may still identify them.
- Interacting, conforming: if $S_{ik} \gg 0$ and the errors in the measurements are consistent with each other. In this case, the method may fail to identify the bad data.

In order to address the weakness of normalized residual test in case of interacting bad data, two techniques using Hypothesis Testing Identification (HTI) are proposed [37, 38]. In the proposed techniques, measurements errors are estimated and used as random variables of concern, then decision is made based on hypothesis testing, which takes into account the statistical properties of the measurements errors estimates. HTI can identify error in the cases of single, multiple non-interacting and interacting bad data. It can also be easily implemented on any WLS-based state estimators [38].

Other Techniques

Numerous algorithms are proposed for detection and identification of bad data. Merrill and Schweppe [39] propose a bad data suppression (BDS) estimator based on a non-quadratic cost function which is changed to the least-squares estimator in the absence of bad data. Their basic idea is to change the cost function so that the effect of bad data is minimized and reasonable state estimates can be obtained. Handschin *et al*. [40] expand this idea to a number of different non-quadratic cost functions. They also suggest associating lesser weight values to measurements with large residuals, to improve the performance of the bad data processing in state estimation. D. M. Falco *et al*. [41] suggest that the method can also be used for tracking state estimation to improve the robustness of the estimator in presence of bad data. Irving *et al*. [42] show how linear programming (LP) can be used to suppress the effect of bad data while estimating the state of the system. The attempts to develop one-step algorithms that simultaneously reject bad data and estimate the state of a power system lead to weighted least absolute value (WLAV) algorithm. The basic difference between WLAV and WLS approximation is that the WLAV approximation is obtained by minimizing the sum of the absolute values of the residuals whereas the LS approximation is obtained by minimizing the sum of the squares of the residuals. The WLAV method is more robust than WLS in presence of bad data, but it is computationally more expensive. Since in this dissertation WLS is used, LAV is not elaborated. For further details refer to [43].

C. Summary

A detailed background of power system static state estimation is provided in this chapter. The models for transmission lines and measurements used in power system state estimation are briefly reviewed. State estimation using weighted least squares technique is reviewed followed by a general summary of observability analysis, bad data detection, identification and elimination techniques. It is described that, in the conventional state estimation a snap shot of the measurements are used to provide the operator with the best estimate of the state. Therefore, real time monitoring of the system is achieved by performing state estimation for every new measurement. Since consecutive runs of state estimation with new measurements are computationally expensive, tracking state estimators are introduced under certain assumptions. Next chapter reviews the applications of tracking state estimation in power systems.

CHAPTER III

POWER SYSTEM TRACKING STATE ESTIMATION

A. Introduction

Tracking state estimators are proposed in 1970s [24] for faster state estimation solutions. These estimators assume that the change in system state between two successive instants of time is extremely small. Therefore, the state vector is simply updated by an amount proportional to the measurement residual at every instant. The difference between the static and tracking estimators is better summarized by the block diagrams shown in Figure 3.1 and 3.2 [44].

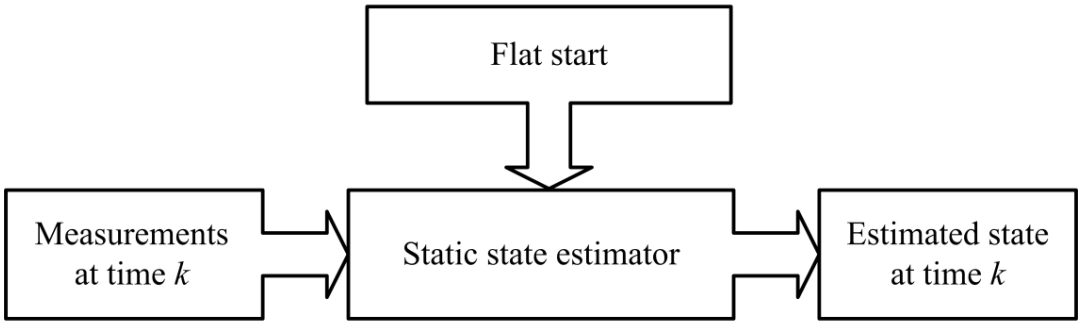

Figure 3. 1. Static state estimator

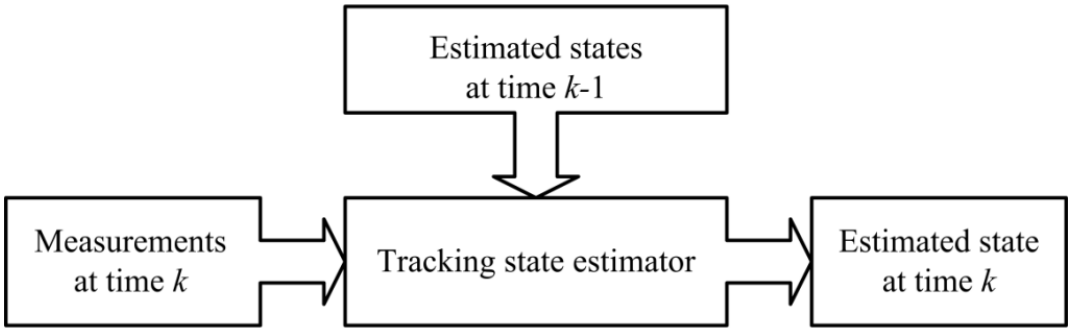

Figure 3.2. Tracking state estimator

Tracking estimator updates the estimated state at the previous time step and does not need the full execution of the state estimation algorithm at that instant. Hence, this type of estimator provides the operator with an easy and fairly accurate monitoring of the power system. Various techniques are proposed in the literature to improve the accuracy and the implementation of tracking state estimation. In the next sections, the models and approaches for the tracking state estimation are reviewed.

B. Assumptions and Network Modeling

Power system modeling and assumptions for the network components and measurements are the same as in static state estimation (Chapter II.1). Power system state changes are assumed to be extremely small in a short span of time (seconds to a few minutes). Therefore, the states are assumed to change linearly in short timeframes. If we know the estimated state, $\underline{\hat{x}}_k$, at the instant of *k*, the mathematical model for updating the states at time *k*+1 using the arrival measurements at time *k*+1 can be written as follows:

$$\underline{x}_{k+1} = \underline{x}_k + \Delta\underline{x}_{k+1} + \underline{w}_k \tag{3.1}$$

where $\Delta\underline{x}_{k+1}$ represents the change in the state *k*+1 due to the changes the power system undergoes within the time interval, Δk.

C. Methods of Updating the State

Different methods are introduced for updating the power system state in the tracking state estimation. The formula for calculating $\Delta\underline{x}_{k+1}$ based on WLS method is as follows [24]:

$$\Delta\underline{x}_{k+1} = G_k^{-1} H_k^T R^{-1} [\underline{z}_{k+1} - \underline{h}(\underline{\hat{x}}_k)] \tag{3.2}$$

where G_k is the same as the gain matrix in state estimation Equation (2.12) which remains constant, *h* is the measurement function in Equation (2.4), *H* is the Jacobian of *h* and *R* is the measurement error covariance matrix. It is observed that, the equation is the same as Equation (2.15). However, $\underline{z}_{k+1}$ is replaced with $\underline{z}_k$. In this way, the state vector for the next sampling time is updated using the new measurement at time *k*+1. Hence, the computationally heavy G_k matrix and its inverse are calculated only once for updating the estimation using the new measurements. In this way, tracking estimation provides an easy update to the existing state vector. The decoupled nature of the power system states can also be used in tracking state estimation under decoupling assumptions. The procedure is the same as in static state estimation except the right hand side where $\underline{z}_{k+1}$ is replaced with $\underline{z}_k$ in Equation (2.15).

The linear programming (LP) technique, which is used in static state estimation, is also suggested for tracking state estimation. In [42], the authors use LP methods to minimize the sum of the measurement residuals. This technique is more robust in presence of bad data. W. W. Kotigua discusses LAV based tracking algorithm in [43]. A.K. Sinha *et al.* [45] proposed a

tracking estimator, which uses the second order Taylor's expansion of the measurement function, $\underline{h}(\underline{x})$. This technique uses a constant gain matrix (for a given topology) resulting in faster convergence. Applying second order Taylor expansion to $\underline{h}(\underline{x})$ around $\underline{x}_0$ results in:

$$\underline{h}(\underline{x}) = \underline{h}(\underline{x}_0) + H(\underline{x}_0)\Delta\underline{x} + S(\underline{x}_0) \tag{3.3}$$

where $H(\underline{x}_0)$ is the Jacobian evaluated at $\underline{x}_0$ and $S(\underline{x}_0) = \frac{\partial^2 h(x)}{\partial^2 x}\Delta x^2$ is the matrix involving second order derivatives of the measurement function at x_0. Substituting equation (3.3) into (2.13) results in:

$$[G(\underline{x}^k)]\Delta\underline{x}^{k+1} = H^T(\underline{x}^k)R^{-1}\left(\underline{z} - \underline{h}(\underline{x}^k) - S(\underline{x}^k)\right) \tag{3.4}$$

In the filtering stage, $S(\underline{x}^k)$ is updated and the gain and Jacobian matrices remain constant. In this way, measurement nonlinearities are considered and tracking estimation can be performed fast enough.

In this section, three alternative approaches for updating the power system using TSE are briefly reviewed. LAV and LP are used to improve robustness of the tracking state estimator in presence of bad data. And, second order Taylor's series expansion are used to incorporate the nonlinearities of the measurements model. In the next section, the advantages of using TSE in EMS are described.

D. Advantages

Tracking state estimation enables the operator to receive a real-time update of the system without actually performing the entire state estimation. A brief summary of the advantages of tracking state estimators is as below [44]:

- Continuous tracking and monitoring of the system helps the system operator to take better and faster decisions especially in case of an emergency,
- The lighter computational load of this approach make it feasible for fast and online implementation,
- Sequential processing techniques can be used to monitor the system if scanning of measurements takes a long time.

E. Summary

In this chapter, capabilities of tracking state estimation is reviewed. These estimators can be used for fast and real-time monitoring of power systems due to their lower computational complexity compared to the conventional static state estimators. However, these estimators are not able to provide the operator with a forecasted state and their performance deteriorates in presence of sudden changes in system states between two consecutive scans. Tracking state estimators are not reliable in presence of distributed, large- and small-scale intermittent renewable penetration and microgrids. In the next chapter, forecasting-aided state estimators, which have the potential to be reliable in the future grid if they employ more accurate state transition models, are reviewed.

CHAPTER IV

FORECASTING-AIDED POWER SYSTEM STATE ESTIMATION

A. Introduction

Static and tracking state estimators are briefly reviewed in the previous chapters as they are the necessary functions to efficiently monitor the network conditions and to determine necessary preventive and corrective control actions in a traditional grid with conventional generation sources. Debs and Larson address the lack of forecasting ability in static and tracking state estimators in 1970. In [6] they propose a 'dynamic' state estimator (DSE) using Extended Kalman Filter (EKF). Note that, here 'dynamic' term is not associated with the stability concept or with the oscillatory response of the system during a transient period in which the states of power system are defined as the rotor angles and the machine frequencies [7, 15, and 46]. In this dissertation, the dynamics of concern are those imposed by variations in load and generation. It is assumed that the frequency oscillations caused by these changes are already damped out. The timeframe of interest is in the order of several seconds to few minutes. The system is supposed to be in a post-dynamic, balanced, sinusoidal and in steady-state [47].

In order to avoid confusion with dynamic state estimators concerned with the machine dynamics [7-15, 46], DSE is also referred to as forecasting-aided state estimation (FASE) [16]. FASE incorporates the present (and sometimes previous) state of the power system along with system's physical model for state vector forecasting at the next time instant. Once the measurements at the next instant arrive, the predicted state is filtered to obtain the best estimate. FASE improves the real-time operations by providing a 'predictive' database of power system states. The predictive database increases the measurement redundancy by providing pseudo-measurements and helps re-establishing observability due to loss of measurements, bad data detection and elimination, and topological error identification [16-19]. Moreover, a predictive database improves the real-time power system operation by providing short-term trajectory of the system state for short-term scheduling and congestion management [16, 20]. Availability of a predictive database also provides the system operator with a longer time for taking preventive or corrective actions [47]. The analysis of the innovation vector along with normalized residuals in FASEs (and in early DSEs) is used to detect sudden load and topology changes [21]. The innovation vector is also

used to identify presence of bad data in critical measurements, which is not possible by analyzing the residual vector in traditional state estimators [19]. FASE can provide optimal estimates of interconnected network states for systems in which some measurements are delayed due to a problem in communication channels [22]. In the following section, FASE using Extended Kalman Filter (EKF) is briefly described.

B. Overview of FASE using Extended Kalman Filter (early DSE)

The well-known procedure for EKF-based FASE (early DSE) is comprised of three main stages:

1. State modeling
2. State forecasting
3. State filtering

The flowchart of FASE is shown in Figure 4.1. The network modeling and assumptions are same as in the case of static and tracking state estimation. The details are provided in Chapter II.1.

In addition to the observation equation given by (2.4), FASE also requires a dynamic model, which represents the state transition. The general form of the model is:

$$\underline{\tilde{x}}_{k+1} = f\left(\underline{\hat{x}}_k, \underline{u}_k, \underline{w}_k\right) \tag{4.1}$$

where, k is the time sample, $\underline{\hat{x}}_k$ is the filtered/estimated state vector with the dimension of ($N\times1$) for a n-bus system with N states ($2n$-1), $\underline{\tilde{x}}_{k+1}$ is the forecasted stated vector, $\underline{w}_k$ represents the uncertainties in the model, f represents an n-dimensional non-linear state transition model, and $\underline{u}_k$ accounts for any unknown control actions on the system. Control variables are ignored, since their effect is much faster than the timeframe of interest. The early dynamic estimators are based on Kalman filter algorithm using the measurements and system models. Since the measurement vector is nonlinear in nature, extended Kalman filter (EKF) is used in most cases [44]. This provides the linear minimum variance estimate of the state vector $\underline{x}$.

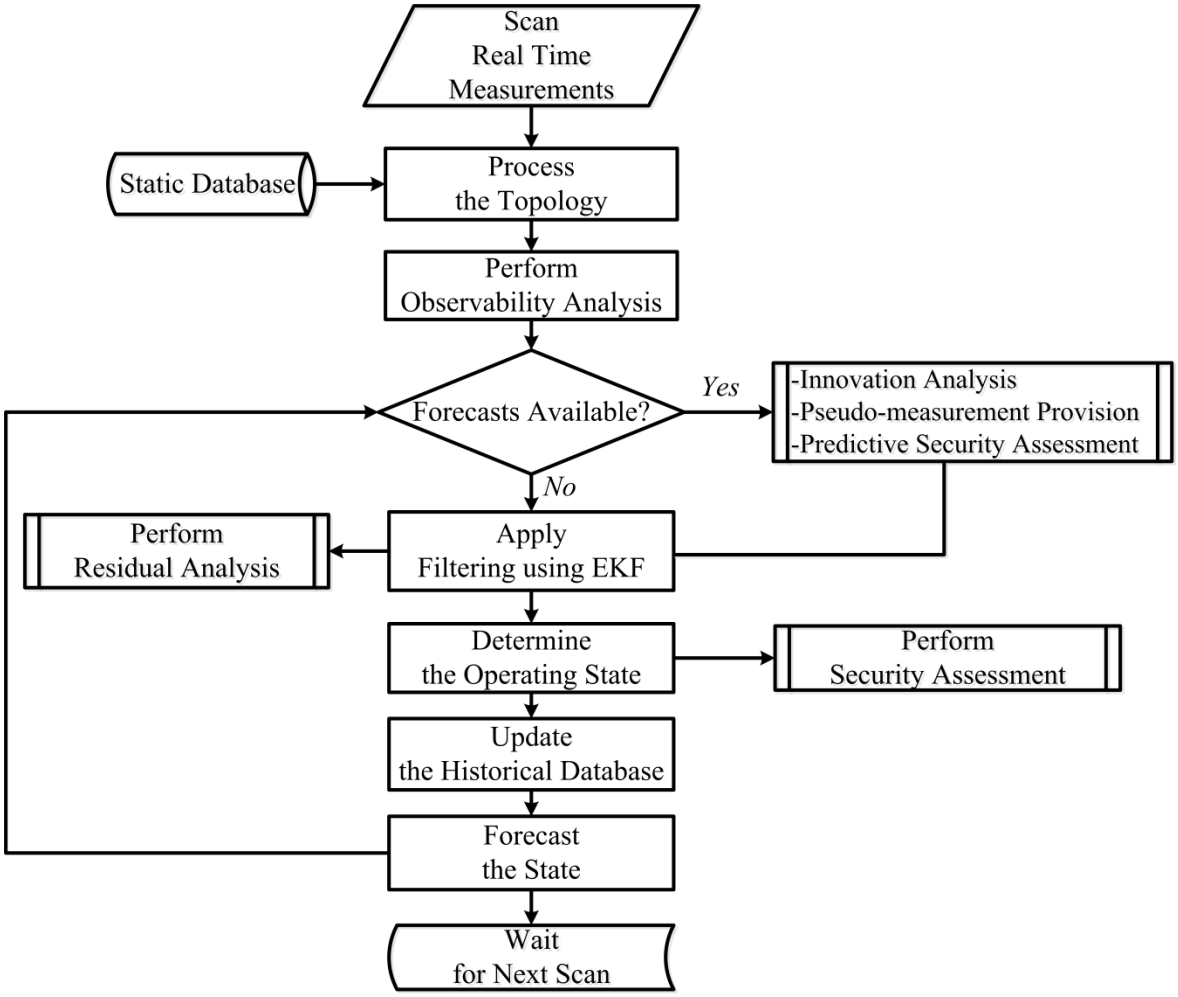

Figure 4.1. FASE algorithm

The corresponding equations of EKF in the state forecasting step are formulated as follows [16]:

$$\underline{\tilde{x}}_{k+1} = F_k \underline{\hat{x}}_k + \underline{g}_k \tag{4.2}$$

$$M_{\tilde{x}_{k+1}} = F_k \Sigma_{\hat{x}_k} F_k^T + Q_k \tag{4.3}$$

where F_k is the state transition matrix with the dimension of ($N \times N$) and $\underline{g}_k$ is associated with the trend behavior of the state trajectory. The forecasted state's error covariance matrix, $M_{\tilde{x}_{k+1}}$, as well as filtered state's error covariance matrix $\Sigma_{\hat{x}_k}$ are updated in this stage. Under normal operating conditions of a quasi-dynamic power system, the F_k is assumed to be diagonal and determined once and kept constant over time. $\underline{g}_k$ is either ignored, determined once and kept constant or updated in certain intervals. Q_k can be obtained offline for different normal operating conditions considering the maximum rate of change each state variable presents, which can be

derived from historical data of the system state behavior [47, 48]. However, in most applications Q_k is assumed to be diagonal with constant entries such as 10^{-6} [21, 47]. In addition, initial values for the diagonal model error covariance matrix $\Sigma_{\hat{x}_k}$ are considered as 10^{-6} [21]. Once the predicted state vector for the next time instant, k+1 is obtained, it can be filtered as the measurements arrive at k+1. The predicted state vector is used for obtaining the final estimate in FASE, the objective function is taken as sum of measurement residuals and the predicted state vector residuals. The filtering step can be seen as an extended WLS estimation in a sense that the received measurements $\underline{z}$ are processed together with the available forecasted state $\underline{\tilde{x}}$ to produce filtered state $\underline{\hat{x}}_k$. The state filtering stage aims at minimizing the following objective function [16]:

$$J(\underline{x}) = \left[\underline{z} - \underline{h}(\underline{x})\right]^T R^{-1}\left[\underline{z} - \underline{h}(\underline{x})\right] + [\underline{x} - \underline{\tilde{x}}]^T M_{\tilde{x}_{k+1}}^{-1}[\underline{x} - \underline{\tilde{x}}] \tag{4.4}$$

where $\underline{z}$ is the measurement vector, $\underline{h}$ is the power flow equations relating the power system state to the measurements and R is the measurement error covariance matrix. The state filtering state is performed as below:

$$\underline{\xi}_{k+1} = \underline{z}_{k+1} - \underline{h}(\underline{\tilde{x}}_{k+1}) \tag{4.5}$$

$$K_{k+1} = M_{\tilde{x}_{k+1}} H^T(\underline{\tilde{x}}_{k+1})\{H(\underline{\tilde{x}}_{k+1}) M_{\tilde{x}_{k+1}} H^T(\underline{\tilde{x}}_{k+1}) + R\}^{-1} \tag{4.6}$$

$$\Sigma_{\hat{x}_{k+1}} = (I - K_{k+1}H_{k+1}) M_{\tilde{x}_{k+1}} \tag{4.7}$$

$$\underline{\hat{x}}_{k+1} = \underline{\tilde{x}}_{k+1} + K_{k+1}\underline{\xi}_{k+1} \tag{4.8}$$

$$\underline{r} = \underline{z}_{k+1} - \underline{h}(\underline{\hat{x}}_{k+1}) \tag{4.9}$$

where H is the Jacobian of h, K is the gain matrix, I is identity matrix and $\underline{r}$ is the residual. The term $\underline{\xi}_{k+1}$ in Equation (4.5) is generally referred as the 'innovation vector' and Equations (4.6) to (4.9) constitute the state filtering or refining stage based on EKF. The filtering procedure starts with $\underline{x}^{(0)} = \underline{\tilde{x}}_{k+1}$ and iterate l times until convergence is attained (i.e. filtered state $\underline{x}^{(l)} = \underline{\hat{x}}_{k+1}$ together with its error covariance matrix $\Sigma^{(l)} = \Sigma_{k+1}$) [16]. The FASE procedure based on iterative EKF is shown as the following flow chart including the Equations (4.2)-(4.8):

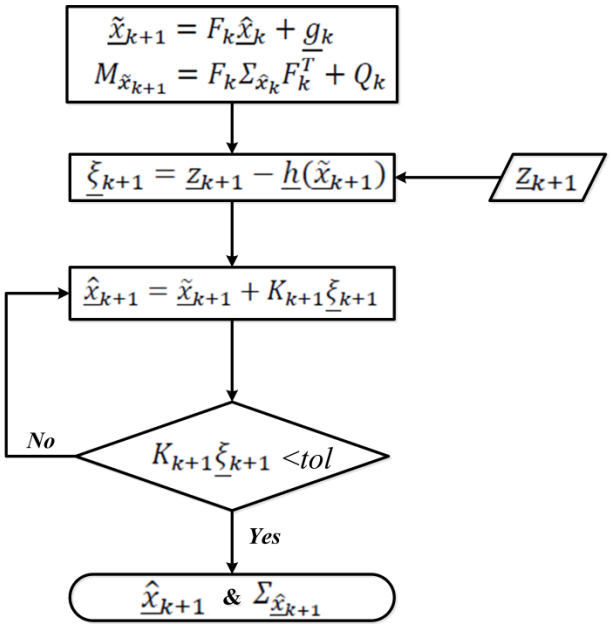

Figure 4.2. FASE algorithm based on iterative EKF

Innovation and residual vector

The innovation vector, $\underline{\xi}_{k+1}$ in equation (4.5), is used to detect and identify bad measurements. The innovation vector is the difference between the incoming measurements and the previously predicted values. Under the assumption that the measurement errors are Gaussian-distributed, innovation vector has zero-mean Gaussian distribution with a covariance as follows [21]:

$$\begin{aligned}\Psi = cov\left(\underline{\xi}_{k+1}\right) &= HM_{\tilde{x}_{k+1}}H^T + R \\ &= H[F_k\Sigma_{\hat{x}_k}F_k^T + Q_k]H^T + R\end{aligned} \tag{4.10}$$

where H is the measurement Jacobian, M is the covariance matrix of $\underline{\tilde{x}}$ and R is the covariance matrix of the measurement error.

$$H[F_k\Sigma_{\hat{x}_k}F_k^T + Q_k]H^T + R$$

The hypothesis that there is no gross error or sudden change will thus be accepted if:

$$\xi_{i_N} = |\xi_i|/\sqrt{\Psi_{i,i}} < \lambda \quad \forall i$$

where the threshold λ is chosen in accordance with the diagonal entries of (4.10). The innovation vector has the following noteworthy property: in the presence of one or multiple bad data at time k, only the components of $\underline{\xi}$ relevant to the erroneous measurements will exhibit abnormal values. In other words, there will be no smearing effect, unlike in conventional static state

estimators, where a gross error in a single measurement may cause several residuals to grow beyond the statistical threshold. Equation (4.5) shows that $\underline{\xi}$ is the direct estimate of incoming measurement error and it relies on the *a priori* information, $\underline{\tilde{x}}_{k+1}$. Post-filtering procedures are based on the measurement residuals:

$$\underline{r} = \underline{z}_{k+1} - \underline{h}(\underline{\hat{x}}_{k+1}) \tag{4.11}$$

The covariance matrix of the residual vector is given by [18]:

$$P = R - H\Sigma_{\hat{x}_{k+1}}H^T \tag{4.12}$$

The hypothesis that there is no gross error or sudden change will thus be accepted if

$$r_{i_N} = |r_i|/\sqrt{P_{i,i}} < \gamma \quad \forall i$$

The related literature on FASE techniques is reviewed in the following section.

