Real Time Dynamic State Estimation: Development and Application to Power System

by

SAFOAN AL-HALALI

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2018

© SAFOAN AL-HALALI 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Hossam Gaber
                     Professor

Supervisor: Ramadan El-Shatshat
                     Lecturer

Internal Member: Kankar Bhattacharya
                     Professor

Internal Member: Sagar Naik
                     Professor

Internal-external Member: Ali Elkamel
                               Professor

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Since the state estimation algorithm has been firstly proposed, considerable research interest has been shown in adapting and applying the different versions of this algorithm to the power transmission systems. Those applications include power system state estimation (PSSE) and short-term operational planning. In the transmission level, state estimation offers various applications including, process monitoring and security monitoring. Recently, distribution systems experience a much higher level of variability and complexity due to the large increase in the penetration level of distributed energy resources (DER), such as distributed generation (DG), demand-responsive loads, and storage devices. The first step, for better situational awareness at the distribution level, is to adapt the most developed real time state estimation algorithm to distribution systems, including distribution system state estimation (DSSE). DSSE has an important role in the operation of the distribution systems.

Motivated by the increasing need for robust and accurate real time state estimators, capable of capturing the dynamics of system states and suitable for large-scale distribution networks with a lack of sensors, this thesis introduces a three state estimators based on a distributed approach. The first proposed estimator technique is the square root cubature Kalman filter (SCKF), which is the improved version of cubature Kalman filter (CKF). The second one is based on a combination of the particle filter (PF) and the SCKF, which yields a square root cubature particle filter (SCPF). This technique employs a PF with the proposal distribution provided by the SCKF. Additionally, a combination of PF and CKF, which yields a cubature particle filter (CPF) is proposed. Unlike the other types of filters, the PF is a non-Gaussian algorithm from which a true posterior distribution of the estimated states can be obtained. This permits the replacement of real measurements with pseudo-measurements and allows the calculation to be applied to large-scale networks with a high degree of nonlinearity. This research also provides a comparison study between the above mentioned algorithms and the latest algorithms available in the literature. To validate their robustness and accuracy, the proposed methods were tested and verified using a large range of customer loads with 50% uncertainty on a connected IEEE 123-bus system. Next, a developed foretasted aided state estimator is proposed. The foretasted aided state

estimator is needed to increase the immunization of the state estimator against the delay and loss of the real measurements, due to the sensors malfunction or communication failure. Moreover, due to the lack of measurements in the electrical distribution system, the pseudo-measurements are needed to insure the observability of the state estimator. Therefore, the very short term load forecasting algorithm that insures the observability and provides reliable backup data in case of sensor malfunction or communication failure is proposed. The proposed very short term load forecasting is based on the wavelet recurrent neural network (WRNN). The historical data used to train the RNN are decomposed into low-frequency, low-high frequency and high frequency components. The neural networks are trained using an extended Kalman filter (EKF) for the low frequency component and using a square root cubature Kalman filter (SCKF) for both low-high frequency and high frequency components. To estimate the system states, state estimation algorithm based SCKF is used. The results demonstrate the theoretical and practical advantages of the proposed methodology.

Finally, in recent years several cyber-attacks have been recorded against sensitive monitoring systems. Among them is the automatic generation control (AGC) system, a fundamental control system used in all power networks to keep the network frequency at its desired value and for maintaining tie line power exchanges at their scheduled values. Motivated by the increasing need for robust and safe operation of AGCs, this thesis introduces an attack resilient control scheme for the AGC system based on attack detection using real time state estimation. The proposed approach requires redundancy of sensors available at the transmission level in the power network and leverages recent results on attack detection using mixed integer linear programming (MILP). The proposed algorithm detects and identifies the sensors under attack in the presence of noise. The non-attacked sensors are then averaged and made available to the feedback controller. No assumptions about the nature of the attack signal are made. The proposed method is simulated using a large range of attack signals and uncertain sensors measurements.

All the proposed algorithms were implemented in MATLAB to verify their theoretical expectations.

Acknowledgments

All praise is due to Allah almighty, who is the source of all knowledge in this world and whose countless bounties have enabled me to complete this thesis successfully. I would not have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my wife and family. I would like to express my deepest gratitude to my advisor, Professor Ramadan El-Shatshat, for his guidance, care, and patience and for providing me with such a wonderful atmosphere for conducting research. In addition to my advisor, I also wish to thank the rest of my thesis committee: Professor Kankar Bhattacharya, Professor Ali Elkamel, and Professor Sagar Naik for their encouragement and insightful comments. I would like also to express my deepest gratitude to Professor Christopher Nielsen, for his care, and patience and for providing his valuable feedback all along the way. Special thanks also to all colleagues and graduate friends, especially the Distribution Engineering group for sharing the literature, support, and fruitful discussions. I also offer my gratitude to my father and my mother. My research would not have been possible without their help, their constant support and encouragement, and their never-failing good wishes. Finally, I would like to thank my wife Eman, my son Muftah, and my daughter May for always standing by me through both the good times and the bad.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ARE   Voltage Angles Relative Errors

MRE   Voltage Magnitudes Relative Errors

ACE   Area Control Error

AGC   Automatic Generation Control

AMI   Advanced Metering Infrastructure

AMRs   Automated Meter Readings

CKF   Cubature Kalman Filter

CPF   Cubature Particle Filter

DCS   Distributed Control System

DER   Distributed Energy Resources

DG   Distributed Generation

DSCADA   Distribution System Supervisory Control and Data Acquisition

DSE   Dynamic State Estimator

DSSE   Distribution System State Estimation

EKF   extended Kalman Filter

EM	Expected Maximization
EPF	Extended Particle Filter
ESD	Estimated Standard Deviation
EVs	Electrical Vehicles
GPS	Global Positioning System
KF	Kalman Filter
MAE	Mean Absolute Error
MAPE	Mean Average Percentage Error
MC	Monte Carlo
MILP	Mixed Integer Linear Programming
MM	Markov Model
NNs	Neural Networks
PF	Particle Filter
PI	Proportional-integral
PMU	Phasor Measurement Unit
PSO	Power System Operator
PSSE	Power System State Estimation
RESs	Renewable Energy Sources
RI	Relative Increment
RNN	Recurrent Neural Network

RTU	Remote Terminal Unit
SCADA	Supervisory Control and Data Acquisition
SCKF	Square Root Cubature Kalman Filter
SCPF	Square Root Cubature Particle Filter
SE	State Estimator
SIR	Sequential Importance Re-sampling
SIS	Sequential Importance Sampling
SM	Smart Meter
UKF	Unscented Kalman Filter
UPF	Unscented Particle Filter
VSTLF	Very Short Term Load Forecasting
WLS	Weighted Least Square
WNN	Wavelet Neural Network
WRNN	Wavelet Recurrent Neural Network

Chapter 1

Introduction

The electric power industry worldwide has undergone significant changes over the last decade. The rapidly growing demand for electric power relative to the available supply, the environmental impact of conventional plants and the fears of the depletion of conventional energy resources (oil and gas), have begun to push energy sources towards distributed generations (DGs), such as wind and solar [35]. In the province of Ontario in Canada, for example, the output power generated by these renewable energy resources will reach 10 GW by end of 2018 [108]. With the increased penetration of DGs into the distribution system, reliable system operation and control has become increasingly difficult as the power flow no longer remains unidirectional. Moreover, the random variations in power produced by weather-based generation cause different levels of uncertainty that affect a variety of decisions. This stochastic nature of the sources may results in unpredictable power flows, greater voltage fluctuations, and different network reactive power characteristics [59]. Therefore, advanced real-time monitoring, and control are needed in order to operate and control the system efficiently and to ensure high reliability and safety.

An accurate real time estimation of the distribution system states is the first step towards real-time monitoring, proper system assessment, and efficient control actions. A distribution system state estimation (DSSE) is a systematic procedure for processing a set of real and pseudo-measurements in order to estimate bus voltages and angles and, hence, the current state of a system. State estimator is composed of two stages: prediction and filter-

ing stage. The first stage predicts the system state before the measurements arrived and the second one is to filter out the predicted values upon the receiving of the measurements. Once the information from the state estimator are received, and with the knowledge of the system components, the second step is to assess the situation of the distribution system and to provide the operator with knowledge about the state of the system ahead of time in order to take proper control actions.

As another type of application of the real time state estimation in the transmission level, is the cyber security of the automatic generation control (AGC) system in either single-area or multi-area AGC system. The AGC system is a fundamental control system used in all power networks to keep the network frequency at its nominal value (50 or 60 Hz), and maintaining tie-line power exchanges at their scheduled values. Several cyber-attack incidents incidents in power networks appears in [114, 42]. Since the AGC is a sensitive component in the power system and is vulnerable to the cyber attack, there is a need to protect this valuable component from the cyber attacks.

1.1 Motivation

State estimation is an essential stage in the monitoring and assessment of electrical distribution networks. In light of todays large-scale distribution systems characterized by limited information coupled with a high degree of fluctuation and uncertainty related to generation and demand, the development of an efficient DSSE has become a necessity. The literature has reported two schemes for estimating the states of a distribution system: centralized and decentralized. In centralized schemes, all measurement data must be sent to a central unit for processing in order to obtain the estimated states. However, the determination of the states over the whole distribution system has become challenging and difficult because of the large size of todays systems, which are expected to expand even further with the anticipated extensive proliferation of renewable energy sources (RESs) and electrical vehicles (EVs). In the second, decentralized, type of scheme, the entire distribution system is divided into a number of zones, each with its own estimator that coordinates with other zone estimators. In view of the ongoing development of the smart grid, more research on state estimator algorithms is needed to meet the challenges that smart grid functionalities

present. Among others, environmental compliance, reliability, and security will impose additional constraints on the state estimation and require improved performance in terms of response time and robustness [28]. In order to meet these constraints, both the prediction and the filtering stages need to be further improved. For the filter stage, most of the state estimator algorithms available in the literature are based on the static approach, which has been proved not compatible especially because of the load fluctuation. The rest of the filtering stages are based on approximations that lead to inaccurate estimation. Moreover, most of the existing state estimation algorithms are either based on the Kalman filter (KF) or the weighted least square principle. Neither these two approaches are capable of coping with the nonlinear dynamics of the distribution system, nor can they deal with the non-Gaussian distribution of the pseudo-measurements. Therefore, proposing a state estimator that can tackle these issues is crucial in order to accomplish the first step of comprehensive monitoring of the distribution system. For the prediction stage, most of the algorithms are based on the regression, or forecasting-based method. Moreover, to have a robust state estimator, not only the filtering algorithm needs to be improved but also the predictor needs to be enhanced. In one hand, improvement of the prediction stage has a huge impact on the accuracy of the state estimator, and on the other hand it can give the operator a quick view about the situation in the distribution system. From the above, it seems that there is a need for a robust real time state estimator that can cope with severe load fluctuations and the huge uncertainty in the pseudo-measurements, loss/delay/ lack of measurements.

In recent years, several cyber-attack incidents have been reported [53]. A detailed survey of different cyber-attack incidents was provided in [102] and a detailed elaboration on cyber-attack incidents in power networks appears in [114, 42]. Little work has been conducted with respect to attack resilient measures that are used to detect, identify, and mitigate corrupted real-time measurements in the feedback loop of AGC. The accuracy and reliability of real-time measurements has a significant impact on system's real-time operation. In smart power grids, real-time measurements for AGC are transmitted using computer networks [104]. These computer networks might be an attractive space for cyber-attackers, e.g., disgruntled employees, insiders, nation states or terrorist organizations [101]. Through the computer network, the attacker can also learn the parameters

of the system using the algorithm shown in [104]. The compromised measurements may lead to very severe and adverse effects on the management and control of a smart grid. For example, the corrupted measurements can cause a rapid decline in the system frequency that leads to trigger load shedding schemes or generators disconnecting. This is because the compromised measurements can mislead the AGC, which requires minimal supervision and intervention by human operators, to increase generation through a false impression of increasing demand. The imbalance between power generation and demand leads to deviation of the grid frequency from its nominal value. The channels through which the attacker can corrupt sensor measurements include the physical sensors, sensor data communication links, and data processing programs at the control center. Accessing and attacking geographically distributed physical sensors is tedious and hard to coordinate. However, hacking a computer program at the strongly protected control center is not impossible [53], [102]. From the above, it seems that there is a need for an algorithm based real time state estimator for mitigation of cyber-physical attacks in multi-area automatic generation control.

1.2   Problem Statement and Research Objective

The proposed research broadly aims at incorporating the concept of the intelligent distribution system into operations and control. With an accurate and timely state estimation result, we could monitor system performance, and diagnose malfunctions in both the monitoring system and the grid. Therefore, the state estimator plays an important role in reliable and secure intelligent distribution system operations. The proposed state estimator may operate under a lack of measurements due to communication failures or delays. Moreover, due to the limited number of measurements, the proposed estimator should deal with high uncertainty and in all above circumstances should provide accurate information about the distribution system states. During operations under any configuration, the objectives and constraints should be taken into consideration. For example, a low computational time has to be a main goal. As a result, on-line situation assessment can be achieved, and based on that, the system performance can be enhanced. The key tasks in

the proposed research are as follows:

1. To develop a real time state estimation algorithms for the electrical distribution system without any approximation in terms of the error distribution function.

2. To develop a very short term load forecaster for the forecast aided state estimator.

3. To design an attack resilient control scheme based real time state estimator for on both single and multi area AGC systems.

1.3   Organization of the Thesis

This thesis is organized as follows:

1. Chapter 2 reviews the literature on discrete-time filters applied in the electrical distribution system state estimation. We also provide a review of the literature on the prediction algorithm applied in the electrical distribution system state estimation. Moreover, we review the literature on cyber security algorithms used in automatic generation controls.

2. In Chapter 3, we introduce three filtering algorithms which are the square root cubature Kalman filter (SCKF), the cubature particle filter (CPF), and the square root cubature particle filter (SCPF) for the area of distribution system state estimation. We utilize the μphasor measurement unit to estimate the distribution system states. The heuristic test load profile is generated based on the expected load shape due to the distributed generation (DG) proliferation. The sampling time (T) is assumed to be 5 sec.

3. In Chapter 4, pseudo measurement generator based very short term load forecasting (VSTLF) is developed for distribution system state estimation. The test load profile is generated based on real data. The sampling time (T) is assumed to be 5 min.

4. In Chapter 5, an algorithm based state estimator is developed to detect a cyber attack on both single and multi area AGC systems, and identify the sensors under attack. A simple switching technique is proposed to mitigate the effect of the cyber attack allowing the controller to make the right decision.