C. Literature Review

A comprehensive literature survey on FASE techniques is provided in this section which is divided into three main parts. In the first part, state modeling approaches are reviewed. In the second part, alternative methods for state filtering are described and finally in the third part, alternative methods for FASE are surveyed.

1. State Modeling Approaches

The first step in FASE involves the identification of an approximate mathematical model for the time behavior of the power system in steady-state operation. The challenge is that there is no such deterministic state transition model for power systems in steady-state operation. Finding a deterministic state transition model that represents the state variation due to change in load and generation is not possible due to the high non-linearity of the relationship between the input (i.e. load and generation) and the output (i.e. complex bus voltages.) Considering the general mathematical model used for a nonlinear dynamic system (Equation 4.1) and the assumptions for EKF application, a generic linear model for FASE is assumed as [16]:

$$\underline{x}_{k+1} = F_k\underline{x} + \underline{g}_k + \underline{w}_k \tag{4.13}$$

where, F_k is the state transition matrix with the dimension of ($N \times N$), $\underline{g}_k$ is associated with the trend behavior of the state trajectory and $\underline{w}_k$ represents the uncertainties in the model with a white Gaussian noise with zero mean and covariance of Q_k as in (4.3). Various methods are proposed in the literature to identify the parameters of F_k. In this section, these methods are reviewed.

a) Naïve model

Debs and Larson [6], credited with the original paper on FASE (as early DSE), assume a simple linear model with F_k assumed to be an identity. In this model, changes in state vector are considered to be extremely small that can be replaced by a zero mean Gaussian noise (i.e. $\underline{g}_k$ assumed to be zero). Hence, the Equation (4.13) reduces to:

$$\underline{x}_{k+1} = \underline{x}_k + \underline{w}_k \tag{4.14}$$

Nishiya *et al.* [49] propose a trend for the state variation as follows:

$$\underline{x}_{k+1} = \underline{x}_k + \underline{g}_k + \underline{w}_k \tag{4.15}$$

where, $\underline{\hat{g}}_k = \underline{\hat{x}}_k - \underline{\hat{x}}_{k-1}$.

A major inadequacy of this model is its consideration of the trend factor as a deterministic input while in reality it is random in nature. In the next section, application of exponential smoothing methods for power system state modeling is reviewed.

b) Silva's model

Holt's two-parameter linear exponential smoothing (HES) technique is widely used for state modeling. The use of this method is first proposed by Silva in [47]. Authors in [18, 50-51], use Silva's approach to obtain the values of F_k and $\underline{g}_k$. However, in order to reduce the computational burden, both the state transition matrix and the covariance matrix are assumed to be diagonal. This means that the system states (i.e. complex bus voltages) are assumed to be independent of each other. Three equations and two smoothing constants are used in the model as follows [52]:

- The exponentially-smoothed series or current level estimate are calculated by:

$$\underline{s}_k = \alpha \underline{y}_k + (1-\alpha)(\underline{s}_{k-1} + \underline{b}_{k-1}) \tag{4.16}$$

- The trend is estimated by:

$$\underline{b}_k = \beta(\underline{s}_k - \underline{s}_{k-1}) + (1-\beta)\underline{b}_{k-1} \quad (4.17)$$

- And, the forecast of *m* periods into the future are calculated by:

$$d_{k+m} = \underline{s}_k + m\underline{b}_k \quad (4.18)$$

where, $\underline{s}_k$ is the estimation of the level of the series at time *t*, α is the smoothing constant for the data, $\underline{y}_k$ is the new observation or actual value of series in period *t*, β is the smoothing constant for trend estimate, $\underline{b}_k$ is the estimate of the slope of the series at time *t*, m is the periods to be forecast into the future, and d_{k+m} is forecast m periods into the future. The weight α and β can be selected subjectively or by minimizing a measure of forecast error such as root-mean-square error (RMSE). Large weights result in more rapid changes in the component and small weights result in less rapid changes. The initialization process for Holt's linear exponential smoothing requires to estimate the values for $\underline{s}_1$ and $\underline{b}_1$. One alternative for $\underline{s}_1$ is to set $\underline{s}_1 = \underline{y}_1$. Initial value of $\underline{b}_1$ can be zero, $y_1 - y_2$ or $\frac{y_4 - y_1}{3}$. The application of HES to determine the state transition is as follows [47]:

Let $\hat{x}_k(i)$ be the i^{th} component of the filtered state vector $\underline{\hat{x}}_k$. If $\underline{\tilde{x}}_k(i)$ and $\underline{\tilde{x}}_{k+1}(i)$ are predictions at time *k* and *k* + 1, respectively, HES method gives:

$$\tilde{x}_{k+1}(i) = a_k(i) + b_k(i) \quad (4.19)$$

where,

$$a_k(i) = \alpha(i)\hat{x}_k(i) + (1-\alpha(i))\tilde{x}_k(i) \quad (4.20)$$

$$b_k(i) = \beta(i)[a_k(i) - a_{k-1}(i)] + (1-\beta(i))b_{k-1}(i) \quad (4.21)$$

Substituting (4.20) and (4.21) in (4.19) and re-arrange the resulted equation as (4.3) we have:

$$\tilde{x}_{k+1}(i) = F_k(i)\hat{x}_k + g_k(i) \quad (4.22)$$

where,

$$F_k(i) = \alpha(i)(1+\beta(i))$$

$$g_k(i) = \big(1+\beta(i)\big)\big(1-a_k(i)\big)\tilde{x}_k(i) - \beta(i)a_{k-1}(i) + \big(1-\beta(i)\big)b_{k-1}(i)$$

In [47], α and β are set to 0.8 and 0.5, respectively. In the next section, applications of artificial neural networks in the forecasting stage of FASE are reviewed.

c) Artificial Neural Network based model

Artificial Neural Network (ANN) is used in [19, 21] for FASE. The adopted ANN architecture, corresponding to a linear model, is shown in Figure 4.3. ANN is used for forecasting one state variable at a time and therefore only one output neuron is considered.

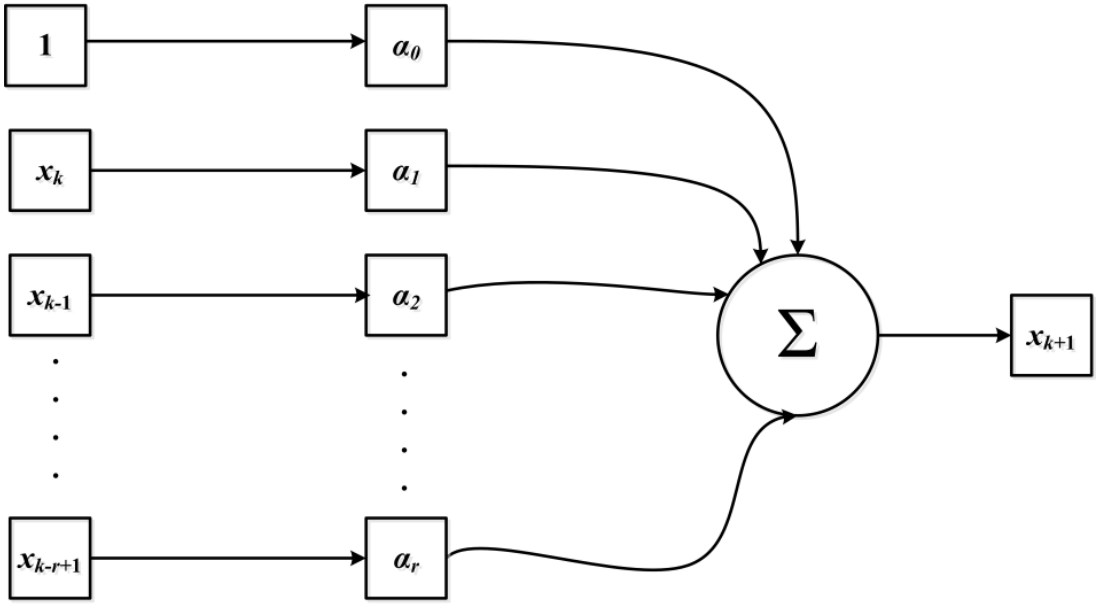

Figure 4.3. ANN architecture for power system state modeling

Considering the series of r values for the state vector, i.e. $\underline{x}_k, \underline{x}_{k-1}, \dots, \underline{x}_{k-r+1}$. The state transition equation can be written as:

$$\underline{x}_{k+1} = \underline{\alpha}_0 + \alpha_1 \underline{x}_k + \alpha_2 \underline{x}_{k-1} + \cdots + \alpha_r \underline{x}_{k-r+1} + \underline{w}_k \tag{4.23}$$

The parameters $\alpha_0, \alpha_1, \cdots, \alpha_r$ are the ANN interconnection weights and they are computed during training procedure. The number of input variables is r+1. In [19], the above equation represents (4.13), and each state is only forecasted by its own historical values. However, the number of history terms used in the training procedure is not specifically mentioned. ANN is also applied in [53] in power system state modeling stage, where two different ANN models have been used for state-forecasting and state-filtering steps. In the state forecasting step, 1,000 training patterns are generated using power flow simulations. 800 patterns are used for training and the remaining patterns are used for testing. In order to provide a broad range of operating points, the demand at each bus is varied linearly from 20% to 120% of the base load. Note that, impacts of adjacent buses are not considered in both [19] and [53] during the training procedure

and the transition matrix is assumed to be diagonal. In the next section, application of load forecasting for power system state forecasting would be reviewed.

d) Load-based models

It can be observed that the dynamic models mentioned above suffer from a common drawback: the state variables are assumed to vary independently from each other. This assumption is not realistic in electric power systems, where a change in a load at a specific node affects voltages at several nodes. In order to address this deficiency, Durgaprasad [54] assumes that the complex voltage at any bus at any instant is not only dependent on its value at the previous instant but also on the last available change in voltages at the incident buses. Application of nodal analysis at the bus *i* at time instant *k* and *k*+l yields [54]:

$$V_{k+1}^{i} = V_{k}^{i} - \frac{1}{Y_{ii} + Y_{k}^{Li} + Y_{k}^{Gi}} \sum_{j=1}^{n} Y_{ij} \left[{V_{k+1}}^{j} - {V_{k}}^{j} \right] + w_{k}^{i} \ , \qquad j \neq i \tag{4.24}$$

The above equation can be written in matrix form as follows:

$$\underline{V}_{k+1} = \underline{V}_{k} - \alpha_{k} \left[\underline{V}_{k+1} - \underline{V}_{k} \right] + \underline{w}_{k} \tag{4.25}$$

where, the matrix α_k is defined as [54]:

$$\alpha_{k} = \begin{cases} 0, & if\ i = j\ or\ bus\ i\ and\ j\ not\ connected \\ \dfrac{Y_{ij}}{Y_{ii} + Y_{k}^{Li} + Y_{k}^{Gi}}, & if\ i \neq j\ and\ bus\ i\ and\ j\ are\ connected \end{cases}$$

Y_k^{Li} and Y_k^{Gi} are the equivalent load and machine admittances at the i^{th} bus, respectively. Based on (4.25), the effects of the adjacent buses are included in the trend behavior of the state trajectory (i.e. $\underline{g}_k$ in Equation 4.2). Note that, the formulation in (4.25) has a fundamental error: Y_{ij} cannot be taken outside the summation in (4.24). In this model, the main driver of change in state from one sampling instant to the next is the uncertain demand change.

In a proposed alternative approach in [47], load value at time *k*+1 is assumed to be equal to the value at time *k*. In this approach, the state model that is used for state identification is as follows [47]:

$$\underline{x}_{k+1} = F_{k} \underline{x}_{k} + \underline{w}_{k}$$

where F_k is diagonal and time variant and its diagonal elements are used to define a parameter vector $\underline{c_k}$. A model for time evolution of $\underline{c_k}$ is defined based on the measurement model as follows:

$$\underline{c}_{k+1} = \underline{c}_k \tag{4.26}$$

$$\underline{z}_k = B_k \underline{c}_k + \underline{\varepsilon}_k \tag{4.27}$$

where, $B_k = H_k A_{k-1}$, A_{k-1} is diagonal matrix using elements of $\hat{\underline{x}}_{k-1}$ and $\underline{\varepsilon}_k = H_k \underline{w}_{k-1} + \underline{v}_k$ is white Gaussian noise. Then, the Kalman filter technique is used to find out the value of $\underline{c_k}$ due to linearity of the equations. Once $\underline{c_k}$ is found, F_k is formed and Kalman filter technique is applied to the dynamic system.

In [55], load prediction is used to build the model for state forecasting. The hourly load at every node is determined by a forecasting procedure. The complex voltage at each node for every hour is calculated by load flow. Then, a nominal (forecasted) hourly behavior of state variables with the following state transition matrix is obtained for a day in advance:

$$\begin{bmatrix} \dfrac{Real\,\{V_{k+d}\}}{Real\,\{V_k\}} & 0 \\ 0 & \dfrac{Img\{V_{k+d}\}}{Img\,\{V_k\}} \end{bmatrix}$$

where *d* is an integer from the hourly interval (0, 23) and V_k is the complex voltage. Using the above equation, the requested value of the state transition matrix is calculated at each hour. The state variables between two known points are calculated by interpolation methods. In [56], load forecasting is used to calculate the trend behavior of the state trajectory in the state forecasting model as follows:

$$\underline{x}_{k+1} = \underline{x}_k + J^{-1} \underline{\Delta \Gamma}_{k+1} + \underline{w}_k \tag{4.28}$$

where J^{-1} is the power injection Jacobian matrix, $\underline{\Delta \Gamma}_{k+1}$ represents the nodal injection change (real/active power injection) and $\underline{w_k}$ represents the uncertainties in model linearization and process noise modeled as a Gaussian noise with zero mean. The simulation is performed for 15-minute intervals and Jacobian is calculated on an hourly-base. Note that, adding $\underline{\Delta \Gamma}_{k+1}$, which represent information of *k*+1 in the right hand side of (4.28) is erroneous.

Some authors propose that instead of building a model for power system state, we should try to predict the bus load and then determine the state for the next time instant by power flow or static state estimation equations [48, 57]. This school of thought believes that it is a more realistic approach in the model determination stage of FASE because, 1) loads actually drive the system dynamics, 2) loads are more independent of each other on the contrary to bus voltages and 3) dynamics of bus loads can be identified through simple and accurate load forecasting models [16, 20]. In this method the prediction step is performed on the basis of bus power injections instead of the conventional state transition model. Active and reactive loads for each bus including the load at generator bus (if any), is predicted for next time instant and the corresponding state is calculated using power flow or state estimation equations. In this approach, alternative state vector $\underline{sv}$, composed of nodal injections is defined as follows [20]:

$$\underline{sv} = [\underline{P}^L, \underline{Q}^L, \underline{P}^G, \underline{V}^G]^T$$

where $\underline{P}^L, \underline{Q}^L$ are P-Q buses active and reactive loads, respectively and $\underline{P}^G, \underline{V}^G$ are active load and voltage of the P-V buses. $\underline{V}^G$ are general controlled variables and they can be used instead of $\underline{Q}^G$. Notice that the vector $\underline{sv}$ has the same number of components as the vector $\underline{x}(|V|, \theta)$, i.e. (*2n*-1). The predict nodal injections $sv(P, Q)$, are related to bus voltages $x(|V|, \theta)$, through set of *2n*-1 non-linear using power flow equations ($i.e.\ \underline{sv} = l(\underline{x})$). For the filtering and forecasting steps of the EKF, we need to have the covariance matrix of the state prediction error at the filtering step, $\Sigma_{\hat{x}_k}$, and the covariance matrix of the injection filtering error, at the prediction step $M_{\tilde{x}_{k+1}}$. These matrices are calculated as:

$$M_{\tilde{x}_{k+1}} = L_k^{-1} \Sigma_{\hat{x}_k} {L_k^{-1}}^T$$

$$\Sigma_{\hat{x}_k} = L_k M_{\tilde{x}_{k+1}} {L_k}^T$$

where L is the Jacobian of l respect to $\underline{x}$ [20].

The load dynamic is considered to be naïve or can be forecasted with the load forecasting methods. For example, the multivariate Box and Jenkins approach and the adaptive Kalman filter are used for the active and reactive bus load forecast, respectively [58]. However, the major difficulty to be faced by this alternative approach is in mapping the uncertainties of forecasted bus power injections to nodal voltages. In [59], a stochastic search technique combined with

scenario approach is used for state forecasting using load prediction. In the paper, it is tried to propose a solution for the discussed challenge in using nodal loads as the states.

In [57,60], artificial neural network (ANN) is used to predict bus load. The method uses ANN based short-term bus load prediction for the prediction step in the FASE. In the next section, application of fuzzy methods in FASE would be reviewed.

e) Application of fuzzy logic

In order to reduce the linearization errors in EKF based techniques, the nonlinearities in the measurements are taken into account. The proposed methods aim at reducing linearization errors during sudden load changes. A sliding-surface enhanced Fuzzy logic-based technique is proposed by Jeu-Min Lin *et al.* [51], where the error signals and the rate of error are integrally used. The input variable is defined by the linear equation as shown below:

$$v = \lambda e + \dot{e} \tag{4.29}$$

where λ is a positive constant, e is the state error and $\dot{e}$ is the rate of state error. The output variable is expressed as u_c. Here v_c is assumed to be an element of input vector $\underline{v}$ and u_c an element of output vector $\underline{u}$. The input and output are partitioned and tabulated along with the fuzzy rules. Then, the input values are fuzzified by assigning an appropriate membership function. Next step is the fuzzy inference, where the min-max method is used to obtain the fuzzy output for each input [51]. The next step is called the 'defuzzification' where the fuzzy inference for each input variable is converted to an output value u_c. This variable u_c for each input, serves as an element of the output vector $\underline{u}$. The equation for a new state vector is proposed as:

$$\underline{\tilde{x}}_{k+1} = F_k \underline{\hat{x}}_k + \underline{g}_k + \underline{u} \tag{4.30}$$

In (4.30), Silva's approach given in (4.22) is used to obtain the F_k and $\underline{g}_k$.The term $\underline{u}$ is close to zero when the system is in steady-state but becomes large when there are drastic changes in load and hence accounting for the nonlinearities. The method shows an improvement during sudden load changes, bad data and topology error, however, introduces higher computational burden. In [61], the same approach involving anomalous data mining is proposed. In the next section, alternative approaches used in the state filtering stage of FASE are reviewed.

2. State Filtering Approaches

In presence of drastic load/generation changes, the linear approximation is no longer valid. In fact, the performance of linearized EKF decreases as soon as the prediction error is of the same order with measurement noise. In case of sudden load variations, prediction is no longer accurate and the prediction error becomes important. Hence, effect of higher order terms in linearization cannot be omitted. In [62], two schemes are proposed to incorporate non-linearities into the Kalman filter-based FASE. The first scheme is a modification of the linearized EKF in which local iterations are performed at each time sample. This scheme considerably improves the reference trajectory and the estimate in presence of measurement nonlinearities. In this scheme, Equation (4.5) changes to:

$$\underline{\xi}_{k+1} = \underline{z}_{k+1} - h\left(\underline{\tilde{x}}_{k+1}\right) - H\left(\underline{\tilde{x}}_{k+1}\right)\left(\underline{\hat{x}}_{k+1} - \underline{\tilde{x}}_{k+1}\right) \tag{4.31}$$

In the second scheme, the nonlinear measurement function, $h(x)$ is linearized by using higher order terms in by using 2nd order Taylor expansion. The Taylor expansion of the measurement function is as follows:

$$h\left(\underline{x}\right) = h\left(\underline{\hat{x}}\right) + H\Delta\underline{x} + \underline{S} \tag{4.32}$$

where $\underline{S} = \frac{\partial^2 h(x)}{\partial^2 x}\Delta x^2$.

In this scheme, only 2nd order terms, $\underline{S_i}$, are updated in each local filtering iterations [62-63].And, Equation (4.5) changes to:

$$\underline{\xi}_{k+1} = \underline{z}_{k+1} - h\left(\underline{\tilde{x}}_{k+1}\right) - H\left(\underline{\tilde{x}}_{k+1}\right)\left(\underline{\hat{x}}_{k+1} - \underline{\tilde{x}}_{k+1}\right) - \underline{S} \tag{4.33}$$

As mentioned before, the EKF-based method cannot handle large changes in load and generation. Hence, Shih *et al.* in [50] propose an algorithm where the weight vector, ($R^{-1} = W_k$), is replaced with the $W_k e^{-\left|\underline{z}_k - h(\underline{\tilde{x}}_k)\right|}$. If the load variation is large, the weight associated with that measurement automatically reduces as the term $\left|\underline{z}_k - h(\underline{\tilde{x}}_k)\right|$ become a negative. Under normal conditions, the value of the residual remains insignificant, making exponential term close to one and allowing normal weights for the measurements. In the next section, alternative approaches to implement FASE are reviewed.

3. Alternative Approaches to Implement FASE

In addition to challenge with determining the state transition model, EKF-based FASE suffers from the computationally demanding nature of the filtering in Kalman procedure due to the high dimensions in power networks. The authors in [18, 20] suggest simplifying the filtering process by using a fast, sparse state estimation (SE) algorithm instead of the Kalman procedure. It consists of successive SEs followed by state predictions. With such a scheme, the prediction capability of FASE is still maintained while some heavy computations are avoided. Hierarchical estimation is proposed in [20] to reduce the dimensionality of the problem. This method consists of decomposing the system, performing independent local state estimations and coordinating the latter so as to obtain the required complete and coherent database [20]. This approach is also called two-level estimation technique [64]. First level is called the low level estimation, and the overall network is partitioned into several small sub-systems, each with small number of buses. The second level is called the upper level estimation, and it is only for coordination among the various sub-systems to obtain a single state vector for the entire system. This technique is comparable to single level estimator in accuracy, but is much faster and in most cases the algorithm converges in the first iteration. Successive SE and the hierarchical algorithm are elaborated in the following sections.

a) Succesive SE-based FASE

In successive SE-based FASE, state forecasting stage for an *n*-bus system with *N* (2*n*-1) states is performed as follows [18, 20]:

$$\underline{\tilde{x}}_{k+1} = F_k \hat{x}_k \tag{4.34}$$

The state filtering aims at minimizing the following objective function $J(\underline{x})$ in (2.10). The minimum value of $J(\underline{x})$ can be obtained after the convergence of the following iterative process in the state filtering stage [18, 20]:

$$\underline{\hat{x}}(l+1) = \underline{\hat{x}}(l) + K(l)[\underline{z} - \underline{h}\left(\underline{\hat{x}}(l)\right)] \tag{4.35}$$

$$K(l) = \Sigma(l+1)H^T(l)R^{-1} \tag{4.36}$$

$$\Sigma(l) = [H^T(l)R^{-1}H(l)]^{-1} \tag{4.37}$$

where l=0,1,2,… is the number of iteration until the convergence is attained and $\underline{\hat{x}}(0) = \underline{\tilde{x}}_{k+1}$. H is the Jacobian of h at $\underline{\hat{x}}(l)$ and K is the gain matrix.

b) Hierarchical FASE

In order to apply hierarchical algorithm, the network is decomposed into r non-overlapping areas $A^1, A^2, \ldots, A^r$ connected by tie-lines. The two ends of each tie-line, called boundary nodes, belong to two distinct areas and define the interconnection area. In the i^{th} area, one defines $\underline{x}^i$, the local state vector is composed of the voltage magnitudes and phase angles at the subsystem's nodes. This vector is decomposed into two sub-vectors [20]:

$$\underline{x}^i = [\underline{x}_r^i, \underline{x}_c^i]^T$$

where $\underline{x}_r^i$ relates to the internal nodes and $\underline{x}_c^i$ to the boundary ones. Local references are considered for the phases of each subsystem. Hence, to determine the entire system's state, a vector $\underline{u}$ is defined as the angular difference between subsystems[20]:

$$\underline{u} = [\underline{u}^1, \underline{u}^2, \cdots, \underline{u}^{r-1}]^T$$

where u^i represents the phase angle of the reference node of area A^i with respect to that of A^r taken arbitrarily as the general reference for the whole system. Moreover, the measurements are divided into internal measurements and interconnection measurements. Internal measurements consist of measurements involving local state variables only (voltage magnitudes, active/reactive power flows, and active/reactive power injections) and interconnection measurements consist of power flows on tie-lines. The measurement vector is thus decomposed into the r + 1 sub-vector [20]:

$$\underline{z} = [\underline{z}^{1T}, \underline{z}^{1T}, \cdots, \underline{z}^{rT}, \underline{z}_u^T]^T$$

where $\underline{z}^i$ represent internal measurements within subsystem A^i and $\underline{z}_u$ is the interconnection measurements. It must be pointed out that the hierarchical scheme is only needed for the **filtering step** of the FASE. After the prediction step, the resulting forecasted vector is partitioned to sub-vectors which are ready to be used in coordination level. The filtering step consists of a two-level procedure [20,64]:

1. First (lower) level: local estimation. This level provides local state vector estimates $\underline{\hat{x}}_r$ as well as the corresponding error covariances. As discussed, the interconnection measurements $\underline{z}_u$ are

ignored at this level which results in the loss of optimality. However, in practice, this loss is negligible.

2. Second (upper) level: coordination. In this level, vector $\underline{u}$ is estimated and power flows on tie-lines are evaluated. This estimation is performed using components of $\underline{z}_u$ which are ignored at the first level. Another important objective of this level is to coordinate the various local state vector estimates using voltages at the boundary nodes $\underline{x}_c$. This step is necessary to make $\underline{x}_c$ comply with the additional information contained in $\underline{z}_u$. Therefore, $\underline{x}_c$ is treated as a pseudo-measurement. Finally, having an estimation of $\underline{u}$ and $\underline{x}_c$ the optimal estimate of the system states are obtained using the EKF equations. The hierarchical estimation scheme is shown for a system with three areas in Figure 4.4. This concept can be generalized to *l*-level FASE. More details can be found in [20, 48, and 64].

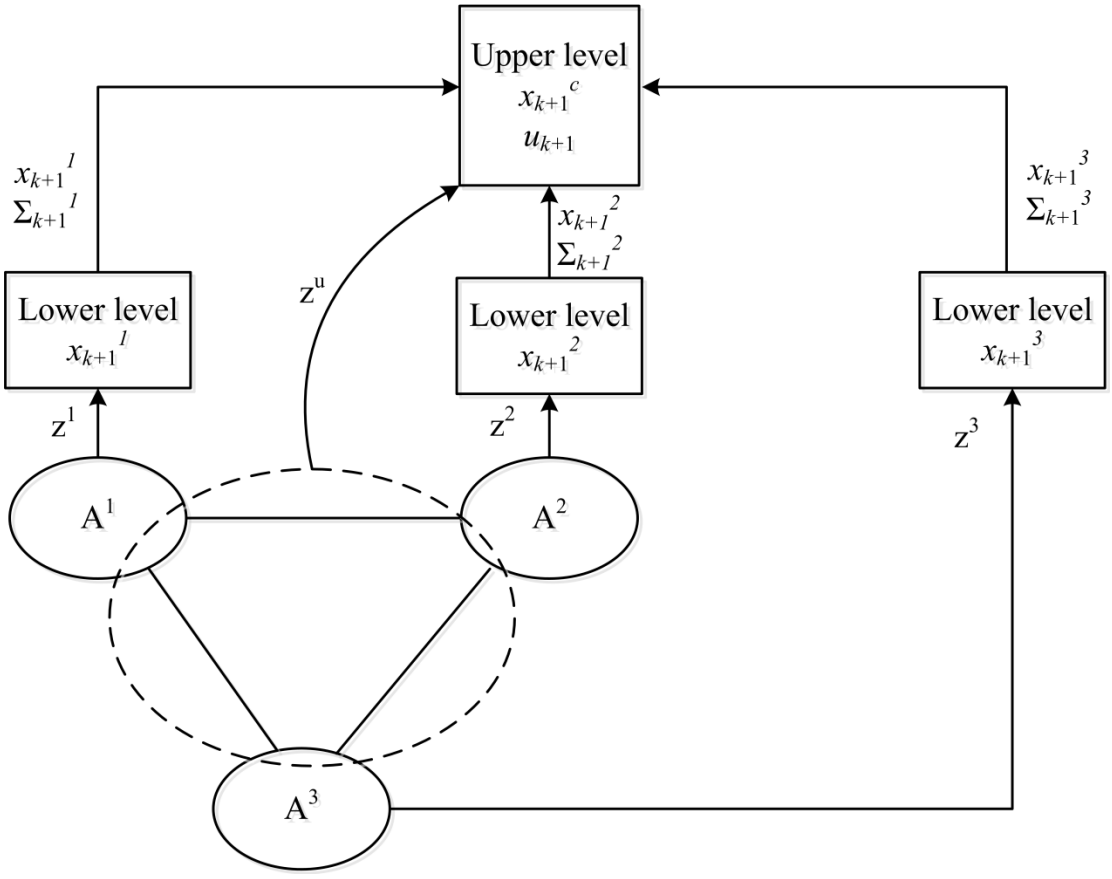

Figure 4.4. Principle of a two-level FASE for a system with three areas [48]

c) Application of Unscented Kalman filter

In order to avoid linearizing the measurement equations application of unscented transformation (UT) combined with the use of Kalman filter, UKF, is proposed in [65]. UKF attempts at approximating the probability distributions of the measurements. The idea of UKF can be summarized as follows [65]:

The mean and covariance of $\underline{\tilde{x}}_{k+1}$ is (4.2) are assumed to be μ_x and Σ_x, respectively. The $2n$+1 points X_i (composed of state vectors) and weights W_i are chosen such that:

$$\underline{\mu}_x = \sum_{i=0}^{i=2n+1} W_i \, X_i \tag{4.38}$$

$$\Sigma_x = \sum_{i=0}^{i=2n+1} W_i \, (X_i - \underline{\mu}_x)(X_i - \underline{\mu}_x)^T \tag{4.39}$$

Then, the resulted points are propagated to produce z_i using the non-linear state function Equation (2.4). For FASE, the UKF algorithm used to estimate the state of a power system $\underline{x}$. The initial state vector $\underline{x}_0$ and the corresponding covariance matrix Σ_0 is considered for the starting point (time sample k=0). The set of $2n$+1 points are calculated to form X_k at time k as follows:

$$\underline{\tilde{x}}_k = \sum_{i=0}^{i=2n+1} W_i \, \hat{X}_k^i \tag{4.40}$$

$$\tilde{\Sigma}_k = \sum_{i=0}^{i=2n+1} W_i \, (X_i - \mu_x)(X_i - \mu_x)^T + Q_k \tag{4.41}$$

where $\underline{\tilde{x}}_k$ is the predicted state mean vector, $\tilde{\Sigma}_k$ is the forecasted state covariance matrix, and $\hat{X}_k^i$ is the ith column of $\hat{X}_k$. Then, the forecasted set it provided for the next time sample as follows:

$$\tilde{X}_{k+1}^i = F\hat{X}_k^i + g_k \tag{4.42}$$

where $\tilde{X}_{k+1}^i$ and $\hat{X}_k^i$ are the ith column of matrices $\tilde{X}_{k+1}$ and $\hat{X}_k$. The points are grouped in matrix $\tilde{X}_{k+1}$ and then propagated through the measurement function (Equation 2.4). The mean and covariance matrices of the measurement are obtained as follows:

$$\underline{z}_k = \sum_{i=0}^{i=2n+1} W_i \hat{Z}_k^i \tag{4.43}$$

$$P_k = \sum_{i=0}^{i=2n+1} W_i \left(\hat{Z}_k^i - \underline{\mu}_z\right)\left(\hat{Z}_k^i - \underline{\mu}_z\right)^T + R_k \tag{4.44}$$

where the cross covariance matrix, Kalman filter gain, the states and the covariance matrix are calculated using the following equations, respectively :

$$C_k = \sum_{i=0}^{i=2n+1} W_i \left(\hat{X}_k^i - \mu_x\right)\left(\hat{Z}_k^i - \mu_z\right)^T + R_k \tag{4.45}$$

$$K_k = C_k {P_k}^{-1} \tag{4.46}$$

$$\underline{x}_k = \underline{x}_k{}^- + K_k(\underline{z}_k - \mu_z) \tag{4.47}$$

$$M_k = M_k{}^- - K_k P_k {K_k}^T \tag{4.48}$$

The procedure above is repeated for every time instant *k*. From the above mathematical description of the UKF, it is observed that it does not require the computation of the Jacobian matrix. Therefore, the implementation of the method becomes easier due to lighter computational load. Since there is no need to calculate any derivate or Jacobian, the UKF is easier to implement for nonlinear systems compared to the EKF. The simplicity of the new estimator and its low computational demand can be mentioned as the advantages of this approach.

In this section, three alternative approached to implement FASE are briefly reviewed. Successive SE, hierarchical methods and UKF are proposed to reduce the computational load of FASE procedure while retaining its prediction capability. In the next section, incorporation of PMU measurements in FASE procedure is reviewed.

4. Incorporating PMU measurements

New measurement technologies such as phasor measurement units (PMUs) provide time-stamped, synchronized complex voltage and current information from various locations throughout the grid. This information enables the system operators to be able to monitor the system with measurements at unprecedented frequencies (unlike the regular measurements, which arrive at every few seconds, the measurements provided by PMUs arrive at control centers every 1/30th of a second.). In addition, PMUs provide the angles of bus voltages with respect to a

reference unlike the conventional measurements, which are only capable of providing the voltage magnitudes. Hui Xue *et al*. [66] propose a technique using the load forecast and combines the PMU data for state prediction. In [66], the state forecasting procedure consists of two steps: (1) load forecast is performed using a short-term nodal load forecasting technique; (2) the predicted state vector is obtained through a decoupled state estimation procedure. For the first step, the nodal power injections are modeled as follows:

$$\underline{\tilde{P}}_{k+1} = \underline{\hat{P}}_k + \underline{\Delta P}_k$$

$$\underline{\tilde{Q}}_{k+1} = \underline{\hat{Q}}_k + \underline{\Delta Q}_k$$

where $\underline{\tilde{P}}_{k+1}$and $\underline{\tilde{Q}}_{k+1}$ are forecasted nodal active/reactive powers at time *k*+1, and $\underline{\hat{P}}_k$ and $\underline{\hat{Q}}_k$ are the instant measurements of powers at time *k*. $\underline{\Delta P}_k$ and $\underline{\Delta Q}_k$ are determined by the following equations [66]:

$$\underline{\Delta P}_k = \frac{1}{n}\sum_{k=1}^{n} \underline{\Delta P}_k$$

$$\underline{\Delta Q}_k = \frac{1}{n}\sum_{k=1}^{n} \underline{\Delta Q}_k$$

where *n* is the number of days used as historical data. In other words, ΔP at time *k* for the desired day is the average of historical data at time *k*. The measurement vector $\underline{z}$ is formed as follows [66]:

$$\underline{z} = \begin{bmatrix} \underline{\tilde{P}_i} \\ \underline{\tilde{P}_{ij}} \\ \underline{\theta_i} \\ \underline{P} \\ \underline{\tilde{Q}_i} \\ \underline{\tilde{Q}_{ij}} \\ \underline{V_i} \\ \underline{Q} \end{bmatrix}$$

where, $\underline{\tilde{P}_i}$, $\underline{\tilde{P}_{ij}}$, $\underline{\tilde{Q}_i}$ and $\underline{\tilde{Q}_{ij}}$ are the forecasted active/reactive powers, $\underline{\theta_i}$ and $\underline{V_i}$ are the angle and voltage from PMU and neighboring bus, $\underline{P}$ and $\underline{Q}$ are pseudo-measurements incident to PMU buses. The simulation results show that the PMU based FASE reduces the estimation errors in comparison to the case of FASE without PMUs. However, the strength of PMUs are not

exploited effectively since the authors do not use every measurement provided by PMUs every $1/30^{th}$ of a second. In the following section, a summary is of this chapter is provided.

D. Summary

In this chapter, the fundamentals of forecasting-aided state estimation are reviewed. A comprehensive literature review about the three stages (i.e. model identification, forecasting, and filtering) of FASE is provided. In addition, alternative methods for reducing the computational burden of FASE are briefly reviewed. The available transition models for the forecasting stage of FASE suffer from three main drawbacks:

1. Time correlation is not addressed and exploited effectively. The impact of high percentage integration of intermittent generation on the system states can be exploited to capture the short-term behavior of states. The proper exploitation of the increased randomness in system states by using statistical models can help improving state transition model under certain assumptions.
2. The power system states are assumed to vary independently from each other, which leads to a diagonal state transition matrix. This assumption is not realistic in power systems in the presence of microgrids, distributed and intermittent power generation, which sometimes are connected to the grid through radial buses. The grid of the future is becoming more decentralized resulting in the decrease of number of buses with large amount of power injections (so called 'electrical hubs'.) Therefore, these electrical hubs lose their significance and dominance in the grid, and the spatial correlation among the states throughout the grid become more observable. Spatial correlation defines the level of interaction of a state with the states of the neighboring buses. Taking spatial correlation into account enables the operator to restore observability when a state is unobservable due to loss of telemetry.
3. The existing state transition models are not effectively utilizing this valuable information provided by the PMUs, which provide time-stamped, synchronized complex voltage and current information from various locations throughout the grid.

Next chapter is dedicated to identification of temporal and spatial correlation in power networks. Three metrics are reviewed since the proposed state transition model utilizes these metrics to identify the structure of the state transition model.

CHAPTER V

TEMPORAL AND SPATIAL CORRELATION OF POWER SYSTEM STATES IN STEADY-STATE OPERATION

A. Introduction

This chapter presents methods to analyze the temporal and spatial correlation in and among the power system states (i.e. complex bus voltages.) The results of these analyses are used to identify the structure of the state transition model described in Chapter VII. Each method enables the operator to analyze a large network by using simple metrics. These metrics help the operator to identify 'important' nodes as well as the impact of these nodes on the rest of the network. The metrics are:

- Electrical connectivity
- Electrical centrality
- Electrical node significance

Electrical connectivity is used to determine the buses with strong/weak electrical connections to the rest of network. This metric, along with *electrical centrality,* is used to determine the relative degree of correlation in time within the historical terms of each bus voltage. The history terms of bus voltages with relatively small electrical centrality and/or small electrical connectivity (such as radial buses) are not as correlated in time as the history terms of voltages at buses with relatively large electrical connectivity and/or relatively high centrality. One example to this phenomenon is the electric hubs and their adjacent buses. Identifying electrical hubs using electrical connectivity helps to determine relative degree of correlation in space (i.e. spatial correlation) among the time series of the bus voltages. Note that, due to the strong connection of electrical hubs to the rest of the grid, amount of power flowing through these buses is significant. Therefore these buses and their adjacent buses are dominant in the variation of the states in the rest of the network. In other words, the states of electrical hubs and their neighboring states vary like each other as a 'super-node' during load and generation variations. Hence, the quantity of electrical hubs in a given network has an adverse effect on the relative degree of spatial correlation among history terms of bus voltages in the rest of the network. This phenomenon is further exploited by using another metric; *electrical node significance*.

Electrical node significance is used to determine the distribution and the quantity of important (significant) nodes in a given network by ranking the nodes based on the amount of injected power. This metric along with electrical centrality is used to obtain the relative degree of spatial correlation among the time series of bus voltages. For example, if several buses with relatively large and similar degrees of electrical node significance are spread throughout a network, the degree of spatial correlation in that network is less than a network with a few significant nodes which are physically close to each other, with different degrees of significance. Following sections provide the necessary theoretical background and demonstrate the uses of these metrics through examples.

B. Electrical Connectivity

In a power grid, connections of components are not only dependent on physical topology but also on electrical properties. The connectivity of the components can be assessed using sensitivity matrices. For power systems, sensitivity matrices such as power transfer distribution factor matrices [67] is used to determine the amount of influence of one component on another. Distance matrix, which is denoted by the impedance matrix, Z_{bus} is used as the complement of a sensitivity matrix. The distance matrix is calculated by inverting admittance matrix Y_{bus}. Y_{bus} is a sparse matrix and its inversion, for a fully connected electrical power network, results in a dense matrix. The equivalent electrical distance between two nodes is represented by the magnitude of the relevant entry of the Z_{bus} which is relatively small for component pairs that have high electrically proximity, and a large number for component pairs that have very little influence on one another [68]. Based on the discussions in [68], the distance matrix indicates that power grids possess some nodes that have strong electrical proximity with several other nodes. These nodes are called ‘electrical hubs’ and are connected by two to five transmission lines to the grid; however, they have small electrical distance (i.e. strong electrical connectivity) to at least twenty buses. Hence, the presence of electrical hubs cannot be identified by using only physical topology and it is necessary to exploit the properties of Z_{bus}. Since this matrix is dense, the number of possible electrical connections is $\frac{n(n-1)}{2}$, where n is number of buses in a network. For example, there are 6,903 possible electrical connections in IEEE 118-bus network. In order to have a better insight about the electrical connectivity, the electrical connections of 118-bus system are sorted in ascending order. The first 179 connections, which are the strongest 179

electrical connections, are shown in Figure 5.1. Note that, 118-bus has 179 branches and the first 179 electrical connections are selected to provide a comparison between physical and electrical topology. Physical topology of 118-bus [69] network is shown in Figure 5.2. In Figure 5.1 the size and color of the nodes indicate their relative importance in the network, with electrical connectivity decreasing as the size decreases. The largest green nodes represent the nodes with highest connectivity (it has the highest number of electrical connections compared to the rest), followed by the cyan and the purple nodes. The small red nodes are the nodes with very small number of electrical connections. In Figure 5.3, buses are sorted in descending order based on their number of electrical connections and the first 70 buses are given for 118- and 300-bus systems due to space limitations. The number of connections for each system is considered to be equal to the number of branches in that system. Therefore, 78, 179 and 411 connections are considered for IEEE 57-, 118- and 300-bus networks, respectively. In order to compare the electrical connectivity and the number of electrical hubs of different networks; a threshold is set as $\kappa_{hub} = 20$. In Figure 5.1, it is observed that there are seven buses with more than κ_{hub} in 300-bus network constituting the electrical hubs.

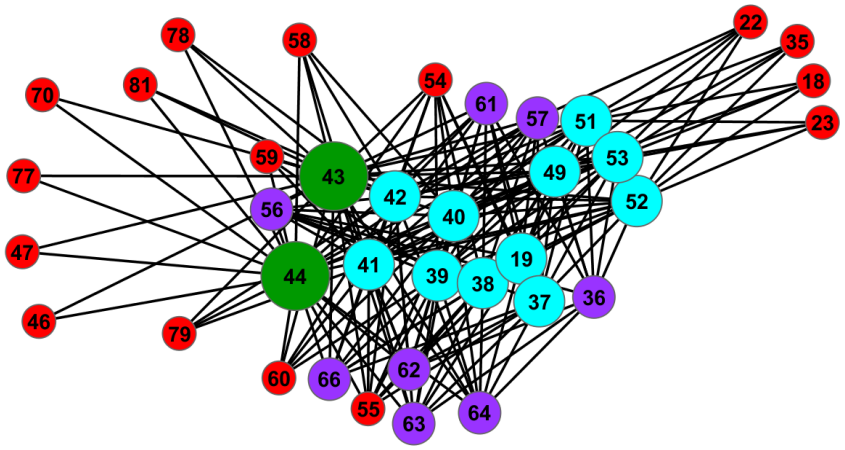

Figure 5.1. Electrical connectivity of the IEEE 118-bus network. The large nodes in green are the most electrically central nodes, followed by the slightly smaller nodes in cyan, and then in purple [70].
Used with permission per email from S. Pahwa April 2014, attached.

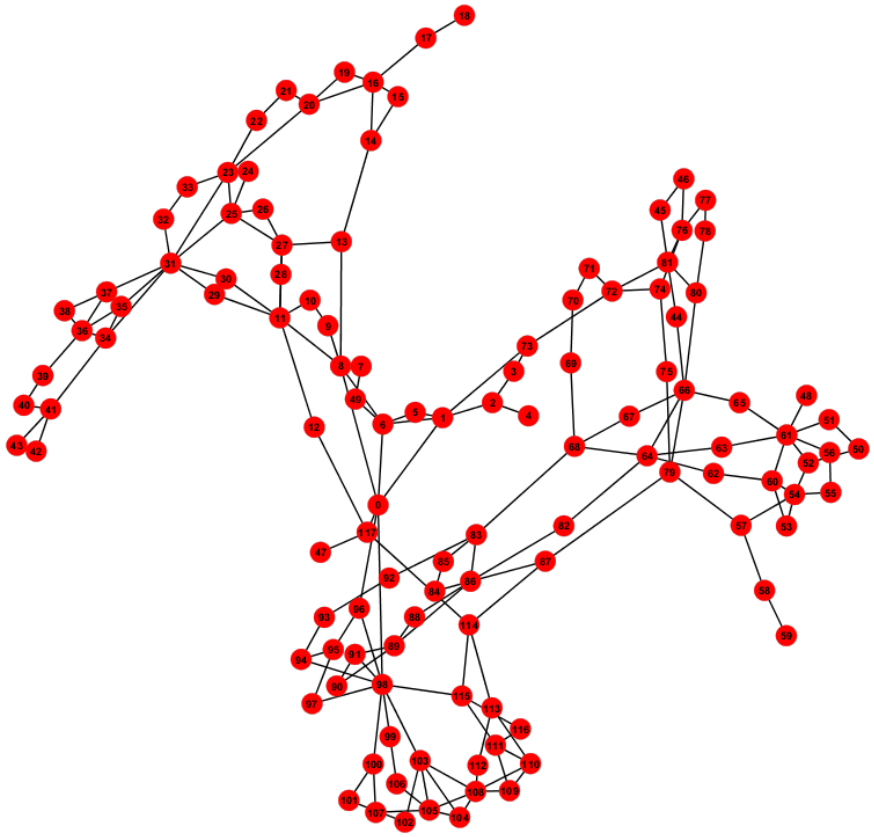

Figure 5.2. Physical structure of the 118-node network [70]
Used with permission per email from S. Pahwa April 2014, attached

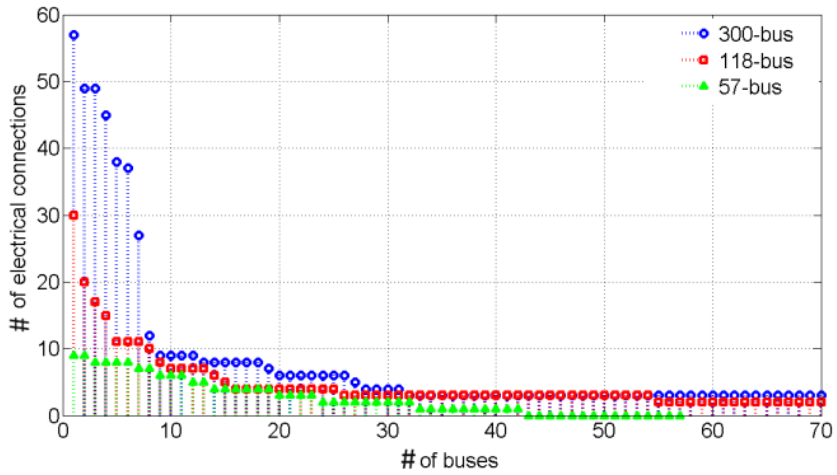

Figure 5.3. The sorted buses with strongest connections for IEEE 57-, 118- and 300-bus networks.
Only the first 70 buses are shown for IEEE 118- and 300-bus networks.