5. Chapter 6 is divided into two parts: in the first part, a number of key attributes of the thesis are summarized. The second part outlines a few interesting research topics we would like to develop in the future.

Chapter 2

Background and Literature Review

This literature review is focused on the stages related to main state estimator (SE) architecture schemes and algorithms, with particular emphasis on the stages involved in active distribution systems. The primary components associated with a state estimator are first clarified, and the decentralized algorithm schemes for both power and distribution systems are then explained. The research related to mathematical modeling of the prediction and filtering stages for both power systems and active distribution systems is then highlighted. Studies that introduce techniques for generating pseudo-measurements with respect to distribution systems are also reviewed. The main focus in this research is the prediction of the voltage and angle in each bus in order to facilitate advance warning prior to failure, thus providing the opportunity for optimal decisions. Most state estimators described in the literature are employed at the transmission level, whose features, in contrast to those of the distribution level [23], include a radial or tree-like topology; a large R/X ratio of cables; unbalanced loading; asymmetrical construction; a small number of real-time measurements characterized by limited reliability (for both analog and digital); limited accuracy with respect to the load data obtained from the load curves; automated meter readings (AMR) or advanced metering infrastructure (AMI); and discrepancies in real-time measurements, load data, and static network data such as line impedances or transformer ratings. In contrast to those in a power system, state estimators in a distribution system need additional investigation because of their limited application and because of the new challenges

presented by the requirements anticipated for the grid of the future [94]. State estimators in active distribution systems perform poorly because of sparse measurements, but placing a measurement device at each node is economically inviable.

2.1 State Estimator Components

To provide essential background information about SE components, the conventional components used in traditional distribution systems are first introduced. New components that represent technological innovations associated with the smart grid concept are then briefly analyzed from the perspective of their influence on the conceptual design of future SE architectures [111].

2.1.1 Remote Terminal Units

A remote terminal unit (RTU) is an electronic device controlled by a microprocessor. The device interfaces between physical objects and a distributed control system (DCS) or a supervisory control and data acquisition (SCADA) system by transmitting telemetry data to the system.

2.1.2 Phasor Measurement Unit

As defined in [90], a phasor measurement unit (PMU) is an instrument that measures voltage and/or current waveforms at the point of connection to the grid using synchronized sampling based on a common time reference for all locations, commonly facilitated through a Global Positioning System (GPS) satellite time reference.

2.1.3 Smart Meter

According to [111], smart meter (SM) systems are becoming increasingly available for medium- and low-voltage distribution grids. In this context, the term smart meter is

used for identifying either single metering devices or data concentrators, which collect and process the active or reactive power measurements provided by the aggregated loads at distribution feeders or secondary substations.

2.2 Background

State estimation is an essential stage in the monitoring and assessment of electrical transmission and distribution networks. It represents a systematic procedure for processing a set of real and pseudo-measurements in order to estimate bus voltages and angles and, hence, the current state of a power system. This section provides a review of the literature related to both centralized and decentralized SE architecture schemes and to both prediction and filtering algorithms. The review begins with a summary of reports that discuss SE architecture schemes at the transmission level, where they were originally introduced, and those in distribution networks, along with SE algorithms used in both transmission and distribution networks. Fig. 2.1 illustrates the state of the art as well as the organization of the literature review with respect to both the transmission and distribution levels. State estimators were originally established at the transmission level to insure the stability of a power system. However, SEs have several additional applications, such as fault detection (rather than protection), isolation, and reconfiguration, all with the goals of enhancing reliability, optimizing and rescheduling generation based on the estimated states, and the ability to compare estimated flows with voltage rating or switching conditions in order to prevent overloading.

2.3 State Estimator Architecture

The literature includes reports of two schemes for estimating the states of a transmission and distribution system: centralized and decentralized. In centralized schemes, all measurement data must be sent to a central unit for processing in order to obtain the estimated states. However, the determination of the states over the whole distribution or transmission system has become increasingly challenging because of the large size of todays systems,
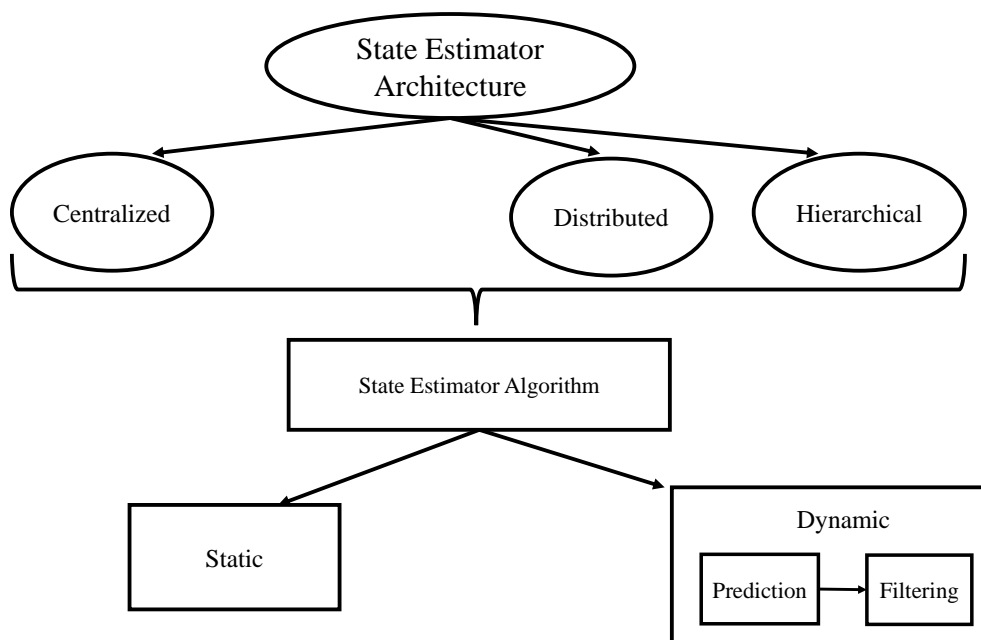
Figure 2.1: State of art with respect to state estimators

Central Unit

Zone 1

Zone n

Zone 2

Zone 1

Zone n

Zone 2

(a)

(b)
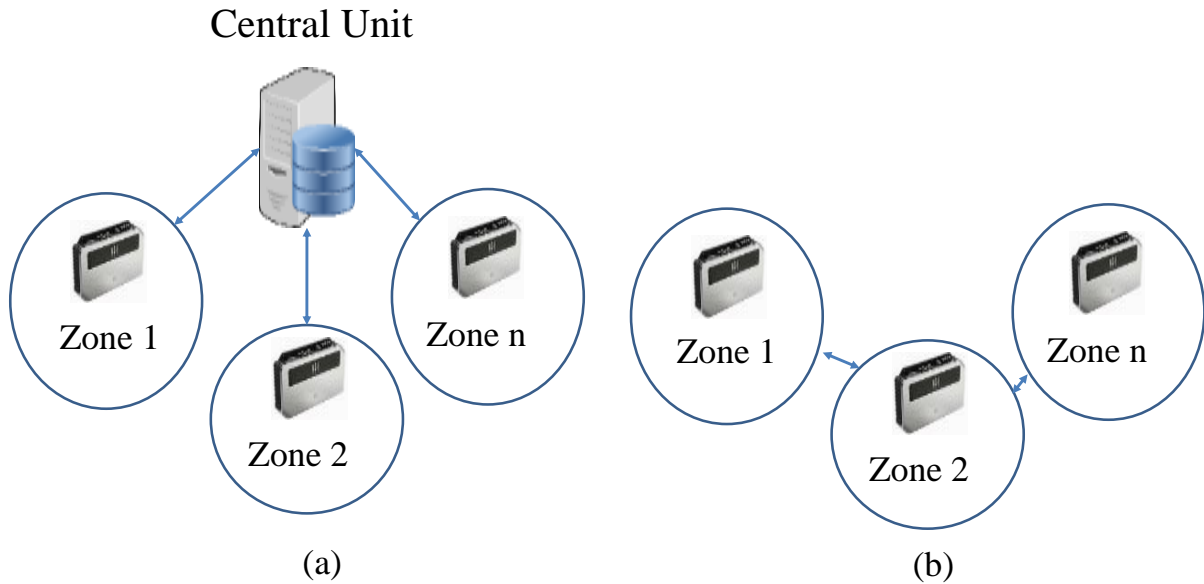
Figure 2.2: Decentralized schemes: (a) Hierarchical; (b) Distributed