However, in 118- and 57-bus networks the number of hubs is reduced to one and zero, respectively due to the chosen threshold, κ_{hub}. It can be concluded that the degree of spatial correlation among the states of 57-bus is more than two other networks while the degree of time

correlation among the time series of bus voltages in 118- and 300-bus is more than that of 57-bus network. According to our analysis, the time correlation among the states of the networks with several numbers of hubs is more prevalent than space correlation.

C. Electrical Centrality

The distance matrix is used to determine another metric in network which is called 'electrical centrality' of each bus. Electrical centrality provides another way to express which nodes are more electrically central to the system. The authors in [71] extended the idea of 'resistance distance' to represent the marginal effect of an active power transaction between buses *i* and *j* on voltage phase angles differences. A network of resistors with current injections at each node can be described by a conductance matrix *G*, such that the current injection at node *i* is as follows:

$$I_i = \sum_{j=1}^{n} g_{ij} V_j \tag{5.1}$$

In presence of no ground connection, *G* has rank *n*−1. By considering a reference voltage, *G* defined as a Laplacian matrix for the resistive network. Node *r* is specified as a voltage reference node, with V_r=0 to make *G* non-singular. The sub-matrix of *G* associated with the non-reference nodes (nr) is full-rank; therefore, we can compute a sub-matrix, G_{nr}^{-1} such that,

$$V = G_{nr}^{-1} I \tag{5.2}$$

If the diagonal element of G_{nr}^{-1} associated with node *i* is ${g_{i,i}}^{-1}$, it indicates the change in voltage between nodes *i* and *r* as a result of a current injection at node *i*, which must be withdrawn from node *r*. This resistance distance gives the sensitivity between current injections and voltage differences. Therefore, the resistance distance between *i* and *r* is equal to ${g_{i,i}}^{-1}$. In order to measure the distance between a pair of nodes *i* and *j*, where $i \neq j \neq r$, one can evaluate:

$$d_{ij} = {g_{i,i}}^{-1} + {g_{j,j}}^{-1} - {g_{i,j}}^{-1} - {g_{j,i}}^{-1} \tag{5.3}$$

which gives the voltage difference between *i* and *j* after injecting 1 A at *i* and withdrawing 1 A from *j*. Using this metric, a measure can be defined that is roughly similar to node degree, but for a fully connected network with continuous weights for each node pair. This measure is determined by taking the average distance from each node *i* to other nodes in the system [68]:

$$ad_i = \sum_{j=1}^{n} \frac{d_{ij}}{n-1} \tag{5.4}$$

Inverting it leads to the metric of *electrical centrality* [68].

$$c_i = \frac{1}{ad_i} \tag{5.5}$$

The electrical centralities of buses in 118-bus network are shown in Figure 5.4. In this dissertation, imaginary part of Y_{bus} is used instead of G matrix in (5.2) to determine the electrical centralities. Buses with relatively small electrical centrality (e.g. radial buses) have weak electrical connectivity. Therefore, voltages at these buses are not as correlated in time as the voltages at those buses with strong electrical connectivity. On the other hand, buses with relatively large electrical centrality have strong electrical connectivity to the rest of buses (e.g. electrical hubs). And, their time series are highly correlated in time.

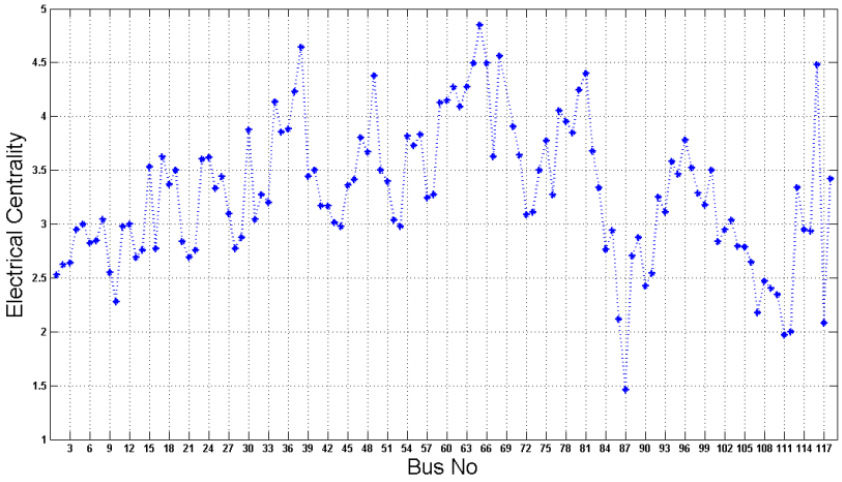

Figure 5.4. Electrical centrality of buses of IEEE 118-bus

The electrical centralities of the IEEE testbeds (i.e. IEEE 57-, 118- and 300-bus) are determined and statistical properties of each are represented in Table 5.1.

Table 5.1. Statistical Analysis of Electrical Centrality of IEEE Test Beds

IEEE Testbeds	Mean	Median	Max	Min
57	2.2682	2.4060	3.2346	0.7369
118	3.2841	3.2699	4.8520	1.4640
300	3.4816	3.1848	9.6467	0.1339

It is observed that 57-bus has the lowest mean, median and maximum electrical centrality. In other words, the electrical connectivity among the buses of 57-bus system is less than those in the other networks which leads to weaker time correlation among the bus voltages than those in the other two networks. 118-bus network has the largest minimum electrical centrality compared to the other two networks. This is due to the lack of buses with strong electrical connectivity. In

the meantime 300-bus network has the smallest minimum and largest maximum electrical centralities. The large value of electrical centrality indicates the presence of electrical hubs as observed in Figure 5.3. And, small electrical centralities indicate the presence of radial buses with a very small degree of physical connectivity to the rest of the network

D. Electrical Node Significance

The concept of electrical node significance is used to rank the nodes with respect to significance (dominance), depending on the quantity of power injected by them to the rest of the network. This metric, combined with other metrics, is used in [72] to quantify the robustness of power networks against cascading failures by targeted attacks. [73] uses this metric along with the electrical connectivity to determine the best location for distributed generation. Electrical node significance is defined as follows [70]:

$$es_i = \frac{P_i}{\sum_{j=1}^{n} P_j} \tag{5.6}$$

where P_i is the injected power by node i and n is the number of buses in the network. The significant nodes for 118-bus system are shown in Figure 5.5. The relative importance of the nodes is demonstrated by their size and color. The biggest node in green is the most significant node, followed by the cyan node, purple nodes, and red nodes. All the small yellow nodes have small electrical significance. Note that significant nodes are spread throughout the network. In order to compare the number of electrically significant nodes of different networks; a threshold is set as $\kappa_{e.n.s.} = 1$. It is observed that there are 14 significant nodes in 118-bus network. In other words, injected power from 12% of the buses play a significant role in the network. The absolute values of electrical node significances of 57-, 118- and 300-bus networks are sorted in descending order in order to have a better overview. The first seventy buses (buses with the biggest node significance), for 118- and 300-bus systems, are plotted in Figure 5.6. It is observed that the pattern and number of significant nodes for 118- and 300-bus systems are more similar to each other than those of 57-bus network. The difference between the first and second significant nodes in 57-bus is around 4. However, this difference is 1 in both 118- and 300-bus networks. Note that there is a significant difference among the quantity of buses with electrical node significance greater than $\kappa_{e.n.s.}$ in 118- and 300-bus networks than that of 57-bus. It can be concluded that there are only a few significant nodes with different degree of significances in 57-

bus network. However, in both 118- and 300-bus networks, there are several significant nodes that are spread throughout the network. It is expected that the spatial correlation among the time series of bus voltages in 57-bus system is more observable than it is in two other networks. In addition, the time correlation among time series of bus voltages in 118- and 300-bus systems are more observable than it is in 57-bus network. It is worthwhile to note that the integration of distributed intermittent generation and microgrids has considerable impact on the electrical node significance of a network and on the spatial correlation. This integration changes the distribution pattern of load and generation throughout a network and therefore nodes significance will drastically change. Microgrids and distributed generations are connected to the grid sometimes through radial buses. This leads to a less centralized grid resulting in the decrease of number of buses with large amount of power injections. Furthermore, those buses lose their significance and dominance in the grid and the spatial correlation among the states throughout the grid become more observable. Table 5.2 summarizes the results of the discussions.

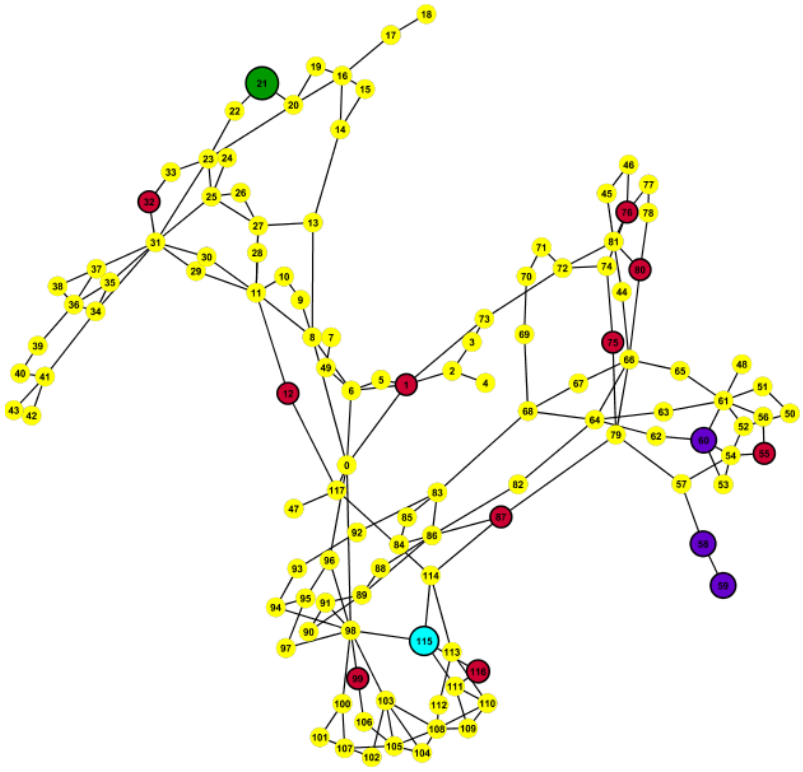

Figure 5.5. Electrical significance of nodes for the 118-node network.
The relative importance of the nodes is shown by their size and color [70]
Used with permission per email from S. Pahwa April 2014, attached

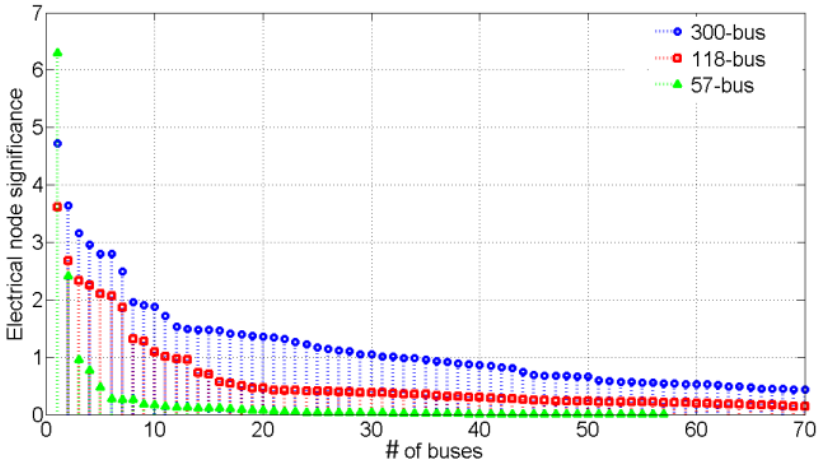

Figure 5.6. Sorted absolute value of electrical node significance of IEEE 57-, 118-, and 300-bus systems for the first 70 buses.

Table 5.2. Summary of Interpretation of the Three Metrics for Identification of Time/Space Correlation

Metric	Correlation
Strong*/weak electrical connectivity*	**High**/low time correlation
Large*/small electrical centrality*	**High**/low time correlation
*Several/**fewer** electrical hubs*	Low/**high** spatial correlation
*Concentrated/**distributed** node significance*	Low/**high** spatial correlation

E. Summary

This chapter reviews three metrics to identify and assess the temporal and spatial correlation between system states (i.e. bus voltages) in power networks. Electrical centrality, electrical network connectivity and electrical node significance are used in this dissertation to determine the relative degree of the temporal and spatial correlation between bus voltages. The information obtained from these metrics is used to determine the structure of the proposed state transition model for FASE using time series analysis techniques. In the next chapter, a brief review of time series analysis fundamentals is provided.

CHAPTER VI

FUNDAMENTALS OF TIME SERIES ANALYSIS

A. Introduction

Time series analysis aims at finding statistical features of data points taken over time. Time series is a successive order of values of a variable at uniform time intervals. The time series models can be used for the followings applications [74]:

- Determine the underlying structure of the observed data,
- Fit a model and provide basis for forecasting, monitoring or even feedback and feedforward control.

There are many methods for model fitting such as exponential smoothing and unvariate/vector autoregressive methods. The appropriate technique is determined based on the application and user preferences. It is beyond the realm and intention of this chapter to cover all these methods. In this chapter, a brief review of common approaches to model univariate and vector time-series is presented.

B. Univariate Time Series Models

Univariate time series consist of single (scalar) observations recorded sequentially over equal time intervals. Some fundamental terms are defined as follows [74]:

Stationarity

In many time-series techniques, it is assumed that the data are stationary. For a stationary process, the mean, variance and autocorrelation structure is constant over timeframe of interest. In addition, stationary time-series does not contain periodic fluctuations and a trend.

Seasonality

Seasonality can be observed in many time-series, especially data in economics, and it is interpreted as the presence of periodic fluctuations. For example, retail sales tend to peak in Christmas season and then decline after Christmas season. In the followings, a few of the most common approaches to model univariate time series are described.

1. Moving Average and Smoothing Techniques

It is known that some form of random variation is available in any kind of collected data and there are methods that can be used for eliminating or reducing the effect of random variation. A popular approach is 'smoothing'. Smoothing methods can be categorized into two main groups:

- Averaging
- Exponential smoothing

a) Averaging: The averaging method which takes the mean of all past observations can be used for forecasting if there are no trends. This method weight all past data equally. Alternatively, average of successive smaller sets can be calculated. This process is continued by dropping the first data, advancing one period and calculating the average of the considered set of data. The general expression for averaging is as follows [74]:

$$m_t = \frac{x_t + x_{t-1} + \cdots + x_{t-N+1}}{N} \tag{6.1}$$

where *N* is the number of considered data points. As mentioned, the performance of averaging methods is degraded in presence of trends. Therefore, exponential smoothing techniques are used to take the trends into account. In the next section, a brief review of exponential smoothing is provided.

b) Exponential Smoothing: Exponential smoothing does not assign equal weight to all the observations. In this method, the weights exponentially decreasing as the observations getting older. Therefore, more weights are assigined to the recenet observation. Exponential smoothing is first suggested by C.C. Holt in 1957 and is proposed for non-seasonal time-series with no trend [52]. He later proposed a procedure that does handle trends [52]. Winters generalizes the method to include seasonality, and the technique is called 'Holt-Winters Method'. There are mainly three types of exponential smoothing schemes: single, double an triple. In the followings, the first two are breifly reviewed.

Single Exponential Smoothing scheme can be initialized by setting s_2 to y_1, where s_i is the smoothed observation at time *i* and y_t is the observation at time *t*. The average of the first four or five observations can also be used as the initial value. The smoothed value at time *t* is determined using the following equation [52]:

$$s_t = \alpha y_{t-1} + (1-\alpha)s_{t-1} \quad 0 < \alpha \leq 1 \;\; t \geq 1 \tag{6.2}$$

where α is called the smoothing constant. Equation (6.2) is expanded by substituting for s_{t-1} in as follows:

$$s_t = \alpha y_{t-1} + (1-\alpha)[\alpha y_{t-2} + (1-\alpha)s_{t-2}]$$

$$= \alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + (1-\alpha)^2 s_{t-2}$$

If we continue to substitute for s_{t-2}, s_{t-3}, and so forth, we reach to s_2 (which is y_1 or the initial value). The expanded equation can be written as:

$$s_t = \alpha \sum_{i=1}^{t-2} (1-\alpha)^{i-1} y_{t-i} + (1-\alpha)^{t-2} s_2, t \geq 2 \tag{6.3}$$

It is observed that (6.3) illustrates an exponential behavior because of the geometric decrement of weights, $\alpha(1-\alpha)^t$, for the older observations. Therefore, older observations have less effect on the smoothed value compared to the recent ones. The speed of decrement of the effect of older observations on the smoothed value is directly related to α. Usually, the best value for α (i.e. the value which results in the smallest mean square error) depends on the application. Based on (6.2), forecasting formula using single exponential smoothing is as follows [74]:

$$s_{t+1} = s_t + \alpha\varepsilon_t \tag{6.4}$$

where ε_t is error term. According to (6.4), the new forecast is the old one plus an adjustment of the error of the last forecast. As discussed, the performance of single smoothing methods is not good in presence of trend. In the next section, double exponential smoothing; which can successfully handle the trends is reviewed.

Double exponential smoothing can perform better than single smoothing methods in presence of trends due to having a second equation. The two equations associated with double exponential smoothing are as follows [52]:

$$s_t = \alpha y_{t-1} + (1-\alpha)(s_{t-1} + b_{t-1}) \;\; 0 \leq \alpha \leq 1 \tag{6.5}$$

$$b_t = \gamma(s_t - s_{t-1}) + (1-\gamma)b_{t-1} \qquad 0 \leq \gamma \leq 1 \tag{6.6}$$

Initial values for s_t and b_t can be assigned in several ways. Usually, s_1 is set to y_1 and b_t is chosen using one of the following equations [46].

- $b_1 = y_2 - y_1$
- $b_1 = \frac{1}{3}[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]$
- $b_1 = \frac{y_n - y_1}{n-1}$

In Equation (6.5), the trend of the previous period, b_{t-1}, is considered for s_t by adding it to s_{t-1} term. Equation (6.6) is used to updates the trend. In this equation, the difference between the last two smoothed values are determined and added to the previous calculated trend. The forecasted value using double exponential smoothing for the next *m* period ahead is given by [46]:

$$f_{t+m} = s_t + mb_t \tag{6.7}$$

In order to have a comparison, plots of forecasting results using single and double smoothing methods are depicted in the Figure 6.1.

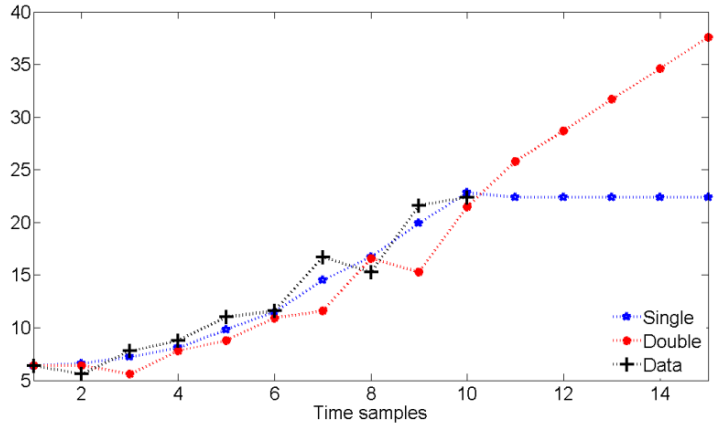

Figure 6.1. Demonstration of forecasting results using single and double smoothing

It is observed that double smoothing method follows the data better than single smoothing. Single smoothing provided a straight horizontal line for the forecasted values, which cannot be true according to the observed data. In the next section, autoregressive models are briefly reviewed.

2. Autoregressive (AR) Models

Autoregressive (AR) models are one of the common approaches for modeling univariate time series [75]:

$$y_t = \delta + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t \tag{6.8}$$

where e_t is a white noise and it is uncorrelated with y_t and t=0,1,2,…, n. φ_1, φ_2 ,…, φ_p are deterministic parameters. And,

$$\delta = \left(1 - \sum_{i=1}^{p} \varphi_i\right)\mu \tag{6.9}$$

where μ is the process mean. According to (6.8), an autoregressive model is basically a linear regression of y_t against p previous terms. p is called the order of the AR model. In the followings, moving average (MA) model is briefly reviewed.

3. Moving Average (MA) Models

The moving average (MA) model is represented as [75]:

$$y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \tag{6.10}$$

where μ is the process mean. e_{t-q} are white noise terms and θ_1, θ_2 ,…, θ_p are the parameters of the model. According to (6.10), moving average model can be interpreted as linear regression of y_t against the white noise or random shocks of q previous terms. The random shocks are assumed to have Gaussian (Normal) distribution with zero mean. In the next section, Box-Jenkins approach, which combines AR and MA model, is described. Also, the procedure to fit a proper model for the desired observations is reviewed.

4. Box-Jenkins Approach

Box and Jenkins proposed a systematic methodology to combine both AR and MA models [74-75]. The proposed model is called ARMA and is expressed as follows:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \tag{6.11}$$

where the terms in the equation have the same meaning as given for the AR and MA models. The followings should be considered for using this model:

- The time-series is assumed to be stationary. If the considered series in not stationary, some technique such as differencing should be applied to achieve stationarity.
- The time-series is assumed to be zero mean. Therefore, the mean of the series should be subtracted from each data point.

The general form of Box-Jenkins model includes AR, MA, difference operators, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms. Interested readers can refer to [75-77] for further details. There are three primary stages in building a Box-Jenkins time series model.

- Model identification
- Model estimation
- Model validation

Effective fitting of ARMA models requires a moderately long series. In [76], it is recommended to consider at least 50 observations. In the following section, model identification procedure is described.

As mentioned, the first step in developing a Box-Jenkins model is to determine stationarity and seasonality of the series. Once stationarity and seasonality is determined, the next step is to identify the order of the autoregressive and moving average terms, p and q in (6.11). Autocorrelation and partial autocorrelation plots are main tools in this stage. In the following, calculating *sample autocorrelation function* (ACF) is reviewed.

Let $y_0, y_1, y_2, \dots, y_N$ be observations of a time series. The sample-mean of $y_0, y_1, y_2, \dots, y_N$ is [78]:

$$\mu = \frac{1}{N}\sum_{k=1}^{N} y_k \tag{6.12}$$

The sample autocovariance function is estimated by [78]:

$$\hat{\gamma}(l) = \frac{1}{N}\sum_{k=1}^{N-l}(y_{k+l} - \mu)(y_k - \mu)\ ,\ -N < l < N \tag{6.13}$$

where l is the time lag. The sample autocorrelation function is:

$$\hat{\rho}(l) = \frac{\hat{\gamma}(l)}{\hat{\gamma}(0)}\ ,\ -N < l < N \tag{6.14}$$

where $\hat{\rho}(l)$ is between -1 and 1. In order to have autocorrelation plot, $\hat{\gamma}(l)$ is plotted vs. time lag (l=0,1,2,3,…). Confidence bands are also plotted along with the autocorrelation plot. The following equation is used to calculate the confidence bands [74]:

$$\pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{N}(1+2\sum_{i=1}^{l} y_i^2)} \tag{6.15}$$

where l is the lag, N is the sample size, z is the cumulative distribution function of the standard normal distribution and α is the significance level. Based on (6.15), the confidence bands increase as lag increases. The following questions can be answered by autocorrelation plot [74]:

- Is the observed time series random (e.g. white noise)?
- Is the observed time series autocorrelated (i.e. adjacent observation related to each other)?
- Is the observed time series seasonal (periodic)?
- Is the observed time series autoregressive?
- What is an appropriate model for the observed time series?

Examples of the autocorrelation plot for several common situations are given in the followings. The autocorrelation plot for a random time series is depicted in Figure 6.2.

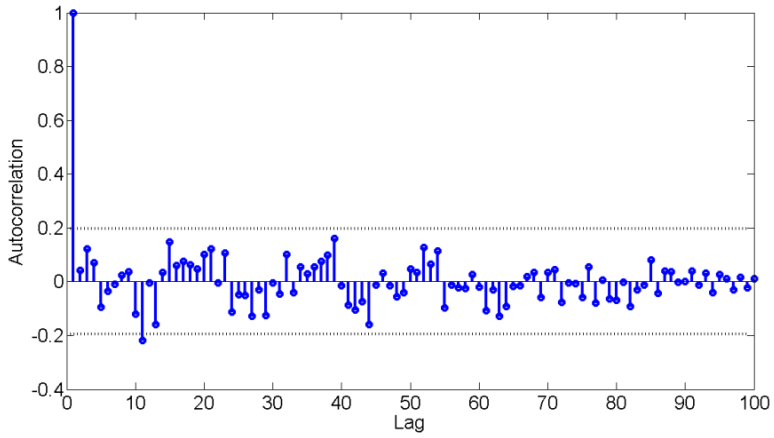

Figure 6.2. Autocorrelation plot for a random time-series

It is observed that disregarding lag 0, which is always 1 by definition, almost all of the autocorrelations fall within the confidence bands. For a non-random series, at least 5% of the lags should be significant. In addition, no apparent pattern (such as the first twenty being positive and the second twenty being negative) is observed in the plot, which is expected for random data. Also, no associations between consecutive values can be interpreted. In other words, there is no correlation between adjacent observations, 'autocorrelation', in this case. In Figure 6.3, autocorrelation plot for a time-series with moderate autocorrelation is depicted.

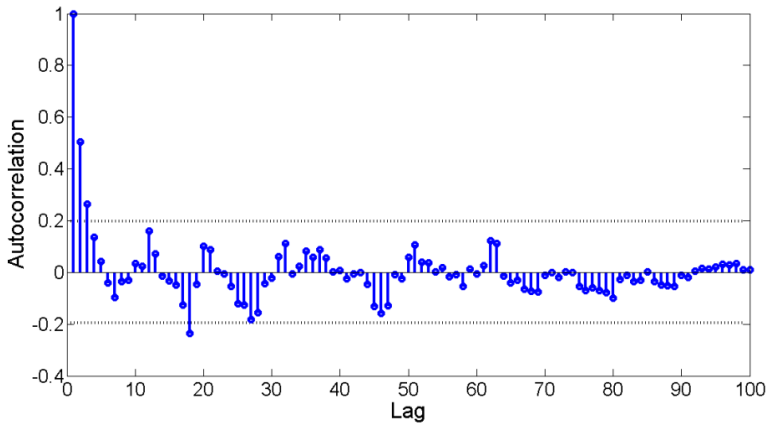

Figure 6.3. Autocorrelation plot for a moderately autocorrelated time-series

It is observed that the autocorrelation at lag 1 is high and decreases linearly at the other lags. Therefore, moderate predictability can be provided using proper AR-based models for this type of series.

In Figure 6.4, autocorrelation plot for a time-series with strong autocorrelation is depicted. It is observed that the autocorrelation at lag 1 is high and decreases slowly at the other lags. However, it becomes negative and starts showing an increasing negative autocorrelation. The variation is usually linear with little noise. In Figure 6.5, the autocorrelation plot for a sinusoidal data is depicted.

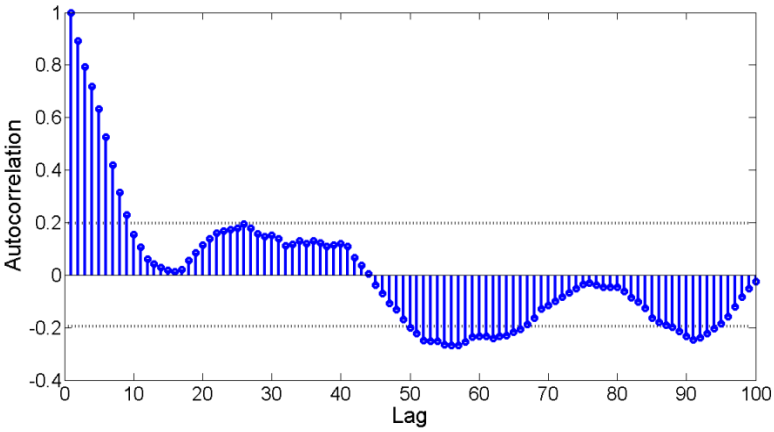

Figure 6.4. Autocorrelation plot for a strongly autocorrelated time-series

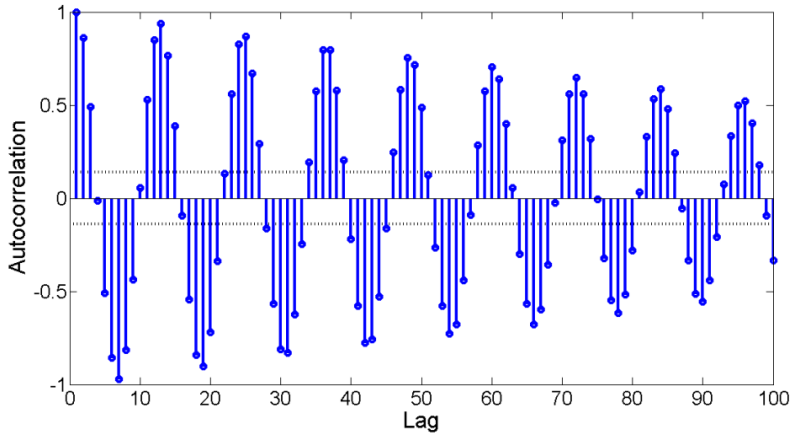

Figure 6.5. Autocorrelation plot for a sinusoidal time-series

An alternating sequence of positive and negative spikes is observed in Figure 6.5 which is not decaying to zero. Table 6.1 summarizes indicted models based on autocorrelation plot [74].

Table 6.1. Model Identification using Autocorrelation Plot [74]

Plot Shape	Indicated model
Exponentially decaying to zero.	Autoregressive
Alternating positive and negative, decaying to zero	Autoregressive
One or a few spikes, rest are close to zero	Moving average
After a few lags, it starts decaying	Mixed autoregressive and moving average
All zero or close to zero	No model can be fitted. Data is random.
High values at fixed intervals	Autoregressive with seasonal terms.
No decay to zero	Series is not stationary. Make it stationary using proper techniques or apply approaches such as co-integration .

Partial auto correlation plot, which is used along with the autocorrelation plot to identify the order of ARMA models, is described below. The partial autocorrelation at lag k is the autocorrelation between y_t and y_{t-k} that is not accounted for by lags 1 through $k-1$. The *sample partial autocorrelation function* (PACF) is defined as follows [78]:

$$\hat{\alpha}(0) = 1$$

And

$$\hat{\alpha}(l) = \hat{\phi}_{ll} \quad l \geq 1$$

where $\hat{\phi}_{ll}$ is the last component of $\hat{\phi}_l = \hat{\Gamma}_l^{-1}\underline{\hat{\gamma}}_l$.

$\hat{\Gamma}_l = [\gamma(i-j)]_{i,j=1}^{l}$ *and* $\underline{\hat{\gamma}}_l = [\gamma(1) \quad \gamma(2) \quad \ldots \quad \gamma(l)]$. $\gamma(l)$ is the autocovariance at lag l and estimated by:

$$\hat{\gamma}(l) = \frac{1}{N}\sum_{k=1}^{N-|l|} y_k\, y_{k+l}$$

where y_k is the observation at time *k*. For example, for *l*=0, 1, 2, and 3 we have:

l=0 , $\hat{\alpha}(0) = 1$

l=1, $\hat{\alpha}(1) = \hat{\phi}_{11}$

$$\hat{\Gamma}_1 = [\gamma(1-1)]_{i,j=1}^{1} = \gamma(0)$$

$$\underline{\hat{\gamma}}_l = \gamma(l)$$

$$\hat{\phi}_1 = \frac{\gamma(l)}{\gamma(0)}$$

l=2, $\hat{\alpha}(2) = \hat{\phi}_{22}$

$$\hat{\Gamma}_2 = \begin{bmatrix}\gamma(1-1) & \gamma(1-2)\\ \gamma(2-1) & \gamma(2-2)\end{bmatrix}_{i,j=1}^{2} = \begin{bmatrix}\gamma(0) & \gamma(-1)\\ \gamma(1) & \gamma(0)\end{bmatrix}$$

$$\underline{\hat{\gamma}}_l = [\gamma(1) \quad \gamma(2)], \hat{\phi}_l = \begin{bmatrix}\gamma(0) & \gamma(-1)\\ \gamma(1) & \gamma(0)\end{bmatrix}^{-1} [\gamma(1) \quad \gamma(2)]^T$$

l=3, $\hat{\alpha}(3) = \hat{\phi}_{33}$

$$\hat{\Gamma}_3 = \begin{bmatrix}\gamma(1-1) & \gamma(1-2) & \gamma(1-3)\\ \gamma(2-1) & \gamma(2-2) & \gamma(2-3)\\ \gamma(3-1) & \gamma(3-2) & \gamma(3-3)\end{bmatrix}_{i,j=1}^{3} = \begin{bmatrix}\gamma(0) & \gamma(-1) & \gamma(-2)\\ \gamma(1) & \gamma(0) & \gamma(-1)\\ \gamma(2) & \gamma(1) & \gamma(0)\end{bmatrix}$$

$$\underline{\hat{\gamma}}_l = [\gamma(1) \quad \gamma(2) \quad \gamma(3)]$$

$$\hat{\phi}_l = \begin{bmatrix}\gamma(0) & \gamma(-1) & \gamma(-2)\\ \gamma(1) & \gamma(0) & \gamma(-1)\\ \gamma(2) & \gamma(1) & \gamma(0)\end{bmatrix}^{-1} [\gamma(1) \quad \gamma(2) \quad \gamma(3)]^T$$

A sample autocorrelation plot is depicted in the Figure 6.6.

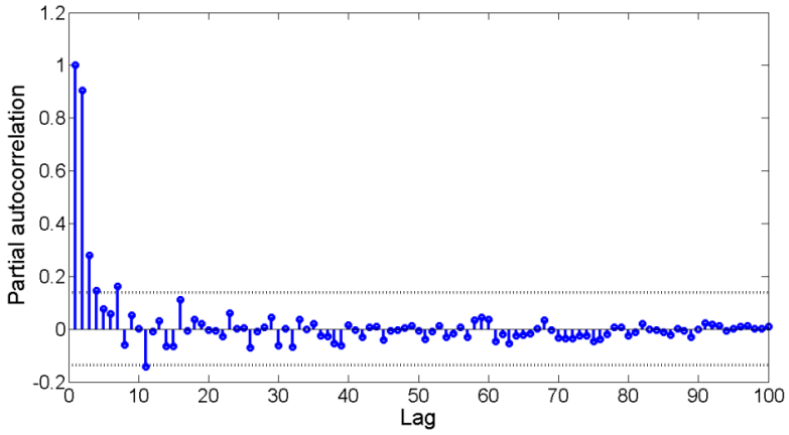

Figure 6.6. Sample partial autocorrelation plot

In Figure 6.6, it is observed that partial autocorrelation is significantly large at lags 1 and 2 (lag 0 is always 1 by definition). The next few lags are a little larger than the confidence limits. If the autocorrelation plot indicates that an AR model is appropriate for this data set, we can choose p=2 and model the data with AR(2) process. AR(3) model can also be checked in the model validation step. In summary, partial autocorrelation of an AR(p) process becomes zero at lag p+1 and greater. This is usually determined by placing a confidence limit on the sample partial autocorrelation plot (6.15); which can be approximated by $\frac{\mp 1.98}{\sqrt{N}}$ [74].

In order to determine the order of MA(q) process, autocorrelation function is used. The function becomes zero at lag q+1 and greater. In practice, the sample autocorrelation and partial autocorrelation functions are random variables which make the model identification more difficult especially for mixed models. Therefore, in recent years information-based criteria such as Final Prediction Error (FPE) and Akaike Information Criterion (AIC) are proposed and used [74].

After identifying the proper model, the next step is to estimate the model parameters. In the following section, parameter estimation for AR model using Yule-Walker (YW) technique is briefly reviewed. By post-multiplying y_{k-1} by (6.8) and getting the expectation we have [78]:

$$E[y_k y_{k-l}] = a_1 E[y_{k-1} y_{k-l}] + a_2 E[y_{k-2} y_{k-l}] + \cdots + a_p E[y_{k-p} y_{k-l}] + E[e_k y_{k-l}]$$

Define: $\gamma(l) = E[y_k y_{k-l}]$

$$\gamma(l) = a_1 \gamma(l-1) + \cdots + a_p \gamma(l-p)$$

For *l=1, 2,…, p* the above equation can be written as :

$$\gamma(1) = a_1\gamma(0) + \cdots + a_p\gamma(1-p)$$

$$\gamma(2) = a_1\gamma(1) + \cdots + a_p\gamma(2-p)$$

$$\vdots$$

$$\gamma(p) = a_1\gamma(p-1) + \cdots + a_p\gamma(0)$$

The above equations in matrix can be formed as:

$$\underline{\gamma} = \Gamma\,\underline{a} \;\rightarrow \underline{a} = \Gamma^{-1}\underline{\gamma}$$

where $\gamma(l)$ is the autocovariance at lag l and is given by:

$$\hat{\gamma}(l) = \frac{1}{N}\sum_{k=1}^{N-|l|} y_k\, y_{k+l}$$

And, Γ is formed as:

$$\hat{\Gamma} = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(1-p) \\ \gamma(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(2-p) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\gamma}(p-1) & \cdots & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

As an example, for auto-regressive model with order one, AR(1), we have:

$$y_k = a_1 y_{k-1} + e_k \tag{6.16}$$

The parameter estimation using Yule-Walker approach having *N* history terms can be performed using the following equation [78]:

$$\widehat{a_1} = \frac{\sum_{k=2}^{N} y_k y_{k-1}}{\sum_{k=2}^{N} {y_k}^2} \tag{6.17}$$

The last stage in modelling procedure is model validation. In order to validate Box-Jenkins models, the error term is assumed to follow a stationary process. Therefore, the difference between the fitted model and the test data, residuals, should be white noise. If this assumption is not valid for the fitted model, we need to restart from the model identification step and try to determine a better model. Usually the analysis of the residuals can provide some clues for a more appropriate model. In the next section, multivariate or vector ARMA model is briefly reviewed.

C. Vector/Multivariate Time Series Models

The multivariate form of the Box-Jenkins univariate models is called Vector Auto-Regressive Moving Average (VARMA) process. If each observation in a time-series is a vector of data points, VARMA model can be used. The VARMA model for a stationary multivariate time series, with a zero mean vector, represented by [78]:

$$\underline{x}_k = \Phi_1 \underline{x}_{k-1} + \cdots + \Phi_p \underline{x}_{k-p} + \underline{e}_t + \Theta_1 \underline{e}_{t-1} + \Theta_2 \underline{e}_{t-2} + \cdots + \Theta_q \underline{e}_{t-q} \tag{6.18}$$

where $\underline{x}_k$ and $\underline{e}_k$ are $n\times1$ column vectors. $\underline{e}_k$ representing multivariate white noise, Φ_k and Θ_k are $n\times n$ matrices for autoregressive and moving average parameters. Like the univariate model, it is assumed that $E[\underline{e}_t]=0$ and $E[\underline{e}_t \underline{e}_{t-k}{}^T]=0$ for $k\neq0$. As an example, for a bivariate series with p=2, and q=1, the VARMA (2,1) model is [74]:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \varphi_{1.11} & \varphi_{1.12} \\ \varphi_{1.21} & \varphi_{1.22} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} \varphi_{2.11} & \varphi_{2.12} \\ \varphi_{2.21} & \varphi_{2.22} \end{bmatrix} \begin{bmatrix} x_{1t-2} \\ x_{2t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{1.11} & \theta_{1.12} \\ \theta_{1.21} & \theta_{1.22} \end{bmatrix} \begin{bmatrix} e_{1t-1} \\ e_{2t-1} \end{bmatrix}$$

If we ignore the MA components, we will have the VAR model given by:

$$\underline{x}_k = \Phi_1 \underline{x}_{k-1} + \cdots + \Phi_p \underline{x}_{k-p} + \underline{e}_k \tag{6.19}$$

The three stages of Box-Jenkins model are also valid for the VARMA model. In order to estimate the model parameters using YW approach, (6.19) is post-multiplied by $\underline{x}_{k-j}{}^T$ for j=0,…,p. By taking expectations, we have [78]:

$$\Gamma(i) = \sum_{j=1}^{p} \Phi_j \, \Gamma(i-j) \; i = 1, \ldots, p \tag{6.20}$$

Given the covariance matrices $\Gamma(0), \ldots, \Gamma(p)$, equation (6.20) can be used to determine the coefficient matrices, $\Phi_1, \ldots \Phi_p$. And, the process noise can then be determined as follows [78]:

$$Q = \Gamma(0) - \sum_{j=1}^{p} \Phi_j \, \Gamma(-j) \tag{6.21}$$

In the following, estimation of the sample covariance matrices is reviewed. These matrices are required in (6.20) and (6.21). A natural estimator of the covariance is $\Gamma(l) = E[(\underline{x}_{k+l} - \underline{\mu})(\underline{x}_k - \underline{\mu})']$, where l is the number of lag (0,1, 2,…, N). The covariance matrix in (6.20) can be estimated as follows [78]:

$$\underline{\mu} = \frac{1}{N}\sum_{k=1}^{N} \underline{x}_k$$

$$\hat{\Gamma}(l) = \frac{1}{N}\sum_{k=1}^{N-l} (\underline{x}_{k+l} - \underline{\mu})(\underline{x}_k - \underline{\mu})' \quad for\ 0 \leq l < N-1 \tag{6.22}$$

where $\underline{\mu}$ is the sample mean. A natural unbiased estimator of the mean vector $\underline{\mu}$ based on the observations $\underline{x}_1, \ldots, \underline{x}_N$ is given by [78]:

$$\underline{\mu} = \frac{1}{N}\sum_{k=1}^{N} \underline{x}_k \tag{6.23}$$

In order to find the ***cross-correlation*** among the time-series, the following equation can be used to estimate $\hat{\rho}_{ij}(l)$ between $y_{k+l}(i)$and $y_k(j)$ [78]:

$$\hat{\rho}_{ij}(l) = \hat{\gamma}_{ij}(l)\sqrt{\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)} \tag{6.24}$$

where $\hat{\gamma}_{ij}$ is the element *i*,*j* of matrix $\hat{\Gamma}(l)$. If *i*=*j* , then $\hat{\rho}_{ij}$ reduces to the sample autocorrelation function of the *i*th series. As an example, for VAR(1), equation (6.19) is changed to :

$$\underline{x}_k = \Phi_1 \underline{x}_{k-1} + \underline{e}_k \tag{6.25}$$

Thus, we only need to calculate Φ_1. Substituting *j*=1 and *p*=1 in (6.20), we have:

$$\Gamma(1) = \Phi_1 \Gamma(0) \tag{6.26}$$

$$\Phi_1 = \Gamma(1)\Gamma(0)^{-1} \tag{6.27}$$

According to equation (6.26), $\Gamma(0)$ should be invertible. In order words, $\Gamma(0)$ should not be singular or ill-conditioned.

D. Summary

In this chapter, fundamentals of time series analysis are described and the common approaches to fit univariate and multivariate models are reviewed. Also, the procedure for identification, estimation and validation of the fitted model is briefly explained for univariate and vector/multivariate autoregressive models. In the next chapter, the described tools in this chapter

and the previous one are used to determine a novel time-series based state transition model for power system FASE.

CHAPTER VII

A STATE TRANSITION MODEL FOR POWER SYSTEM FASE USING TEMPORAL AND SPATIAL CORRELATION OF STATES

A. Motivation and Underlying Assumptions

In this chapter, a new state transition model is presented for power system FASE. The proposed state transition model utilizes time-series analysis. The structure of the model is based on the analyses of temporal and spatial correlation of bus voltages time-series for a given network. This section is dedicated to explain the temporal and spatial correlation within and among the power system states. The causes and effects behind these correlations are discussed followed by the assumptions behind the proposed approach.

It is noted that the concept of state variable as defined in static state estimator is purely algebraic state variable, which does not represent the dynamic state variable of the system. Throughout this section algebraic state variables are also referred to as nodal voltage magnitude and phase angle.

First, let's investigate the temporal correlation in the state time-series:

1. The transition matrix is formed in a pre-determined time interval, ΔT, corresponding to the number of history terms, M, used to fit the model. If each measurement set arrives at the control center with Δk intervals, $\Delta T = M \times \Delta k$.
2. In ΔT, it is assumed that the phasor voltage variations follow a piecewise stationary process and the state transition matrix is held constant. During the time interval, it is assumed that there is no outage, short-circuit, and topology change in the network. The length of the time interval can be in the order of several seconds to few minutes. However, considering the assumption, the approach works effectively for prediction horizon less than a few minutes ahead. The prediction horizon can also be updated with respect to reliable weather forecasts.
3. In ΔT, it is assumed that temperature and humidity do not vary. Consequently, the statistical properties of the load are assumed to be time-invariant. However, the length of ΔT and the prediction horizon can be decreased during on-peak hours, when the demand

significantly increases, and it can be increased during off-peak hours, when the demand drops.

4. Both the load and the intermittent renewable energy have time correlation. As a consequence, due to the deterministic network constraints imposed by circuit laws, each state has correlation in time.