which are expected to expand even further with the anticipated extensive proliferation of renewable energy sources (RESs) and electric vehicles (EVs). In the second, decentralized, type of scheme, the entire system is divided into a number of zones. Researchers have proposed several decentralized schemes based on hierarchical and distributed solution architectures, as shown in Fig. 2.2(a) and (b), respectively. This literature review addresses decentralized schemes based on distributed architectures because this architecture has been chosen for the proposed SE based on its avoidance of any kind of centralization (i.e., sending information to a single unit).

2.4 Decentralized Hierarchical Architecture

In the hierarchical architecture shown in Fig. 2.2(a), the central unit distributes the work among slave processors, each of which is responsible for its own zone states. The central

unit then coordinates the local estimates. In this scheme, distributed processors, which can be located either remotely (distributed architecture) or in the same physical location (parallel architecture), communicate only with the central computer. A detailed survey of hierarchical SE approaches has been provided by Gomez-Exposito et al. [37].

2.5   Decentralized Distributed Architecture

As shown in Fig. 2.2(b), in decentralized distributed SE architecture, the coordination phase performed by the central computer is avoided, so no central computer is needed. Instead, each local processor that is used for estimating its zone states communicates only with neighboring processors, exchanging border information at each iteration. As stated in [38], decentralized distributed architecture is highly dependent on the algorithm used in the estimator. Therefore, as can be concluded from the following studies, all research has been based on the interaction principle. With the introduction of the proposed new estimator algorithm, additional elaboration is required with respect to both the data that must be exchanged between the zones and the number of iterations needed. The taxonomy of multi-area state estimation algorithms is addressed in [38]. In [90], [49], [120], [15], [50], PMUs were utilized for coordinating the voltage angles of the SE solution for each subsystem. The results based on PMUs are quite accurate, specifically with respect to angles. Recently, SE based compressed sensing [2], [66] is proposed in the literature. However, the compressed sensing theory is based on a linear transformation which does not fit in the future smart grid.

2.6   State Estimator Algorithms

Since the method's first proposal in 1970 [88], many studies have addressed the method's inherent problems and treated them in centralized and decentralized schemes. The core algorithm of most of the proposed SE schemes, whether centralized or decentralized, is based on a weighted least square (WLS) [77], [74], [113], [81] [37], [45], [116] and [39] or the Kalman filter (KF) [33] [61]. The WLS methods are computationally efficient, can be

12

implemented easily, and converge quickly. However, it is a non-robust estimator, as any fluctuations in power generation and customers' loads in systems with a high penetration of renewable resources and EVs can severely distort the estimation results [109]. The use of WLS, both at the transmission level [11], [95], [73] and at the distribution level [82], [83], is based on the minimization of the following objective:

$$J = [z - h(x)]^T R_z^{-1} [z - h(x)] \tag{2.1}$$

where z and x are the vectors of the measurements and the state components, respectively. These vectors are assumed to be related according to the equation $z = h(x) + e_z$, in which $h(x)$ is a known function. The measurement error vector is assumed to be a Gaussian random variable with a diagonal covariance matrix $R_z = diag[\sigma_{z1}^2, \sigma_{z2}^2, \sigma_{z3}^2, \cdots, \sigma_{zm}^2]$. Thus, σ_{zi}^2 is the variance of the ith measurement. The minimizer $\hat{x}$ for the WLS objective is obtained as the limit of a sequence of states by means of a Newton-based recursive scheme, one step of which is,

$$\hat{x}(k + \Delta k) = \hat{x}(k) + (H^\top(\hat{x}(k))R_z^{-1}H(\hat{x}(k)))^{-1}H^\top(\hat{x}(k))R_z^{-1}[z - h(\hat{x}(k))] \tag{2.2}$$

An estimate of the state error covariance matrix $\hat{P}(\hat{x}(k))$ at $\hat{x}(k)$ is given by,

$$\hat{P}(\hat{x}(k)) = (H^\top(\hat{x}(k))R_z^{-1}H(\hat{x}(k)))^{-1} \tag{2.3}$$

where $H(\hat{x}(k))$ Jacobian Matrix evaluated at $\hat{x}(k)$;
$\hat{x}(k + \Delta k)$ optimal estimate of $x(k + 1)$, given observations $z(k)$ for sample k;
$\hat{P}(\hat{x}(k))$ positive definite symmetric matrix of order $n \times n$, representing the error covariance matrix for the state estimate at sample k. The authors of [97] demonstrated that the consistency of WLS state estimation relies on the assumptions that $P(k + \Delta k)$ and that the measurements are distributed normally. These assumptions result in the introduction of greater approximation, which leads to divergence because the gain matrix and the covariance matrix are used recursively to calculate the estimated states. WLS-based methods are also considered inefficient [72], especially with respect to the penetration of renewable power resources.

Thus, Kalman filtering techniques are proposed to address and estimate the power system's dynamic states. The authors of [8], [76], [10] estimated system states using

13

a KF, which is considered the optimal filter to use with linear models and Gaussian noise [52], [46], [95]. Dynamic state estimator (DSE) methods have been established as a means of addressing severe power fluctuations in the generation and load. In discrete-time format, their system model can be represented in state-space form as,

$$\hat{x}(k + \Delta k) = ax(k) + W(k)$$
$$z(k + \Delta k) = h(\hat{x}(k + \Delta k)) + V(k)$$

(2.4)

where x state vector of order $n \times 1$;

a $n \times n$ state transition matrix;

h $1 \times n$ observation vector;

W $n \times 1$ vector of white random noise with a covariance matrix Q;

V white random scalar with a variance R uncorrelated with W;

z observed output;

k kth time sample.

A Kalman estimator is implemented using the following recursive equations [55]:

$$K(k) = P(k)(H(\hat{x}(k + \Delta k)))^T[H(\hat{x}(k + \Delta k))P(K)(H(\hat{x}(k + \Delta k)))^T + R]^{-1}$$
$$P(k + \Delta k) = a(P(k) - K(k)H(\hat{x}(k + \Delta k))P(k))a^T + Q$$
$$\bar{x}(k + \Delta k) = \hat{x}(k + \Delta k) + K(k)[z(k + \Delta k) - h(\hat{x}(k + \Delta k))])$$

(2.5)

where $K(k)$ Kalman gain vector at sample k; $\hat{x}(k + \Delta k)$ the predicted states.

However, distribution systems usually behave in a nonlinear manner, and the use of a KF can lead to inferior estimates of system states. In addition, the variance in the KF results for the estimated states must be represented as a normal distribution by $P(k + \Delta k)$ in order to avoid an inappropriate condition for calculating the gain $K(k)$. Because a KF relies on the least square principle, it has been shown [97], [46] that the consistency of a KF state estimation is based on the assumptions that $P(k + \Delta k)$ and that the measurements are normally distributed. As with the WLS technique, these assumptions lead to the introduction of increased approximation, which creates divergence because of the use of the gain matrix and the covariance matrix for the recursive calculation of the estimated states [110], [36].

14

The authors of [94], [100], [75], [34], [103] introduced an extended Kalman filter (EKF) as a means of handling nonlinearity; only the linear expansion terms of a Taylor series are used [34], and the variations in the process and measurement noise are assumed to be Gaussian with a zero mean. Like a KF, an EKF relies on the least square principle, and its state estimation consistency [97], [46] has also been shown to be based on the assumptions that $P(k + \Delta k)$ and that the measurements are normally distributed. Again, greater approximation is introduced, causing divergence since the gain matrix and the covariance matrix are used recursively for calculating the estimated states [110]. An EKF can also result in inaccurate estimations if the system behaves in a highly nonlinear manner. The primary difference between an EKF and a KF is that the value of $\hat{x}(k + \Delta k)$ is used in equation 2.5, which has been derived from the Taylor series of the system equations as a means of updating the estimated value iteratively until the difference between two consecutive estimates equals a very small value. Despite these advances, as with a regular KF, an EKF can also result in inaccurate estimations if the the measurements have a non-Gaussian distribution. Other researchers [109], [84], [14] proposed an unscented Kalman filter (UKF). Unlike an EKF, a UKF provides an enhanced approximates of the state random variable distribution [110]. Recently, a powerful dynamic filter, known as the cubature Kalman filter (CKF), has been added to the KF family [5], [4], and the CKF was introduced for power system state estimation by [92], [93]. Unlike a UKF, a CKF uses a more accurate representation of the state random variable distribution and can work efficiently in a high-dimensional space [5]; however it still represents the posterior state distribution using a Gaussian random variable. The distribution systems' lack or sparseness of measurement sensors makes the use of pseudo-measurements crucial for ensuring system observability. However, pseudo-measurements generate a high degree of uncertainty and a non-Gaussian distribution [67]. All of the above methods (KF, EKF, WLS, UKF, and CKF) are used to estimate the voltage states, assuming a Gaussian representation of the pseudo-measurements. This introduces an additional error due to the approximation of the non-Gaussian distribution function, as the error associated with the pseudo-measurements has a non-Gaussian distribution [67], [63]. Moreover, WLS- and KF-based techniques provide estimated states with a Gaussian distribution, which introduces another source of approximation. The above two sources of approximation may lead the estimator to diverge.

15

A detailed comparative review of a variety of these techniques can be found in [97].

To overcome the drawbacks of high levels of uncertainty, PF-based algorithms have been proposed in [3], [7], [87], [56], [71], [57], [13], [26], [25]. A primary benefit of the PF is that it handles the non-Gaussian measurements [85], [110]. Although the principle of PF has been applied to efficiently estimate the states of a distribution system, the test system used to validate the proposed algorithm is relatively small, containing only 14 buses. Moreover, several issues related to distribution systems have not been included, such as the effect of dynamics of a system, distributed generation (DG), high nonlinearity, and pseudo-measurements.

2.7   Pseudo Measurement Generator

There are two approaches in the literature to predict the states of the electrical distribution system [21],[20]. The first one, achieved by modeling the states dynamic (the evolution of the states over the time), and the second is accomplished by forecasting the load and then predicting the future states through either static or dynamic SE. For the first one, since the method was first proposed in [19], there are several models that have been used to predict the state dynamics of the distribution system. The method proposed in [19] is based on the quasi steady state behavior and slow dynamics of the power system. To overcome the difficulties of modeling the time behavior of the system, it has been assumed that the next state will be the same as the previous state, except for some uncertainties pointed out by Debs et al. [19]. By solving the power flow equation of the system, given the power at each bus, the corresponding time behavior can be obtained for the conventional load profile. So, as in [19], the value of the maximum rate of the change vector has been calculated. In [16], [109] a nonlinear algorithms, was used to model the state evolution of the distribution system without any linearization, which increases the accuracy of the state estimator. The following equation has been used:

$$x_{k+\Delta k} = F_k x_k + g_k + q_k \tag{2.6}$$

where, x_k is the state vector comprising the magnitudes and angles of the nodal voltages. Matrix F_k and vector g_k describe the transition process of the states, and q_k is the white

Gaussian noise vector of the prediction model at time k. With the help of the Holts exponential smoothing, the accuracy of the parameters have been increased. In [22], [43], [44], the impact of the neighbor buses have been studied and included in the model parameters. In [21], [99] the artificial neural network was trained using only information by monitoring the power system for a short amount of time. Livani et al. [64] used historical data about load and wind generation in order to train a Markov model (MM), which describes the evolution of the states over a specific amount of time. The historical data was not enough to train the model, however, a Monte Carlo simulation was used in order to generate enough data. Because the data has some uncertainties in it, a state estimator has been used in order to obtain filtered predicted states. Most of the effort done to model the dynamic of the system state is based on historical and forecast data. Therefore, once an efficient forecasting algorithm is used an efficient transition model of the state can be obtained. Most of the forecasting algorithms available in the literature are based on regression based methods [40], including persistence [32], extrapolation [30], [65], [121] , [118], time series [17], [62], [89], [105], Kalman filters [107], [51], fuzzy logic [18], [117], and neural networks (NNs) [91], [12], [47]. A summary of the most commonly-used methods for load estimation is provided in [21]. Recently, an algorithm of the wavlet neural network (WNN) has been developed by [41] to forecast the load of ISO New England power system by also using it's data set.

2.8 Cyper Security in AGC system

The work in [104] assumes that the frequency sensor is well protected and builds their detection model based on a comparison between the prediction of frequency and the measured one. Therefore, in case of losing the frequency measurement the detection algorithm will be no longer functioning, leaving great chance for the attacker to manipulate the system. Many of the proposed detection and identification algorithms for security of AGCs [102, 101, 6, 9, 106] use limited attack templates that cannot characterize real-world attackers well. In [101] the author proposed a statistical method based on maximizing the likelihood of detecting the attacked sensor. However, many of the statistical methods [58, 68, 69], including [101], have the problem of false positives and false negatives.

The work of [101] has reported 5 % false positives and false negatives. The proposed mitigation strategy in [104, 101] is, upon detecting an attack, to replace the sensor data with forecasting data. Such a strategy can cause large deviation from the nominal frequency due to forecasting error. The work in [104] identifies the sensors under attack after 20 seconds from the onset of the attack. The mitigation algorithm proposed in [104] is based on neglecting the measurements from the sensors under attack. Therefore, the attacker can manipulate the AGC system during the 20 second period, causing at least some profit losses to the owner or even triggering the remedial action. The work in [80], [29] have formulated the state estimation problem using mixed integer linear programming (MILP). However, the authors assume that the system is noiseless. No mitigation solution has been proposed in [80], [29], [27], [78].