Now, let's investigate the spatial correlation:

5. The spatial correlation of the loads, and the spatial correlation of the intermittent renewable generation contribute to the spatial correlation among the states through the network topology.
6. As a result of the analyses of our simulations, only time correlation is considered for voltage magnitudes due to the slow time response of voltages to changes in reactive power. Since the spatial correlation is not considered for the voltage magnitudes, the time correlation is modeled by using AR model inspired by the approaches used in short-term load forecasting. The voltage magnitudes of voltage-controlled buses are not considered in the model.
7. Since a change in active power rapidly affects the voltage angles, the spatial correlations among the voltage angles in a network are highly observable. The structure and the degree of the spatial correlation among the voltage angles are determined by extensive analysis of the network. The details of the temporal and spatial analyses are provided in Chapter V. The time and the spatial correlation among the voltage angles are captured by using VAR model.
8. Our analyses of simulation results show that the cross-correlation between the times-series of voltage magnitudes and phase angles is insignificant and therefore ignored.
9. The resulting state transition matrix, is non-diagonal, time-variant and can be updated using
    a. State estimates, or
    b. PMU-based measurements.

VAR models are recently used in [80, 81] to determine the regression-based state transition matrix and recently in [79] to address the correlation between PMU measurements. A vector autoregressive model of order p and dimension D, ($\mathrm{VAR}_D(p)$) is defined as follows:

$$\underline{x}_k = \Phi_1 \underline{x}_{k-1} + \Phi_2 \underline{x}_{k-2} + \cdots + \Phi_p \underline{x}_{k-p} + \underline{\varepsilon}_k \tag{7.1}$$

where $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^{D\times D}$, the observation vector is $\underline{x}_k \in \mathbb{R}^D$ and the noise is $\underline{\varepsilon}_k \in \mathbb{R}^D$. The noise is assumed to be Gaussian $\mathcal{N}(0, S)$ where the Φ and the positive definite matrix S are not necessarily diagonal. The spatial correlation among adjacent buses are contained in the non-diagonal elements of the estimated $\hat{\Phi}$ terms and the error covariance, $\hat{S}$. The autocorrelation and partial autocorrelation functions (ACF and PACF) are used to analyze the time correlation in a signal [74]. Based on the analysis of IEEE standard networks, the order of the AR/VAR model is determined as 1. Slack bus and the voltage magnitudes at voltage-controlled buses are excluded from the state vector. Under the above conditions the proposed state transition model is as follows:

$$\begin{bmatrix} \theta_2(k+1) \\ \theta_3(k+1) \\ \vdots \\ \theta_n(k+1) \\ v_2(k+1) \\ v_3(k+1) \\ \vdots \\ v_n(k+1) \end{bmatrix} = \begin{bmatrix} f_{\theta_2,\theta_2} & f_{\theta_2,\theta_3} & \ldots & f_{\theta_2,\theta_n} & 0 & 0 & \ldots & 0 \\ f_{\theta_3,\theta_2} & f_{\theta_3,\theta_3} & \ldots & f_{\theta_3,\theta_n} & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{\theta_n,\theta_2} & f_{\theta_n,\theta_3} & \ldots & f_{\theta_3,\theta_n} & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & f_{v_2,v_2} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & f_{v_3,v_3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & f_{v_n,v_n} \end{bmatrix} \begin{bmatrix} \theta_2(k) \\ \theta_3(k) \\ \vdots \\ \theta_n(k) \\ v_2(k) \\ v_3(k) \\ \vdots \\ v_n(k) \end{bmatrix} + \underline{e}(k) \tag{7.2}$$

The determination of the model parameters and the details of incorporating PMU measurements followed by the illustrative results are provided below.

B. Model Parameter Determination

While regression techniques are used to determine the autoregressive model parameters [72, 75, 76], the method of moments, Yule-Walker estimator (YW), is a more efficient estimator for an AR model. YW estimators reach the Cramer-Rao bound asymptotically and offer the maximum estimation accuracy [78, 82]. Note that, YW is not robust and appropriate steps can be taken to improve the robustness [83].

By post-multiplying the equation (7.1) by ${\underline{x}_{k-j}}^T, j = 0, .., p$ and taking expectations gives the equation (7.3).

$$\Gamma(i) = \sum_{j=1}^{p} \varphi_j \Gamma(i-j)\,, i = 1, \dots, p \tag{7.3}$$

Given the covariance matrices $\Gamma(0), \dots, \Gamma(j)$, equation (7.3) can be used to determine the coefficient matrix of voltage angles in (7.2). A natural estimator of covariance is estimated as follows [78]:

$$\hat{\Gamma}(h) = \frac{1}{M} \sum_{k=1}^{M-h} (\underline{x}_{k+h} - \underline{\mu})(\underline{x}_k - \underline{\mu})' \tag{7.4}$$

where $0 \leq h < M-1$, t is the number of considered history terms in the pre-determined interval based on key point 2 in Section A and, $\underline{\mu}$ is the sample mean. Effective fitting of an autoregressive process requires a moderately long time series [76]. A lower bound for the number of history terms, *M*, is determined based on analyses and offline simulations to reduce the computational load while entertaining the assumption in key point 2. In the presence of faster sampling rates (e.g. half a second or less), the only constraint for the lower bound is computational load. However, if slower sampling rates are used (e.g. one minute), considering more than 15 history terms (equal to 15 min) will violate the assumption of piecewise stationarity. The following equation is used to determine the forecasting error covariance matrix [78]:

$$S = \Gamma(0) - \sum_{j=1}^{p} \Phi_j \Gamma(-j) \tag{7.5}$$

The proposed $\mathrm{VAR}_D(1)$ in (7.2) can be rewritten as:

$$\underline{x}_k = \Phi_1 \underline{x}_{k-1} + \underline{\varepsilon}_k \tag{7.6}$$

Therefore, substituting p =1 and j =1 in (7.3) we have:

$$\Gamma(1) = \Phi_1 \Gamma(0)$$

And,

$$\Phi_1 = \Gamma(1)\Gamma(0)^{-1} \tag{7.7}$$

Construction of $\Gamma(0)$ and $\Gamma(1)$ is demonstrated using a 5-bus network in Figure 7.1. Bus 1 is chosen as reference. The procedure for the state transition model using VAR is demonstrated

only for the voltage angle of Bus 3. The same procedure is applied to all other bus angles except the slack bus.

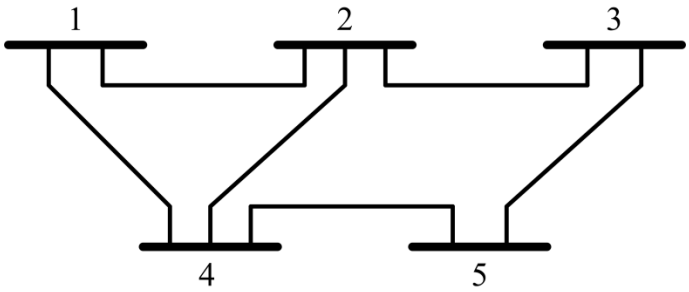

Figure 7.1. 5-bus network

The time series model for bus 3 voltage angle is written as:

$$\begin{bmatrix}\theta_2(k+1)\\ \theta_3(k+1)\\ \theta_5(k+1)\end{bmatrix} = \begin{bmatrix}f_{\theta_2,\theta_2} & f_{\theta_2,\theta_3} & f_{\theta_2,\theta_5}\\ f_{\theta_3,\theta_2} & f_{\theta_3,\theta_3} & f_{\theta_3,\theta_5}\\ f_{\theta_5,\theta_2} & f_{\theta_5,\theta_3} & f_{\theta_5,\theta_5}\end{bmatrix}\begin{bmatrix}\theta_2(k)\\ \theta_3(k)\\ \theta_5(k)\end{bmatrix} + \underline{e}(k)$$

The transition matrix parameters are estimated as follows:

$$\begin{bmatrix}\hat{f}_{\theta_2,\theta_2} & \hat{f}_{\theta_2,\theta_3} & \hat{f}_{\theta_2,\theta_5}\\ \hat{f}_{\theta_3,\theta_2} & \hat{f}_{\theta_3,\theta_3} & \hat{f}_{\theta_3,\theta_5}\\ \hat{f}_{\theta_5,\theta_2} & \hat{f}_{\theta_5,\theta_3} & \hat{f}_{\theta_5,\theta_5}\end{bmatrix} = \hat{\Gamma}(1)\hat{\Gamma}(0)^{-1}$$

The $\Gamma(0)$ and $\Gamma(1)$ for bus 3 are estimated using M history terms using Equation (7.4) as follows:

$$\hat{\Gamma}(0) = \frac{1}{M}\sum_{j=1}^{M}\begin{bmatrix}\theta_2(k-j)-\mu_{\theta_2}\\ \theta_3(k-j)-\mu_{\theta_3}\\ \theta_5(k-j)-\mu_{\theta_5}\end{bmatrix}\begin{bmatrix}\theta_2(k-j)-\mu_{\theta_2}\\ \theta_3(k-j)-\mu_{\theta_3}\\ \theta_5(k-j)-\mu_{\theta_5}\end{bmatrix}^T$$

$$\hat{\Gamma}(1) = \frac{1}{M}\sum_{j=1}^{M-1}\begin{bmatrix}\theta_2(k-j-1)-\mu_{\theta_2}\\ \theta_3(k-j-1)-\mu_{\theta_3}\\ \theta_5(k-j-1)-\mu_{\theta_5}\end{bmatrix}\begin{bmatrix}\theta_2(k-j)-\mu_{\theta_2}\\ \theta_3(k-j)-\mu_{\theta_3}\\ \theta_5(k-j)-\mu_{\theta_5}\end{bmatrix}^T$$

where $\mu_{\theta_2}, \mu_{\theta_3}$ and μ_{θ_5} are sample mean of the voltage angles, which are calculated as follows:

$$\begin{bmatrix}\mu_{\theta_2}\\ \mu_{\theta_3}\\ \mu_{\theta_5}\end{bmatrix} = \frac{1}{M}\sum_{j=1}^{M}\begin{bmatrix}\theta_2(k-j)\\ \theta_3(k-j)\\ \theta_5(k-j)\end{bmatrix}$$

The resulting $\hat{\Gamma}(0)$ has the following form:

$$\hat{\Gamma}(0) = \begin{bmatrix} \hat{\gamma}_{\theta_2,\theta_2} & \hat{\gamma}_{\theta_2,\theta_3} & \hat{\gamma}_{\theta_2,\theta_5} \\ \hat{\gamma}_{\theta_3,\theta_2} & \hat{\gamma}_{\theta_3,\theta_3} & \hat{\gamma}_{\theta_3,\theta_5} \\ \hat{\gamma}_{\theta_5,\theta_2} & \hat{\gamma}_{\theta_5,\theta_3} & \hat{\gamma}_{\theta_5,\theta_5} \end{bmatrix}$$

Note that the diagonal terms of $\hat{\Gamma}(0)$ reflect the temporal correlation and the off-diagonal terms reflect the spatial correlation of the voltage angles. VAR(1) model can result in better state forecasting accuracy compared to AR(1) if the magnitude of off-diagonal terms of the $\Gamma(0)$ and $\Gamma(1)$ matrices are not small compared to the diagonal terms.

Since the spatial correlation among the states of

- adjacent buses of electrical hubs, and
- buses electrically close to electrical hubs

cannot be observed and inclusion of these buses lead to singularity of $\Gamma(0)$; they are not taken into account in the model. As pointed out in Chapter V, the states of groups of buses, which are electrically close to each other with high electrical centrality are not spatially correlated but highly correlated in time since the variations in their states are similar. Thus, VAR model is used for networks with fewer electrical hubs and smaller electrical centrality. The level of non-singularity of $\Gamma(0)$ in (7.7) is used to determine the 'eligible buses' for VAR model. The eligible buses are those buses that have high degree of spatial correlation with their adjacent buses. The eligible buses are determined by using the time series of filtered states and forming the $\Gamma(0)$ for each bus with their adjacent buses. The condition number of each $\Gamma(0)$ is determined, they are sorted, and the value when a significant difference is observed is set as the threshold. Those buses with smaller condition number than the threshold are considered as the eligible buses. Therefore, *F* is non-diagonal for voltage angles of the eligible buses in which spatial correlation is observed, and is diagonal for the rest of buses.

C. Incorporation of PMU Measurements

Given that PMUs provide measurements in much faster rates than the conventional measurements, FASE algorithms can exploit the PMUs in state forecasting stage to capture sudden variations in states. In the proposed model the voltage measurements from PMUs can be used along with the conventional measurements as shown in Figure. 7.2 in both forecasting and filtering stages of FASE.

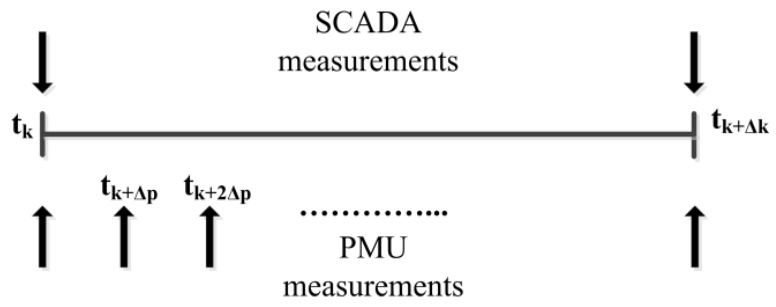

Figure 7.2. Arrival times of SCADA and PMU measurements

As Figure. 7.2 shows; k represents the time instant when conventional measurements are available and Δk is used to present the time interval for conventional measurements or the waiting period, the period between two successive SE runs. Δp is used for the time interval for PMU measurements. Once a voltage measurement from PMU is available at $k + \Delta p$; the PMU measurements are used with the conventional measurements arrived at time k to estimate the state trend by using the weighted least square (WLS) state estimator. Complex voltages obtained from PMUs are included directly into the state estimator with higher weights than the normal measurements. The corresponding entries in the measurement Jacobian matrix (i.e. H_{pmu}) are one and the measurement Jacobian has the following structure:

$$H_{new} = \begin{bmatrix} H_{conventinal} \\ H_{pmu} \end{bmatrix}$$

In addition, the measurement vector can be modeled as:

$$z_{new} = \begin{bmatrix} z_{conventinal} \\ z_{pmu} \end{bmatrix}$$

where,

$$z_{conventinal} = \begin{bmatrix} P_i \\ P_{ij} \\ Q_i \\ Q_{ij} \\ V_i \end{bmatrix}, z_{pmu} = \begin{bmatrix} V_{pmu} \\ \theta_{pmu} \end{bmatrix}$$

P_i,Q_i, P_{ij} and Q_{ij} are the active/reactive powers injection and flows, respectively. V_i is voltage magnitude measurement which is provided by conventional measurement devices. V_{pmu} and θ_{pmu} are the phase angle and voltage magnitude measurements from PMUs. The measurement covariance matrix is modified to:

$$R = \begin{bmatrix} R_{conventinal} & 0 \\ 0 & R_{pmu} \end{bmatrix}$$

where entries of R_{pmu} are much smaller that the entries of $R_{conventinal}$ due to high accuracy of PMU measurements.

The conventional measurements arrived at time *k* are used each time a new PMU measurement is available to calculate the state trend and providing the history terms for the proposed time-series based state transition model. At each time step, *k* the following steps are performed:

1. Forecast the system state at time $k + \Delta k$ by using (7.2).
2. Update the state transition matrix, F_k by using the historical values filtered by the PMU measurements using WLS state estimator. F_k is updated by using the trend states calculated at time steps $(k - \Delta p)$, $(k - 2\Delta p)$,…, $(k - M\Delta p)$ where *M* is the number of historical terms, which is determined based on analyses and offline simulations as mentioned in Section A.

Once new conventional measurements along with PMU measurements are available at $k + \Delta k$, the state is filtered using (4.6)–(4.9). Using PMU measurements to calculate state trends improves the accuracy of the state transition matrix, subsequently improving the state forecasting [84]. In addition, state trend analysis can also be used to detect changes in switch status, which may improve the topology estimation and error detection [85].

In the following section, the illustrative results on IEEE testbeds are provided with detailed discussion on the performance of the proposed model in presence of spatial correlation, with incorporation of PMU measurements, in presence of large-scale wind integration, and during the loss of observability.

D. Illustrative Results

The proposed technique is tested on IEEE 57-, 118- and 300-bus systems in MATLAB environment. The simulations are performed using scaled aggregated 5-min load and wind data from BPA [86]. In order to simulate faster sampling rate, the five-minute interval of the load data is linearly filled with ten data points. The percentage of change in load from onetime step to another is 0.2%. The random fluctuation of load at each bus is simulated as follows:

$$P_i^k = P_i^k + \zeta_i^k \quad \zeta_i^k \sim \mathcal{N}\left(0, \alpha P_i^k\right)$$

where P_i^k is the active load at bus i at time k and ζ_i is a normal random variable with zero mean and standard deviation 0.2% of the load at each bus. Note that, the power factor at each bus is held constant. The actual states are obtained by running power flow. In order to determine the parameters of the VAR and AR models, M is chosen as 20 history terms and based on offline analyses of each network. The spatial correlation of renewable sources such as wind speed and solar irradiation will impact the spatial correlation among the states; however, for the sake of simplicity this correlation is ignored. The measurements are configured such that sufficient redundancy is achieved for state estimation. The measurement configuration and redundancy ratio for each test system is provided in is provided in Table 7.1.

Table 7.1. Measurement Configuration for IEEE Test Networks

Measurement Type	**IEEE 57-bus**	**IEEE 118-bus**	**IEEE 300-bus**
Power flow pair	48	179	411
Power injection pair	41	12	25
Voltage magnitude	2	4	8
Redundancy ratio	**1.59**	**1.64**	**1.45**

The measurement correlation reported in [87, 88] is ignored for simplicity and measurement error covariance matrix, R, is assumed to be diagonal. Refer to Table 7.2 for the selected standard deviations for different types of measurements [28].

Table 7.2. Standard Deviations for Measurements

Type	Flow	Injection	|V|
σ	0.008	0.010	0.004

As explained in key point 2 in Section A, stationarity is assumed over the selected time interval. As an example, the ACF and PACF of bus 19 of IEEE 57-bus are depicted in Figure 7.3 and Figure 7.4, respectively. Note that the signal is not independent identically distributed, i.i.d., but correlated. The ACF and PACF show that the process is stationary for the selected time interval, which comprises M=20 history terms and obeys a short-memory autoregressive moving average (ARMA) model [74]. This observation is in good agreement with the interpretations provided in Table 6.1 in Chapter VI.B.4. In addition, since it decays exponentially, including moving average (MA) terms is not necessary. M=20 history terms provides us with an acceptable forecasting accuracy (the state forecasting accuracy did not improve significantly by using more history terms in our studies). In Figure 7.4, it is observed that only lag zero and lag one are greater than the confidence bands, which are approximately equal to $\mp\frac{1.98}{\sqrt{M}}$, where M is the number of history terms [74]. Therefore, the order of the selected AR process does not need to

be more than one. In Figure 7.5 and Figure 7.6, sample cross-correlations between two electrically far (buses 10 and 57) and close (buses 56 and 57) buses are shown, respectively. Since the existence of dependence is easily observed in Figures 7.5 and 7.6, the cross-correlation should be modeled by a vector AR (VAR) rather than a single scalar cross-correlation term. We emphasize the importance of electrical distance in detecting and modeling time and spatial correlation.

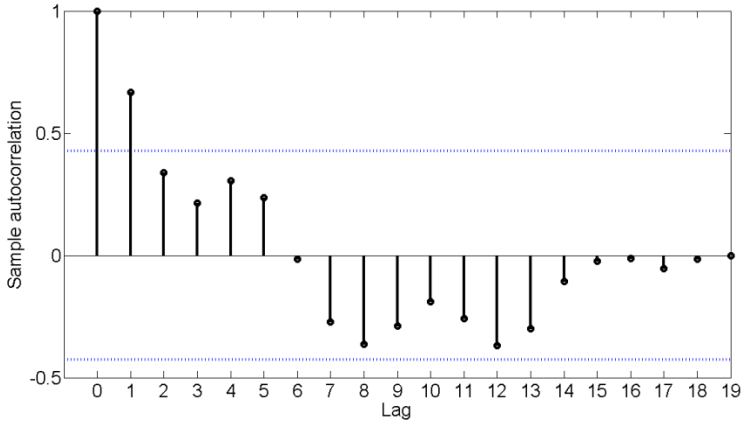

Figure 7.3. Sample autocorrelation of voltage angle time-series (with confidence intervals)

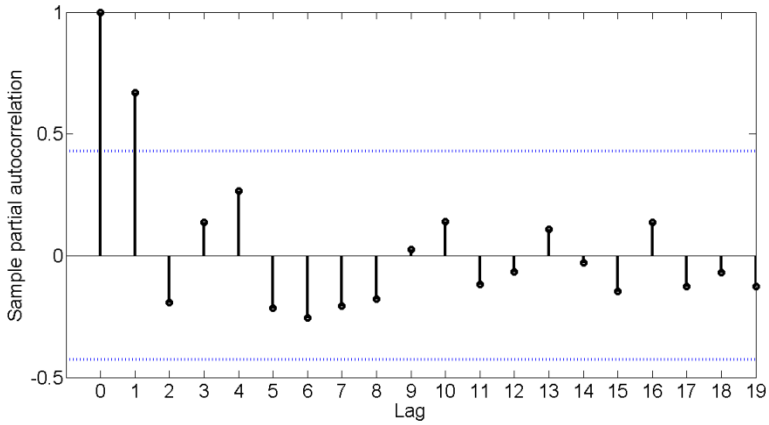

Figure 7.4. Sample partial autocorrelation of voltage angle time-series (with confidence intervals)

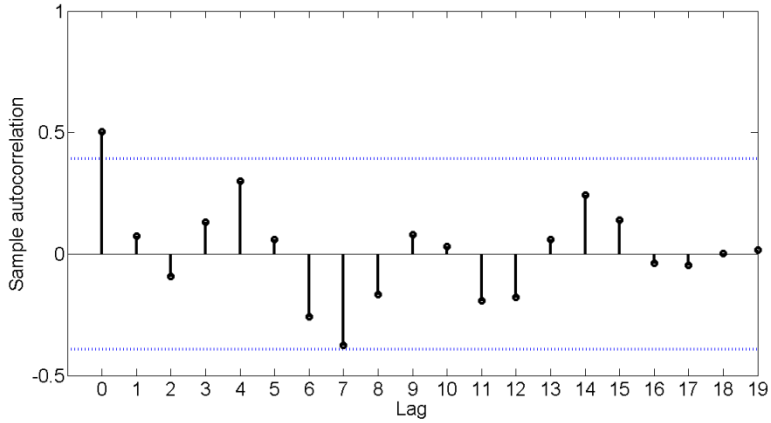

Figure 7.5. Sample cross correlation of voltage angle time-series (with confidence intervals) for two electrically far buses

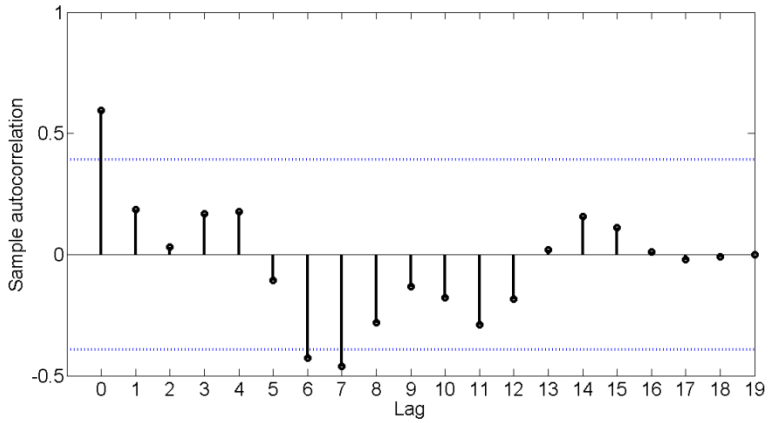

Figure 7.6. Sample cross correlation of voltage angle time-series (with confidence intervals) for two electrically close buses

In order to compare the state forecasting accuracy of the proposed model with the conventional methods, $\aleph_c$ runs of Monte Carlo (MC) simulations are performed and the mean-absolute-error (MAE) is evaluated as follows:

$$\frac{1}{\aleph_c}\sum_{i=1}^{\aleph_c} e_i \tag{7.8}$$

where $e_i = \frac{1}{\aleph_s}\sum_{j=1}^{\aleph_s}\left|x_j - \tilde{x}_j\right|$ is the MAE at *i*th MC run. $\aleph_s$ is the number of states (voltage angles or voltage magnitudes), x_j is the actual state at bus *j* obtained through power flow simulations and $\tilde{x}_j$ is the corresponding forecasted state. The results for $\aleph_c$ =100 for voltage angles, voltage magnitudes and voltage angles of the eligible buses are presented in Tables 7.3, 7.4 and 7.5, respectively.

Table 7.3. Comparison of Forecasting Accuracies of Voltage Angles

Method	MAE [degree]		
	IEEE 57-bus	**IEEE 118-bus**	**IEEE 300-bus**
Debs's	0.1002	0.2034	0.2266
Silva's	0.0931	0.2011	0.2194
AR(1)	0.0317	0.1985	0.1393

Table 7.4. Comparison of Forecasting Accuracies of Voltage Magnitudes

Method	MAE [pu]		
	IEEE 57-bus	**IEEE 118-bus**	**IEEE 300-bus**
Debs's	0.0041	0.0087	0.1335
Silva's	0.0041	0.0087	0.1234
AR(1)	0.0013	0.0086	0.0463

Table 7.5. Comparison of Forecasting Accuracies of Voltage Angles at Eligible Buses

Method	MAE [degree]		
	IEEE 57-bus	**IEEE 118-bus**	**IEEE 300-bus**
Debs's	0.0953	0.2850	0.3698
Silva's	0.0892	0.2828	0.3498
AR(1)	0.0306	0.2853	0.1786
VAR(1)	0.0474	0.2901	0.1929
# of eligible buses	**44**	**93**	**91**

Following are the detailed discussions on the results for lack and in presence of spatial correlation, in presence of large-scale wind integration, for incorporating PMU measurements, and for the performance during loss of observability:

a. *Lack of spatial correlation:* Only AR model is used to capture the time correlation when no spatial correlation is observable. In Table 7.3, it is observed that the improvement in state forecasting accuracy using AR(1) is more significant for 57- and 300-bus networks than it is in 118-bus network. The similar trend is also observed for voltage magnitudes in Table 7.4. The reason is due to the existence of considerable amount of buses with small electrical centrality in 57- and 300-bus networks compared to 118-bus network. In order to have a better insight, electrical centrality of buses in 300-bus network, and the phase angle forecasting accuracies of Debs's method, Silva's method and AR(1)-based model are presented in Figures 7.7 and 7.8, respectively.

   It is observed that forecasting accuracy is significantly improved for buses with relatively small electrical centrality. The reason is that the time-series of states of buses with relatively large electrical centrality are highly correlated in time, and the conventional approaches are only good enough in presence of very high correlation in time (the correlation coefficient of

the existing approaches are fixed and equal to one and 0.8 for Debs' and Silva's technique, respectively [6, 41]). However, time-series of states of buses with relatively small electrical centrality buses, which are highlighted in Figures 7.7 and 7.8, are not highly correlated in time and AR(1) model is more accurate for catching any degree of correlation.

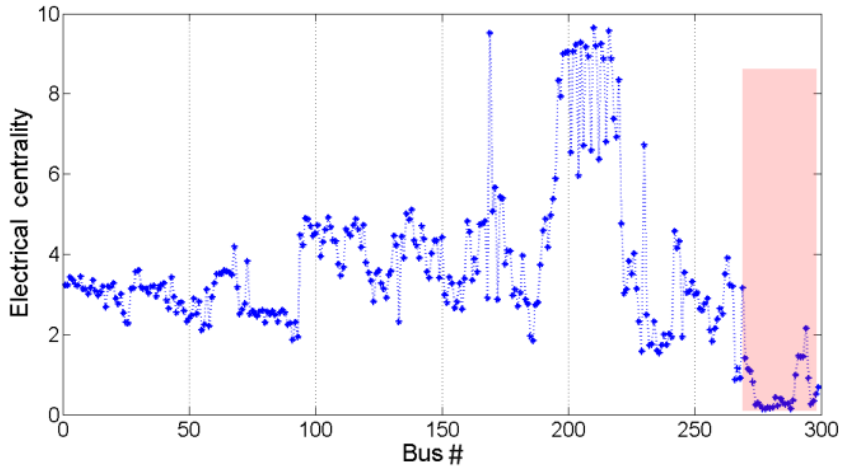

Figure 7.7. Electrical centrality of buses of IEEE 300-bus network

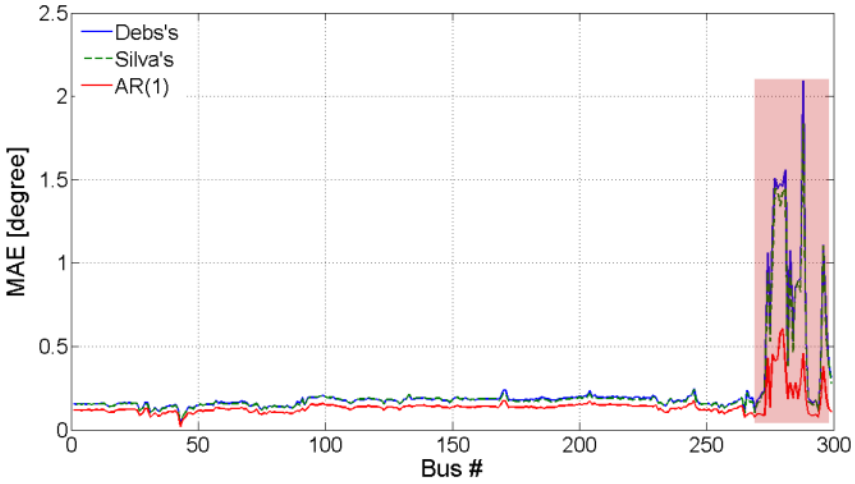

Figure 7.8. Phase angle forecasting accuracy comparison for IEEE 300-bus network

Table 7.5 presents the forecasting accuracies *only* for the eligible buses, and the improvement using VAR(1) is better than the existing methods; however, worse than the improvement using AR(1). This demonstrates the presence of more time correlation rather than spatial correlation among voltage angles at the eligible buses. Furthermore, the percentage of eligible buses in 57-bus is significantly more than in 118- and 300-bus networks. This observation is in good agreement with the discussion in Section B.

The condition number of $\Gamma(0)$ for each bus with its adjacent buses for IEEE 57-bus system is determined, sorted in ascending order, and demonstrated in Figure 7.9 where the threshold is also depicted. It is observed that there are 13 buses with greater condition number than the threshold. The temporal correlation between these buses and their adjacent buses is higher than the desired correlation level. This is the reason for large condition number of $\Gamma(0)$ in Equation (7.7) for those buses. The pattern of voltage angle partition of F_k for this network is shown in Figure 7.10.

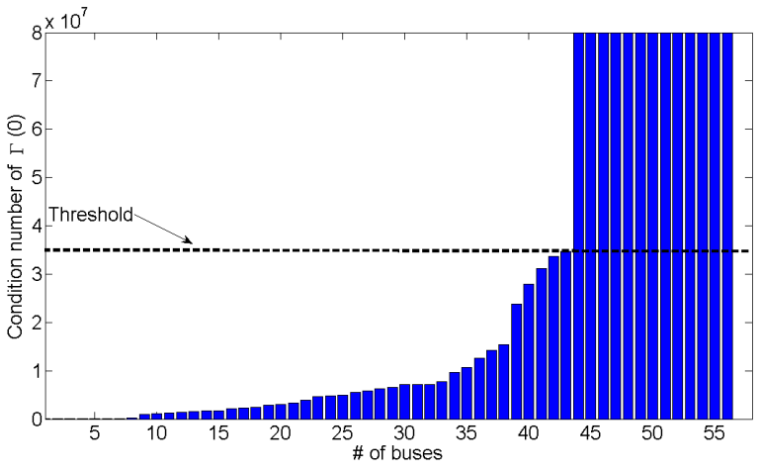

Figure 7.9. Sorted condition numbers of $\Gamma(0)$ for $\mathrm{VAR}_d(1)$ of each bus in IEEE 57-bus network

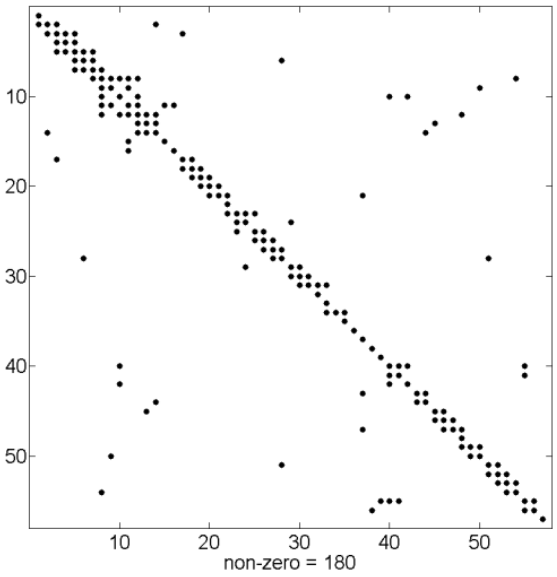

Figure 7.10. The pattern of voltage angle partition of *F* for IEEE 57-bus network

It is observed in Figure 7.10 that the voltage angle partition of F is asymmetric. This feature can be explained with an example. Considering 5-bus network in Figure 7.1, bus 5 is connected to buses 3 and 4 and bus 4 is connected to bus 1, 2 and 5. In order to determine the VAR model for bus 4 and 5, voltage angles of buses 2,4,5 and 3,4,5 are considered, respectively. After estimation of $\Gamma(0)$, if the condition number of $\Gamma(0)$ for (2,4,5) is smaller than (3,4,5), (2,4,5) are used for VAR model of bus 5 instead of (3,4,5). As discussed, smaller condition number can be interpreted as the presence of relatively higher degree of spatial correlation among the states which results in better forecasting accuracy using VAR model.

b. *Presence of spatial correlation*: In order to demonstrate the effectiveness of the method, the degree of spatial correlation among the states is increased by performing modifications on 118- and 300-bus networks. Since the number of electrical hubs in 118-bus network is not significant, the degree of spatial correlation is increased by distributing the significant nodes similar to those in 57-bus network. This is attained by changing the load patterns. In the case of high number of distributed generation and microgrids, the demands at the point of connections are insignificant. This information is the basis of the modification of the 118-bus network.

However, the objective of the modifications in 300-bus network was to reduce the number of electrical hubs in the system. This is attained by topological modifications such as increase in reactances of lines connected to electrical hubs or to buses with high centrality. These changes simulate the proliferation of long transmission lines to transport large-scale wind power to load centers. In addition, strengthened tie-lines among areas with new parallel transmission lines in the future grid are represented with decrease in reactances. The effects of topological changes translate into changes in apparent impedances, which subsequently affect the number of electrical hubs. The absolute values of the electrical node significance of 118-bus and its modified version are sorted in descending order and presented in Figure 7.11 for the first 70 buses. Changing the number and size of significant buses of 118-bus network results in increase in the number of eligible buses demonstrated in Table 7.8.

In Figure 7.12, buses are sorted in descending order based on their number of electrical connections and the first 70 buses are given for 300-bus and its modified version. The

number of electrical connections for the network is chosen to be equal to the number of branches, 411. In order to compare the electrical connectivity and the number of electrical hubs; a threshold is set as $\kappa_{hub} = 20$. It is observed that the number of electrical hubs in the modified 300-bus are reduced to 5 (originally 7). Also, the number of electrical connections for each hub is reduced.

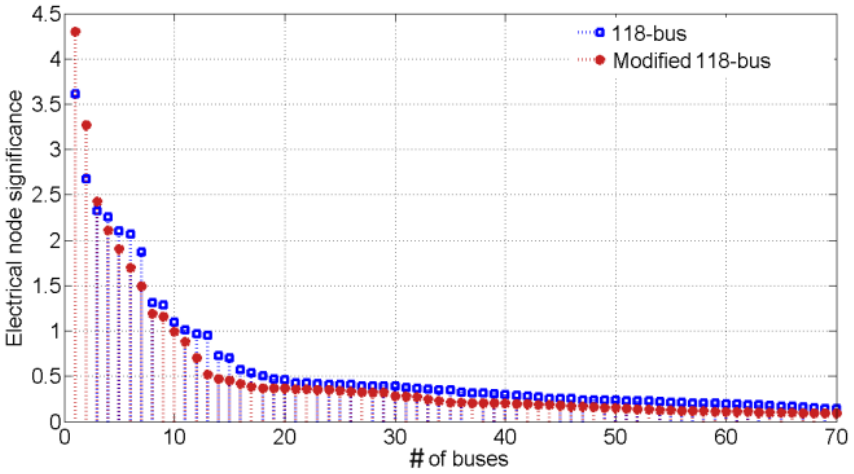

Figure 7.11. Sorted electrical node significance of IEEE 118- and modified 118-bus networks for the first 70 buses

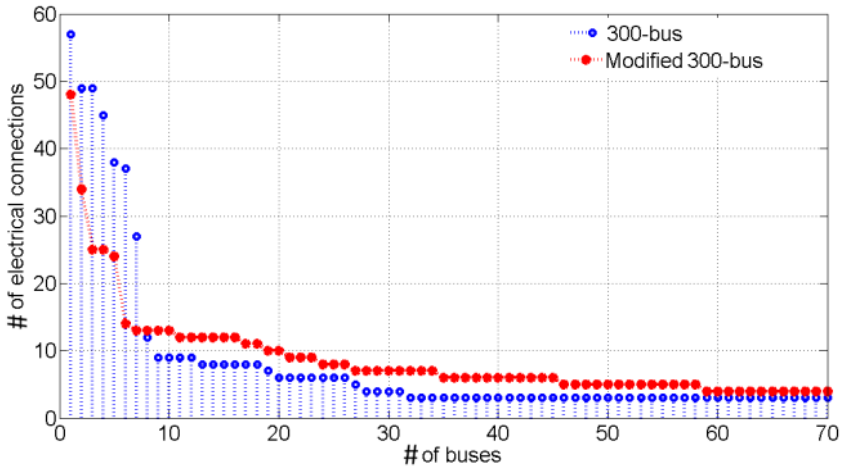

Figure 7.12. The sorted buses with strongest connections for IEEE 300- modified 300-bus networks. Only the first 70 buses are shown

The state forecasting accuracies for the eligible buses in modified networks are increased using VAR(1) model as presented in Table 7.8. The results confirm that the modifications increased the spatial correlation among voltage angles of each eligible bus with those of its adjacent buses. The state forecasting accuracy for the eligible buses is improved due to

decrease in number of electrical hubs and subsequent increase in spatial correlation as well as the number of eligible buses.

Table 7.6. Comparison of Forecasting Accuracies of Voltage Angles at Eligible Buses

Method	**MAE [degree]**			
	118-bus	**Mod. 118-bus**	**300-bus**	**Mod. 300-bus**
Debs' s	0.2850	0.2320	0.3698	0.3761
Silva's	0.2828	0.2321	0.3498	0.3574
AR(1)	0.2853	0.2244	0.1786	0.3297
VAR(1)	0.2901	0.2210	0.1929	0.3065
# of eligible buses	**93**	**96**	**91**	**274**

c. *Presence of large-scale distributed wind:* Integration of large-scale, distributed, and intermittent generation, and/or microgrids will result in an increase of spatial correlation among the states. In this situation, using VAR(1) model will improve the state forecasting significantly for the wind buses and/or for the buses electrically close to the wind connections. Various patterns of wind generation can also increase the spatial correlation among the states and improve the state forecasting accuracy [89]. MAEs of the state forecasting accuracies at eligible buses are demonstrated in Table 7.9 for 57-bus network in presence of 30% wind integration.

Table 7.7. Comparison of Forecasting Accuracies of Voltage Angles at Eligible Buses

Method	**MAE [degree]**	
	No wind	**With Wind**
Debs's	0.0953	0.1083
Silva's	0.0892	0.1035
AR(1)	0.0306	0.0946
VAR(1)	0.0474	0.0848
# of eligible buses	**44**	**42**

Wind generation is assumed to be located at buses 4, 21, 22, 37, 45, and 48. It is noted that the loss of wind is compensated by all the generators according to their participation factors. It is observed that forecasting accuracy of the buses 17, 19, 20, 43, 53, 54, and 55 (those are adjacent to wind buses and have small electrical centrality) are improved using VAR model. Moreover, the quantity of eligible buses changed (reduced in this case). Figure 7.13 shows that the condition numbers of the $\Gamma(0)$s for wind buses and their adjacent buses (e.g. bus 22, 23 and 48) significantly dropped due to an increase in spatial and a decrease in time correlation among the states.

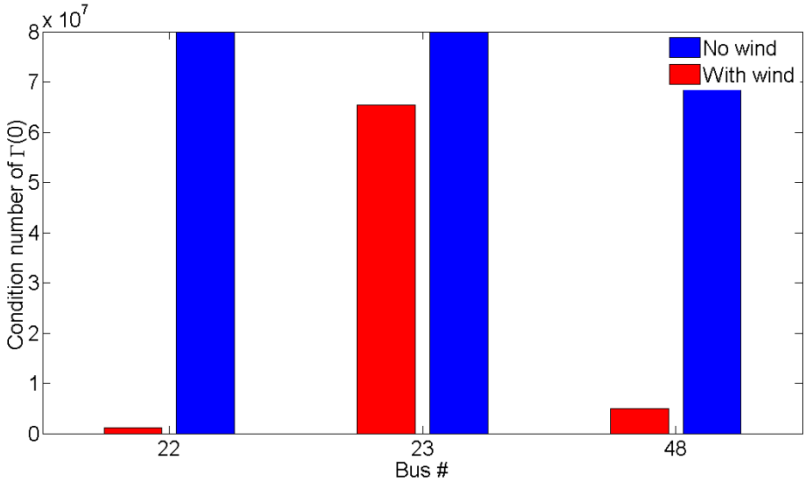

Figure 7.13. Condition numbers of $\Gamma(0)$ for buses 22, 23 and 47 in 57-bus network in presence of and without 30% wind integration

d. Incorporating PMU measurements: The proposed technique is also tested on IEEE 118-bus system in presence of PMU measurements. In addition to the conventional measurements of Table 7.1, 15 voltage measurements are assumed to be provided by PMUs installed on 15 high-voltage (HV) buses (8, 9, 10, 26, 30, 38, 63, 64, 65, 79, 81, 105, 109, and 113). The standard deviation of voltage angles and voltage magnitude measurements are considered to be 0.0001 and 0.004, respectively [90]. In order to simulate the PMU measurements with faster sampling rate, the 5-minute interval of the load data is linearly filled with extra data points. The percentage of change in load from one- time step to another is considered to be 0.02%. In Tables 7.10 and 7.11, it is observed that incorporation of PMU measurements significantly improves the forecasting accuracy. This betterment is more obvious for phase angles since their variations are directly related to active power variations in the system.

Table 7.8. Comparison of Voltage Angle Forecasting Accuracies in presence of PMU Measurements

Method	**MAE [degree]**
	IEEE 118-bus
Debs's	0.0558
Silva's	0.0536
AR(1)	0.0529

Table 7.9. Comparison of Voltage Magnitude Forecasting Accuracies in presence of PMU Measurements

Method	**MAE [pu]**
	IEEE 118-bus
Debs's	0.0070
Silva's	0.0070
AR(1)	0.0068

e. *Performance during loss of observability*: As discussed in Chapter IV, one important benefit of FASE is to provide the system operator with a predictive database, which can be used for enhancing different layers of state estimation (SE) such as observability. In presence of a sufficient degree of spatial correlation among the states, VAR(1) model can be used to forecast the future value of a desired state using the information of only its adjacent buses in presence of loss of observability of the desired bus due to a loss of telemetry. For example, assume that bus *b* is an important bus for the power system operator with large load and/or generator connection. Assume that *b* is connected to buses a_1,.., a_j. In time of loss of observability rendering an unobservable island including bus *b*, the VAR(1) coefficients can be used to provide state forecasting based on the available information of the adjacent buses. The simulation is performed for IEEE 57-bus to demonstrate the capability of VAR(1) model in presence of loss of observability. Bus 12, with the largest generator and load connection, is assumed to be a part of an unobservable island at time sample 101. It is connected to buses 9, 10, 13, 16 and 17. The last available telemetry from this bus (i.e. the filtered state, voltage angle, at time 100) is used in VAR(1) model. The performance of the proposed method is compared with other methods in Figure 7.14.

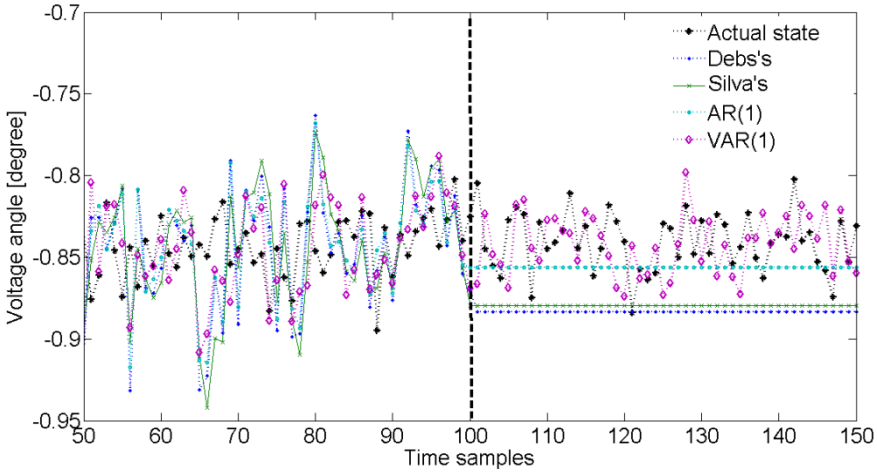

Figure 7.14. VAR(1) performance compare to the conventional methods in presence of loss observability

It is observed that VAR(1)-based model successfully tracks the voltage angle while the other methods cannot. Table 7.12 presents the MAEs for 50 samples following loss of telemetry. Note the significant improvement using VAR(1) model.

Table 7.10. Comparison of Existing Methods with the Proposed Technique

Method	MAE [degree]
Debs's	0.0840
Silva's	0.0716
AR(1)	0.0709
VAR(1)	0.0408

E. Summary

In this chapter, the non-diagonal, and time-variant state transition matrix is proposed using time-series modeling of complex bus voltages. The proposed model is designed for the state forecasting stage of power system forecasting-aided state estimators. The model takes the temporal and spatial correlation in and among the power system states into account. Electrical centrality, node significance, and connectivity in a given network are used to capture the temporal and spatial correlation among the states. This information is used to determine the structure of the proposed state transition matrix. The transition matrix is updated once new filtered states are available in the historical database. The simulations and analyses are performed on IEEE 57-, 118- and 300-bus networks. The proposed technique is independent of network size and gives satisfactory results for large networks especially in presence of distributed large- and small-scale intermittent power generation and microgrids.