Chapter 3

Electrical Distribution System State Estimation For Real Time Operation

3.1 Introduction

The goal of this chapter is to introduce new SEs that can handle the challenges of future distribution grids with higher accuracy. These challenges include the following: the sparseness or lack of sensors due to economic restrictions, the non-Gaussian representation of estimated states and pseudo-measurements, and the high degree of fluctuation in generation and customer loads. The proposed estimators, which are the cubature particle filter (CPF) and square root cubature particle filter (SCPF), address the above challenges and the limitations of the existing tools, and solves the problem in a distributed manner leveraging the distributed algorithm proposed in [74]. The SCPF and CPF techniques are still in its infancy, even in signal processing, and only few publications are available in that area. To the best of the authors' knowledge, only [55] and [54] have implemented the CPF and SCPF in different applications, which are not related to power system state estimations. We also introduce the square root cubature Kalman filter (SCKF), which is the improved version of the CKF recently introduced for power system state estimation by [92], for the first time for the electrical distribution system state estimation. Due to their superiority over the other existing state estimation techniques, the SCKF, CPF, and

19

SCPF are proposed in this chapter. To the best of the authors' knowledge, the three proposed estimators have been introduced for the first time for state estimation applications in the power system area, for either transmission or distribution levels. The CPF and SCPF are capable of processing non-Gaussian representations of pseudo-measurements directly without Gaussian distribution approximation. Moreover, the proposed algorithms treat the problem in a distributed fashion; as a result, the estimators can work for large-scale electrical distribution systems while providing accurate estimations within a short period of time. The distribution system will be clustered according to the algorithm proposed in [74], and the state of each zone will be estimated in parallel. Additionally, this chapter provides a comprehensive evaluation of the proposed SCKF, CPF, and SCPF under normal and abnormal operations. This chapter also provides a comprehensive comparison study between the SCKF, CPF, SCPF and the latest versions of state estimation algorithms. The results show that the proposed SEs have a superior performance compared to the latest estimation algorithms (EKF, UKF, and CKF). The research investigates the performance of these estimators on an IEEE 123-bus test system.

3.2 Nomenclature

Variables

z	Measurement vector
$h(x)$	Nonlinear function (power flow equation relating the injected power to the voltage magnitudes and angles)
v	Measurement noise vector
x	Vector of the state variables
w_k	Particle weight vector of the states
$p\left(z_k | \bar{x}_{k|k}\right)$	Particle likelihood vector

$p\left(\bar{x}_{k|k}|\hat{x}_{k-1|k-1}\right)$ Particle prior estimate of the vector states

$q\left(\bar{x}_{k|k}|z_k, \ddot{x}_{k|k}\right)$ Particle proposal distribution vector of the states

I_k Prior evaluated at time instant k

$\jmath_k$ Likelihood evaluated at time instant k

$\imath_k$ Proposal distribution evaluated at time instant k

w_k Importance weight

$\hat{w}_k$ Normalized weight of each particle

$\hat{x}_{k|k}$ Estimated state of each particle

$\hat{S}_{k|k}$ Estimated variance of each estimated particle

3.3   Zoning

The proposed zoning method is a modified version of the zoning method proposed in [74]. First, consider a distribution system comprising n buses collected as a set of vertices $V :=$ $\{1, 2, \cdots, n\}$, and overhead or underground lines represented by a set of edges $E \subseteq V \times V$. The proposed solution is based on the following assumptions [74].

1. The distribution system is partitioned into $L > 1$ zones $ZO^l \subset V, l \in \{1, 2, \cdots L\}$ and $ZO^l \cap ZO^j = \varnothing$ based on topological or geographical criteria. To have similar execution times for individual zones, the number of buses per zone should be kept approximately equal;

2. If two zones are connected, then at least one of the two zones is extended to share at least one bus with the other zone. After extension, at least one of the shared buses between the two zones is assumed to be fully monitored through a suitable measurement point. This is just to maximize the number of real measurement per zone; and

3. Each extended zone has a local center time-synchronized through the global position-ing system (GPS) where the local estimation is carried out; such computing nodes work in parallel and communicate with the local center of adjacent areas.

Following Assumption 2, we sequentially extend each zone ZO^l to $\overline{ZO}^l$ according to: $\overline{ZO}^l = ZO^l \cup \{i|\,(i,j) \in E, i \in ZO^l, j \in ZO^k, l < k\}$.

The above zone extension procedure is demonstrated in the modified IEEE 123 bus system shown in Fig. 3.1. The IEEE 123 test system is modified by connecting DG at buses 17 and 79, which results in a non-Gaussian distribution of the pseudo-measurement in most of the buses [60]. Furthermore, several links are introduced in the system, highlighted in the dashed line in Fig. 3.1, to upgrade the operation into a mesh configuration, and hence, the power flow becomes bidirectional with high uncertainty. For example, the links between bus 41 and bus 95 and between bus 73 and bus 74 have been added to the original system. Fig. 3.1 shows the $L = 4$ original zones ZO^l and the extended zone $\overline{ZO}^1$, highlighted with dashed and solid lines, respectively. According to the rules of zoning stated previously, ZO^1 is extended to share at least one bus with its neighbor zones, and as a result, the extended version of ZO^1, which is $\overline{ZO}^1$ (highlighted in the solid line in Fig. 3.1), includes bus 17 (to share one bus with ZO^4) and (bus 95 and 79) (to share one bus with ZO^2 and ZO^3 respectively). The extended version of ZO^2, which is $\overline{ZO}^2$ (highlighted in the solid line in Fig. 3.1) includes bus 79 (to share one bus with Z^3). The extended version of ZO^3, which is $\overline{ZO}^3$ (highlighted in the solid line in Fig. 3.1) includes bus 73 (to share one bus with ZO^4). ZO^4 is equal to $\overline{ZO}^4$, as after the extension of the previous zones, ZO^4 shares at least one bus with its neighbour zones.

Each zone estimates its state independently as a first step described in [74] and allows efficient estimation of the state within a controlled time. The zoning algorithm allows the system state to be estimated in a fully distributed manner in two steps (see [74] for more details).
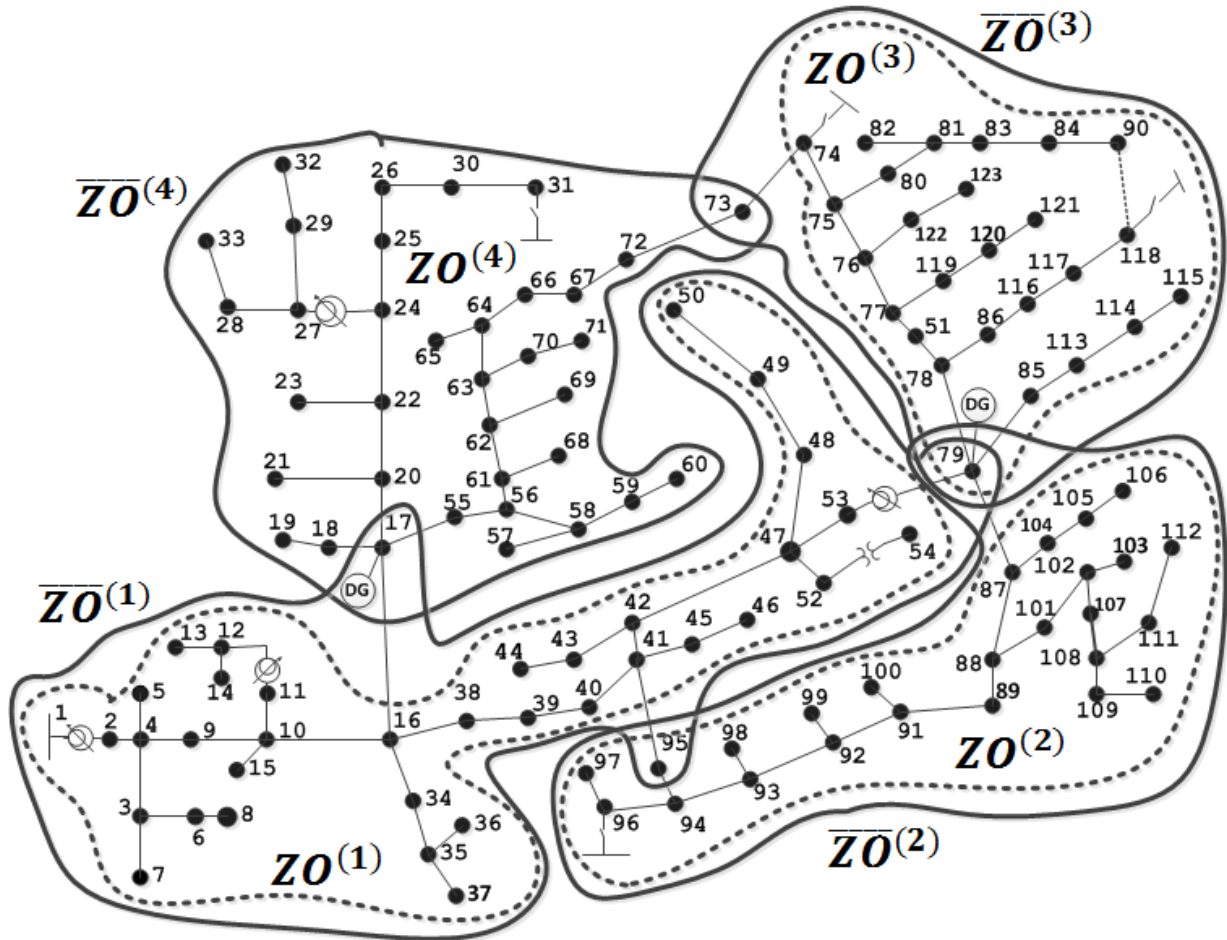
Figure 3.1: IEEE 123-bus system zoned according to the proposed algorithm

3.4   The Proposed Filter Scheme

The objective is to estimate the distribution system bus voltage magnitudes and angles within each zone $\overline{ZO}^l$. The proposed SCPF technique is constructed by a combination of the SCKF and PF. It recursively updates the estimates using the previous estimated value from time $k - 1$ and the measurements (real and pseudo) at time k to provide the final estimation of the bus voltage magnitudes and angles at time instant k.

At each zone $\overline{ZO}^l$, we are given a measurement vector $z := \begin{bmatrix} z_r^\top & z_p^\top \end{bmatrix}^\top \in \mathbb{R}^{n_z}$. $z_r \in \mathbb{R}^{n_z^r}$ is the vector of real power measurements provided by the micro phasor-measurement units (μPMUs). $z_p \in \mathbb{R}^{n_z^p}$ is the vector of pseudo-measurements provided by the load forecaster. $n_z := n_z^r + n_z^p$ is the total number of power measurements. n_z^r and n_z^p are the number of real and pseudo-measurement, respectively. In compact notation, the set of measurement equations that include the injected active and reactive power and power flows between zones can be written as follows:

$$z = h(x) + v \tag{3.1}$$

where, $h(x)$ is a nonlinear function (power flow equation relating the injected power to the voltage magnitudes and angles). v is the measurement noise, with either a Gaussian or non-Gaussian distribution.

$x \in \mathbb{R}^{n_x}$ is the vector of the state variables. n_x is the total number of state variables. The states that going to be estimated are the voltage magnitude and angle at each bus $x := \begin{bmatrix} |\mathbf{v}_1| \cdots |\mathbf{v}_{\frac{n_x}{2}}| & \delta_1 \cdots \delta_{\frac{n_x}{2}} \end{bmatrix}^\top$. μPMUs are assumed to be available on buses 1 (primary substation) and 17 and 79 (DG buses). Five additional μPMUs are optimally located at buses 24, 62, 77, and 89 using the ordinal optimization approach described in [98]. Having a μPMU at bus m means that the voltage $\mathbf{v}_m[k]$ and the current $\sum_{n:m\sim n} \mathbf{i}_{mn}$ are both available, where $m \sim n$ denotes that bus m and n are connected through a line. Instead of current measurements, we can easily derive and use the injected active and reactive power and the power flow in the lines connected to the bus equipped by the μPMU. In the case where there is no μPMU in the bus, the power injected (P and Q) is extracted from a forecast data (pseudo-measurements). Table 3.1 shows the real measurement provided by

Importance Weight → $\boldsymbol{w_{k-1}}$

Resampling to Particles → $\widetilde{\boldsymbol{x}}_{\boldsymbol{k-1}|\boldsymbol{k-1}}$

Predicted States in Particle Form → $\widehat{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k-1}}$

Move all Particles to the Area of Interest → $\overline{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k}}$

Ratio in Eq (3.2)

Importance Weight → $\boldsymbol{w_k}$

Resampling Enhancement Approximation → $\widetilde{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k}}$
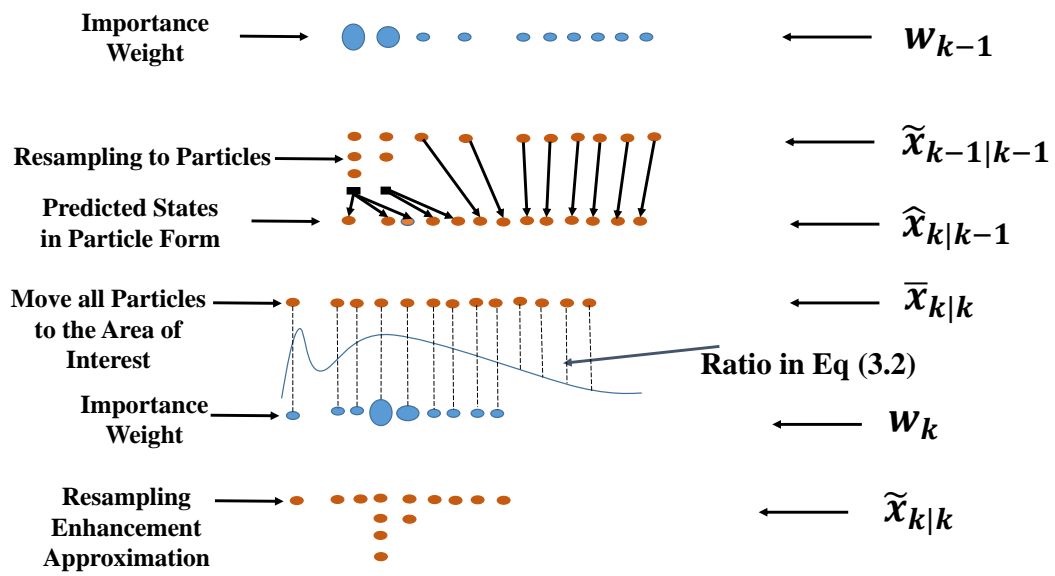
Figure 3.2: The proposed SCPF estimator graphical representation

Table 3.1: Measurements of various quantities at different Buses

Zone.	Measurements
1	$V_1, V_{17}, V_{79}, P_1, P_{17}, P_{79}, Q_1, Q_{17}, Q_{79}$
2	$V_{79}, V_{89}, P_{79}, P_{89}, Q_{79}, Q_{89}$
3	$V_{77}, V_{79}, P_{77}, P_{79}, Q_{77}, Q_{79}$
4	$V_{17}, V_{24}, , V_{62}, P_{17}, P_{24}, , P_{62}, , Q_{17}, Q_{24}, , Q_{62}$

μPMU.