CHAPTER VIII

CONCLUSIONS AND CONTRIBUTIONS

This dissertation tackles with the main challenge in FASE: the lack of a deterministic state transition model in stead-state operation. The represented models suffer from the following drawbacks:

1. Time correlation is not addressed and exploited effectively. The impact of high percentage integration of intermittent generation on the system states can be exploited to capture the short-term behavior of states. The proper exploitation of the increased randomness in system states by using statistical models can help improving state transition model under certain assumptions.
2. The power system states are assumed to vary independently from each other, which leads to a diagonal state transition matrix. This assumption is not realistic in power systems in the presence of microgrids, distributed and intermittent power generation, which sometimes are connected to the grid through radial buses. The grid of the future is becoming more decentralized resulting in the decrease of number of buses with large amount of power injections (so called 'electrical hubs'.) Therefore, these electrical hubs lose their significance and dominance in the grid, and the spatial correlation among the states throughout the grid become more observable. Spatial correlation defines the level of interaction of a state with the states of the neighboring buses. Taking spatial correlation into account enables the operator to restore observability when a state is unobservable due to loss of telemetry.
3. The existing state transition models are not effectively utilizing this valuable information provided by the phasor measurement units (PMUs), a new measurement technology that provides time-stamped, synchronized complex voltage and current information from various locations throughout the grid. This information enables the system operators to be able to monitor the system with measurements at unprecedented frequencies (unlike the regular measurements, which arrive at every few seconds, the measurements provided by PMUs arrive at control centers every $1/30^{th}$ of a second.) In addition PMUs provide the angles of bus voltages with respect to a reference unlike the conventional measurements, which are only capable of providing the voltage magnitudes.

In order to tackle the drawback of existing models a non-diagonal and time-variant state transition matrix is proposed using time-series modeling of complex bus voltages. The proposed model is designed for the state forecasting stage of power system forecasting-aided state estimators. The model takes the temporal and spatial correlation in and among the power system states into account. Electrical centrality, node significance, and connectivity in a given network are used to capture the temporal and spatial correlation among the states. This information is used to determine the structure of the proposed state transition matrix. The transition matrix is updated once new filtered states are available in the historical database. The simulations and analyses are performed on IEEE 57-, 118- and 300-bus networks. The proposed technique is independent of network size and gives satisfactory results for large networks especially in presence of distributed large- and small-scale intermittent power generation and microgrids.

We believe that the following challenges have potential for future research:

1. Recursive Levinson-Durbin algorithm (LD) can be considered instead of YW to improve the computation speed in the state transition model determination.
2. If voltage phasor measurements at each bus are available using PMUs, these measurements can be directly used to determine the state transition matrix. Spatial correlation among the states will also be observed better in this case.
3. A more robust approach, such as ratio-of-medians-based estimator (RME), can be considered instead of YW; however, it is absolutely necessary if PMU measurements are used directly to determine the state transition model.
4. The forecasted states can be used in combination with the most recent measurements to improve the redundancy ratio. This will also improve bad data processing especially in presence of critical measurements.
5. In order to reduce the computation burden in the filtering stage, recursive least-square estimator is recommended.
6. In the model determination stage, it is assumed that the state variations follow a piecewise stationary process within ΔT and the state transition matrix is held constant. Proper approaches such as ‘cointegration’ technique can be used in presence of non-stationary state variations. Also, the non-Gaussian probability density function of intermittent generation may require more complicated models such as generalized

autoregressive conditional heteroskedasticity (GARCH) model instead of AR-based models.

7. In ΔT, the statistical properties of the load are assumed invariant and temperature and humidity do not vary. Proper approaches for varying statistical properties can be considered.
8. The impact of spatial correlation of renewable sources such as wind speed and solar irradiation on the spatial correlation among the states is ignored. Methods to integrate this correlation can be considered to improve forecasting accuracy.

REFERENCES

1. C. Schweppe and J. Wildes, "Power system static state estimation, pt. I: Exact model," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-89, no.1, pp. 120 – 125, Jan. 1970.

2. C. Schweppe and D.B. Rom, "Power system static-state estimation, pt. II: Approximate model," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-89, no.1, pp. 125 – 130, Jan. 1970.

3. C. Schweppe, "Power system static state estimation, pt. III: Implementation*," IEEE Trans. Power Apparatus and Systems*, vol. PAS-89, no.1, pp. 130 – 135, Jan. 1970.

4. U.S.-Canada Power System Outage Task Force, "Final report on the august 14th, 2003 blackout in the United States and Canada: Causes and recommendations," Apr. 2004. Available at: https://www.ferc.gov/industries/electric/indus-act/reliability/blackout/ch1-3.pdf

5. National Conference of State Legislatures (NCSL), "State renewable portfolio standards," Aug. 2010 [Jun. 2014], Available at: http://www.ncsl.org/research/energy/state-renewable-portfolio-standards.aspx

6. S. Debs and R. E. Larson, "A dynamic state estimator for tracking the state of a power system," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-89, no.7, pp. 1670-1678, Sep./Oct. 1970.

7. S. Wang, W. Gao, and A. P. S. Meliopoulos, "An alternative method for power system dynamic state estimation based on unscented transform," *IEEE Trans. on Power Syst.*, vol. 27, no. 2, May 2012.

8. E. Ghahremani and I. Kamwa, "Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements," *IEEE Trans. on Power Syst.*, vol. 26, no. 4, Nov. 2011.

9. W. Miller and J. B. Lewis, "Dynamic state estimation in power systems," *IEEE Trans. on Automatic Control*, vol. 16, no. 6, Dec. 1971.

10. H. Modir and R. A. Schlueter, "A dynamic state estimator for dynamic security assessment," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-100, no. 11, Nov. 1981.

11. K. Gharban and B. J. Cory, "Non-linear dynamic power system state estimation," *IEEE Trans. on Power Syst.*, vol.1, no. 3, Aug. 1986.

12. J. Chang, G. N. Taranto, and J. H. Chow, "Dynamic state estimation using a nonlinear observer for optimal series-capacitor switching control," *Elec. Power & Energy Syst.*, vol. 19, no. 7, pp. 441-447, 1997.

13. A. Vahidnia, G. Ledwich, E. Palmer, and A. Ghosh, "Dynamic equivalent state estimation for multi-area power systems with synchronized phasor measurement units," *Elec. Power Syst. Res.*, vol. 96, pp.170–176, 2013.

14. L. Fan, Z. Miao, and Y. Wehbe, "Application of dynamic state and parameter estimation techniques on real-world data," *IEEE Trans. on Smart Grid*, vol.4, no.2, pp.1133-1141, Jun. 2013.

15. Z. Wang, E. B. Makram, G. K. Venayagamoorthy, and C. Ji, "Dynamic estimation of rotor angle deviation of a generator in multi-machine power systems," *Elec. Power Syst. Res.*, vol. 97, pp.1– 9, 2013.

16. M. B. Do Coutto Filho, J. C. Stacchini de Souza and R. S. Freund, "Forecasting-aided state estimation—part I: Panorama," *IEEE Trans. on Power Sys.*, vol. 24, no.4, pp. 1667-1677, Nov. 2009.

17. M. Leite da Silva, M. B. Do Coutto Filho, and J. M. C. Cantera, "An efficient dynamic state estimation including bad data processing," *IEEE Trans. Power Syst*., vol. 2, no. 4, pp. 1050–1058, Nov. 1987.

18. M. B. Do Coutto Filho, A. M. Leite da Silva, J. M. C. Cantera, and R. A. da Silva, "Information debugging for real-time power systems monitoring," in *Proc. of Inst. Elect. Eng. C*, vol. 136, no. 3, pp. 145–152, May 1989.

19. M. B. Do Coutto Filho, J. C. Stacchini de Souza, R. S. G. Mattos, and M. T. Schilling, "Revealing gross errors in critical measurements and sets via forecasting-aided state estimators," *Int. J. Elect. Power Syst. Res.*, vol. 57, no. 1, pp. 25–32, Jan. 2001.

20. P. Rousseaux, T. V. Cutsem, and T. D. Liacco, "Whither dynamic state estimation?," *Int. Journal of Elec. Power & Energy Syst*., vol. 12, no.2, pp.104-116, Apr. 1990.

21. M. B. Do Coutto Filho, J. C. Stacchini de Souza and R. S. Freund, "Forecasting-aided state estimation—part II: Implementation*," IEEE Trans. on Power Syst.*, vol. 24, no. 4, pp.1678–1685, Nov. 2009.

22. L. Su and C. N. Lu, "Interconnected network state estimation using randomly delayed measurements," *IEEE Trans. on Power Sys*., vol. 16, no. 4, Nov. 2001.

23. B. Sadecky and P. Neuman, "State estimation and short term load prediction as components of the power system security assessment," in *Proc. of* 11th *Power Syst. Computation Conf. (PSCC)*, vol. II, pp. 705–712, Avignon, France, Aug. 1993.

24. R. D. Masiello and F. C. Schweppe, "A tracking static state estimator," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-90, no. 3 Mar. 1971.

25. B. Stott, O. Alsac and A. J. Monticelli, "Security analysis and optimization," in *Proc. of the IEEE*, vol. 75, no. 12, Dec. 1987.

26. L. H. Fink and K. Carlson, "Operating under stress and strain," *IEEE Spectrum*, vol. 15, pp. 48-53, Mar 1978.

27. A. Monticelli, *State estimation in electric power system: A generalized approach*, Kluwe academic publishers, 1999.

28. A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation,* NY: Marcel Dekker, 2004.

29. A. G. Exposito, A. J. Conejo, C. Canizares, *Electric Energy Systems: Analysis and Operation*, CRC Press, Jul. 2008.

30. L. Mili, ECE 5714, Class Lecture, Topic: "Parametric estimation theory," Faculty of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, Spring, 2013.

31. S. Thorp, A. G. Phadke, and K. J. Karimi, "Real time voltage-phasor measurements for static state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 11, pp. 3098–3104, Nov. 1985.

32. M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE Trans. Power. Syst.*, vol. 21, no. 4, pp.1930 – 1937, Nov. 2006.

33. C. Y. Evrenosoglu, "ECE 6304 Power System State Estimation Lecture Notes,'' The Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, Spring 2012.

34. N. G. Bretas, J.B.A, London, "Network observability: the critical measurement identification using the symbolic Jacobian matrix," in *Proc. of Int. Conf. on Power System Technology (POWERCON '98)*, vol.2, pp. 1222-1226, China, 1998.

35. A. Tarali, A. Abur, "Bad data detection in two-stage state estimation using phasor measurements," in *Proc. of IEEE PES Int. Conf. & Exh. on Innovative Smart Grid Technology (ISGT Europe)*, Berlin, 2013.

36. N. G. Bretas, J.B.A, London, L. F. C. Alberto, A. S. Bretas, "A Topological approach to the identification of critical measurements in power-system state estimation," *IEEE Trans. on Circuits & Systems—I: Regular papers*, vol. 52, no. 1, Jan. 2005.

37. L. Mili, Th. Van Cutsem and M. Ribbens-Pavella, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation", *IEEE Trans, on Power Apparatus and Systems,* vol.PAS-103, no. ll, pp.3239-3252, 1984.

38. L. Mili and Th. Van Cutsem, "Implementation of HTI method in power system state estimation", *IEEE Trans. on Power .Systems*, vol.3, no.3, pp.887-893, Aug. 1988.

39. H. M. Merrill and F.C. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-90, no. 6, pp.2718-2725, Nov./Dec. 1971.

40. E. Handschin, F.C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power App. Syst*., vol. PAS-94, no. 2, pp.329-337, Mar./Apr. 1975.

41. D. M. Falcao, P.A. Cooke, and A. Brameller, "Power system tracking state estimation and bad data processing," *IEEE Trans. on Power Apparatus and Systems,* vol. PAS-101, no. 2 Feb. 1982.

42. M. R. Irving, R. C. Owen, and M. J. H. Sterling, "Power system state estimation using linear programming," in *Proc. of IEE*, vol. 125, no. 9, 1978.W.W. Kotiuga and M. Vidyasagar, "Bad data

rejection properties of weighted least absolute value techniques applied to static state estimation," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-101, no. 4 Apr. 1982.

43. W.W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weighted least absolute value techniques applied to static state estimation," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-101, no. 4 Apr. 1982.

44. A. Jain and N. R. Shivakumar, "Power system tracking and dynamic state estimation," in *Proc. of IEEE/PES Power Systems Conference Exposition* (*PSCE*), Mar. 2009.

45. K. Sinha, L. Roy, and H. N. P. Srivaetava," A new and fast tracking state estimator for multi-terminal DC/AC power systems," in *Proc. of 4th IEEE TENCON*, India, Nov. 1989.

46. E. Farantatos, G. K. Stefopoulos, G. J. Cokkinides, and A. P. Meliopoulos, "PMU-based dynamic state estimation for electric power systems," in *Proc. of IEEE/PES General Meeting*, Jul. 2009.

47. M. Leite da Silva, M. B. Filho, and J. F. Queiroz, "State forecasting in electric power systems," *IEE Proc. C, Gen., Trans & Dist.*, vol.130, no.5, pp.273, Sep. 1983.

48. P. Rousseaux, D. Mallieu, T. V. Cutsem, and M. Ribbens-Pavella, "Dynamic state prediction and hierarchical filtering for power system state estimation," *Automatica*, vol. 24, no. 5, pp. 595–618, Sep. 1988.

49. Nishiya, H. Takagi, J. Hasegawa, and T. Koike, "Dynamic state estimation for electric power systems—Introduction of a trend factor and detection of innovation processes," *Elect. Eng. Jpn.*, vol. 96, no. 5, pp. 79–87, May 1976.

50. K. R. Shih and S. J. Huang, "Application of a robust algorithm for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 17, no. 1, pp. 141–147, Feb. 2002.

51. J. M. Lin, S. J. Huang, and K. R. Shih, "Application of sliding surface-enhanced control for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 570–577, May 2003.

52. S. Makridakis and S.C Wheelwright, *Forecasting methods and applications*. NY: Wiley, 1998.

53. M. Vinod Kumar and S. C. Srivastava, "Power system state forecasting using artificial neural networks," in *Proc. of Electric Power Components and Systems*, vol.7, no. 6, pp. 653-664, 1999.

54. G. Durgaprasad and S.S. Thakur, "Robust dynamic state estimation of power systems based on M-Estimation and realistic modeling of system dynamics," *IEEE Trans. on Power Sys.*, vol. 13, no. 4, pp.1331-1336, Nov. 1998.

55. Z. Morvaj, "A mathematical model of an electric power system for dynamic state estimation," *Electric Power System Research*, vol.8, pp. 207– 217, May 1985.

56. A. Blood, B. H. Krogh, and M. D. Ilic, "Electric power system static state estimation through Kalman filtering and load forecasting," in *Proc. of IEEE/PES General Meeting*, Pittsburgh, PA, Jul. 2008.

57. K. Sinha and J. K. Mondal, "Dynamic state estimator using ANN based bus load prediction," *IEEE Trans. Power Syst*., vol. 14, no. 4, pp. 1219–1225, Nov. 1999.

58. M. Falcão and U. H. Bezerra, "Power system operating state forecasting for security analysis and applications," *Elect. Power & Energy Syst.*, vol. 13, no. 6, pp. 330–336, Dec. 1991.

59. M. Nejati, N. Amjady, and H. Zareipour, "A new stochastic search technique combined with scenario approach for dynamic state estimation of power systems," *IEEE Trans. on Power Sys*., vol. 27, no. 4, Nov. 2012.

60. K. Sinha and J. K. Mandal, "Hierarchical dynamic state estimation using a ANN-based dynamic load prediction," in *Proc. of . Inst. Elect. Eng. C*, vol. 146, no. 6, pp. 541–549, Nov. 1999.

61. S. J. Huang and J. M. Lin, "Enhancement of anomalous data mining in power system predicting-aided state estimation," *IEEE Trans. Power Syst*., vol. 19, no. 1, pp. 610–619, Feb. 2004.

62. J. K. Mandal, A. K. Sinha, and L. Roy, "Incorporating nonlinearities of measurement function in power system dynamic state estimation," in *Proc. of Inst. Elect. Eng. C*, vol. 142, no. 3, pp. 289–296, May 1995.

63. S. J. Huang and K. R. Shih, "Dynamic state estimation scheme including nonlinear measurement function considerations," in *Proc. of Inst. Elect. Eng. C*, vol. 149, no. 6, pp. 673–678, Nov. 2002.

64. A. Bahgat, M. M. F. Sakr, and A. R. El-Shafei, "Two level dynamic state estimator for electric power systems based on nonlinear transformation," in *Proc. of Inst. Elect. Eng. C*, vol. 136, no. 1, pp. 15–23, Jan. 1989.

65. G. Valverde and V. Terzija, "Unscented Kalman filter for power system dynamic state estimation," *IET Gen. Trans. Dist.*, vol. 5, Iss. 1, pp. 29–37, 2011.

66. H. Xue, Q. Jia, N. Wang, Z. Bo, H. Wang, and H. Ma," A dynamic state estimation method with PMU and SCADA measurement for power systems," *International Power Engineering Conference (IPEC)*, pp.848-853, Dec. 2007.

67. J. Wood and B. F. Wollengerg, *Power Generation Operation and Control*, 2nd ed. NY: Wiley-InterScience, 1996.

68. P. Hines and S. Blumsack, "A centrality measure for electrical networks," in *Proc. of the 41st Annual Hawaii International Conf. on Sys. Sci.*, Big Island, Hawaii, pp. 185, Jan. 2008.

69. R. D. Christie. Power systems test case archive. University of Washington, 1961-1993. http://www.ee.washington.edu/research/pstca/.

70. S. Pahwa, "Dynamics on complex networks with application to power grids," PhD. Dissertation, Dept. of. Elec. & Comp. Eng., Kansas State University, Manhattan, Kansas, 2013.

71. E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure*," IEEE Systems Journal*, vol. 6, no. 4, Dec. 2012.

72. R. E. Kooij Y. Koc, M. Warnier and F. M. T. Brazier, "An entropy-based metric to quantify the robustness of power grids against cascading failures," *Safety Science*, 59:126, 2013.

73. S. Pahwa, D. Weerasinghe, C. Scoglio, and R. Miller, "A complex networks approach for sizing and siting of distributed generators in the distribution system," In *Proc. of IEEE/PES NAPS*, Sep. 2013.

74. NIST/SEMATECH, e-Handbook of Statistical Methods, http://itl.nist.gov/div898/handbook/, Available online, June 2014.

75. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control*, 3rd ed. Prentice Hall, Englewood Clifs, NJ, 1994.

76. C. Chatfield, *The Analysis of Time Series*, 5th ed., Chapman & Hall, New York, NY. 1996.

77. D. Hamilton, *Time Series Analysis*, Princton, NJ: Princton University Press, 1994.

78. P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd. ed., Springer-Verlang, 2002.

79. Y. Chakhchoukh, V. Vittal, and G.T. Heydt, "PMU based state estimation by integrating correlation," *IEEE Trans. on Power Syst*. vol. pp, no. 99, Oct., 2013.

80. M. Hassanzadeh and C.Y. Evrenosoğlu, "A regression analysis based state transition model for power system dynamic state estimation," in *Proc. of IEEE/PES North American Power Symposium*, Boston, MA, Aug. 2011.

81. M. Hassanzadeh and C.Y. Evrenosoğlu, "Power system state forecasting using regression analysis," *IEEE*/*PES General Meeting*, San Diego, CA, Jul. 2012.

82. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice Hall PTR, Englewood Cliffs, NJ, 1995.

83. Y. Chakhchoukh, P. Panciatici, and L. Mili,"Electric load forecasting based on statistical robust methods," *IEEE Trans. on Power Sys*., vol. 26, no. 3, Aug. 2011.

84. M. Hassanzadeh and C. Y. Evrenosoğlu, "Use of PMUs in regression-based power system dynamic state estimation," *IEEE*/*PES North American Power Symposium*, Urbana, IL, Sep 2012.

85. H. Wu, and J. Giri, "PMU impact on state estimation reliability for improved grid security*,"* in *Proc. of IEEE/ PES Trans. & Dist. Conf. & Exh.*, Dallas, TX, May 2006.

86. Bonneville Power Administration, *Wind Generation & Total Load in the BPA Balancing Authority*, http://transmission.bpa.gov/business/operations/wind, 2013.

87. E. Caro, J. M. Morales, A. J. Conejo, and R. Mínguez, "Calculation of measurement correlations using point estimate," *IEEE Trans on Power Delivery*, vol. 25, no. 4, Oct. 2010.

88. E. Caro, A. J. Conejo, R. Mínguez, M. Zima, and G. Andersson, "Multiple bad data identification considering measurement dependencies," *IEEE Trans. on Power Systems*, vol. 26, no. 4, Nov. 2011.

89. C. Hill, D. McMillan, K.R.W. Bell, and D. Infield, “Application of auto-regressive models to UK wind speed data for power system impact studies,” *IEEE Trans. on Sustainable Energy,* vol. 3, Issue: 1, Jan. 2012.

90. L. Zhao, and A. Abur, “Multi-area state estimation using synchronized phasor measurements,” *IEEE Trans. on Power Sys*., vol. 20, no. 2, May 2005.

APPENDIX A

Table A.1. Modifications to IEEE 118-Bus

Type	Bus No	Original [MW]	Modified [MW]
Load	3	39	0
Load	6	52	0
Load	11	70	0
Load	13	34	0
Load	15	90	0
Load	18	60	0
Load	27	71	0
Load	29	24	0
Load	32	59	0
Load	34	59	0
Load	35	33	0
Load	45	53	0
Load	46	28	0
Load	49	87	0
Load	54	113	0
Load	59	277	77
Load	60	78	0
Load	62	77	0
Load	66	39	0
Load	70	66	0
Load	74	68	0
Load	75	47	0
Load	76	68	0
Load	78	71	0
Load	79	39	0
Load	80	130	0
Load	82	54	0
Load	88	48	0
Load	90	163	0
Load	95	42	0
Load	100	37	0
Load	103	23	0
Load	106	43	0
Load	111	39	0
Load	112	68	0
Load	116	184	0
Gen.	10	450	350
Gen.	25	220	120
Gen.	49	204	0
Gen.	54	48	0
Gen.	59	155	0
Gen.	61	160	0
Gen.	65	391	291
Gen.	66	392	92
Gen.	80	477	0
Gen.	87	4	0
Gen.	89	607	0
Gen.	100	252	152

Gen.	103	40	0
Gen.	111	36	0

Table A.2. Modifications to IEEE 300-Bus

From Bus	To Bus	Original X [pu]	Modified X [pu]
41	92	0.1180	0.5
89	92	0.2740	0.5
40	68	0.0420	0.5
62	73	0.0270	0.5
62	240	0.2013	0.5
29	60	0.0600	0.5
29	63	0.0140	0.5
29	64	0.0290	0.5
67	190	0.1070	0.1
180	183	0.1830	0.1
180	57	0.0980	0.7
57	190	0.2320	0.1
57	66	0.1570	0.2
76	78	0.2530	0.05
76	79	0.0770	0.05
74	76	0.1190	0.05
66	190	0.3750	0.1
67	190	0.1070	0.1
176	190	0.0400	0.1
177	190	0.1090	0.1
182	190	0.0590	0.1
181	190	0.1280	0.1
61	66	0.0520	0.2
65	66	0.0220	0.2
78	79	0.4270	0.5
79	82	0.0530	0.1
79	83	0.0920	0.1
79	84	0.1220	0.1
41	61	0.1450	0.05
59	61	0.0990	0.05
65	66	0.0220	0.08
65	69	0.0360	0.2
64	65	0.0060	0.3
74	76	0.1190	0.2
64	67	0.0700	0.3
64	239	0.2676	0.3
64	241	0.2127	0.3