The PF algorithm forms the core of the proposed estimator. Unlike the KF-based algorithms, a PF is a non-Gaussian algorithm, which matches the requirements of distribution systems. The PF is a non-Gaussian algorithm in the sense that it accepts non-Gaussian measurements without any approximation, and produces non-Gaussian estimates. The PF is based primarily on four sub-algorithms: a perfect Monte Carlo simulation, Bayesian importance sampling, sequential importance sampling (SIS), and re-sampling (SIR) [110] [3]. The details of these sub-algorithms are explained in [3]. These sub-algorithms lead to (3.9), which is used to update the importance weight sequentially:

$$w_k = w_{k-1} \frac{p\left(z_k | \bar{x}_{k|k}\right) p\left(\bar{x}_{k|k} | \hat{x}_{k-1|k-1}\right)}{q\left(\bar{x}_{k|k} | z_k, \ddot{x}_{k|k}\right)} \qquad (3.2)$$

Here, w_k is the particle weight vector of the states of ZO^l [1]. at time instant k, and z_k is the measurement vector drawn from the measurement distribution, which is Gaussian or non-Gaussian. $p\left(z_k | \bar{x}_{k|k}\right)$ is the particle likelihood vector of the states of ZO^l. The vector $p\left(\bar{x}_{k|k} | \hat{x}_{k-1|k-1}\right)$ represents the particle prior estimate of the states of ZO^l, and $q\left(\bar{x}_{k|k} | z_k, \ddot{x}_{k|k}\right)$ is the particle proposal distribution vector of the states of ZO^l, which is provided by the SCKF. Data processing of the proposed algorithm is graphically illustrated in Fig. 3.2. N particles are drawn from the previous posterior $p\left(\hat{x}_{k-1|k-1} | z_{k-1}\right)$. To avoid

---

[1]For this section we abuse the notation by ignoring the superscript l, to avoid overloading the notation. However, all the calculations in this section are done per zone

the degeneration of the SIS, each particle is re-sampled according to its weight. The SIR algorithm is implemented in this chapter, where a sample with a high importance weight regenerates more often than those with a lower importance weight, as shown in Fig. 3.2. The SCKF is used to provide the particle proposal distribution $q\left(\bar{x}_{k|k}|z_k, \ddot{x}_{k|k}\right)$, which is an approximation of the integral of $p\left(z_k|\bar{x}_{k|k}\right)p\left(\bar{x}_{k|k}|\hat{x}_{k-1|k-1}\right)$, which provides the normalization constant for the posterior.

The SCKF is developed using the least-squares method for the cubature Kalman gain and matrix triangular factorization or triangularization for the covariance updates. The least-squares method avoids the explicit computation of matrix inversion, whereas triangularization essentially computes a triangular square-root factor of the covariance without square rooting a squared-matrix form of the covariance. The SCKF essentially propagates the square roots of the predictive and posterior error covariances and offers the following benefits as compared to the CKF introduced in [92]:

1. Preservation of symmetry and positive (semi) definiteness [5]; and

2. Numerical accuracy enhancment [5].

The SCPF algorithm is explained as follows[2]:

  At each time instant k perform the following:
For each particle $j \in \{1, \cdots, N\}$,

1. Evaluate the prior as follows:

$$I_k = \exp(-0.5(\bar{x}_{k|k} - \hat{x}_{k-1|k-1})^\top Q^{-1}(\bar{x}_{k|k} - \hat{x}_{k-1|k-1})) \tag{3.3}$$

where, $\hat{x}_{k-1|k-1} \in \mathbb{R}^{n_x}$ is the previous estimated states. $Q \in \mathbb{R}^{n_x \times n_x}$ is process covariance matrix. $\bar{x}_{k|k} \in \mathbb{R}^{n_x}$ is the propagated particle which is computed as follows:

$$\bar{x}_{k|k} = \ddot{x}_{k|k} + \Gamma_{k|k}. \tag{3.4}$$

---

[2]The CPF is also designed in same manner by including the CKF as sub-algorithm.

where, $\Gamma_{k|k} \in \mathbb{R}^{n_x} \sim \mathcal{N}(0, S_{k|k})$. $S_{k|k}$ is estimated covariance matrix provided by the SCKF. $\ddot{x}_{k|k}$ is the estimated states computed by the SCKF as follows:

$$\ddot{x}_{k|k} = \dot{x}_{k|k-1} + K_k \left(z_k - \hat{z}_{k|k-1} \right), \tag{3.5}$$

where, $\dot{x}_{k|k-1} \in \mathbb{R}^{n_x}$ is the predicted state vector computed as shown in Section 3.4.1. K_k is the Kalman gain computed as shown in Section 3.4.1. $z_k \in \mathbb{R}^{n_z}$ is the measurement vector. $\hat{z}_{k|k-1} \in \mathbb{R}^{n_z}$ is the predicted measurement vector computed as follows:

$$\hat{z}_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} \hat{Z}_{i,k|k-1}, \tag{3.6}$$

$\hat{Z}_{i,k|k-1}$ is the the predicted measurement vector evaluated at cubature point i (see Section 3.4.1 for more details). $m = 2n_x$ is the number of the cubature points.

2. Evaluate the likelihood as follows:

$$\jmath_k = \exp(-0.5(z_k - \hat{z}_{k|k-1})^\top R^{-1}(z_k - \hat{z}_{k|k-1})) \tag{3.7}$$

where, $R \in \mathbb{R}^{n_z \times n_z}$ is measurement covariance matrix.

3. Evaluate the proposal distribution as follows:

$$\imath_k = \frac{\exp(-0.5(\bar{x}_{k|k} - \ddot{x}_{k|k})^\top S_{k|k}^{-1}(\bar{x}_{k|k} - \ddot{x}_{k|k}))}{\sqrt{|S_{k|k}|}} \tag{3.8}$$

4. Calculate the importance weight (w_k) of each particle as follows:

$$w_k = \frac{\jmath_k I_k}{\imath_k} \tag{3.9}$$

5. After calculating the importance weight of each particle[3], we normalize the weight of each particle as follows:

$$\hat{w}_k = \frac{w_k^j}{\sum_{j=1}^{N} \left(w_k^j \right)} \tag{3.10}$$

---

[3]All the above calculations are per particle j. We abuse the notation by ignoring the superscript j, to avoid overloading the notation.

6. After normalizing the importance weight of each particle, the re-sampling algorithm is now applied to copy the fit particle and eliminate the lower-importance particles and generate the random samples $\{\tilde{x}_{k|k}\}$.

7. The expectation of the $x[k]$ can therefore be calculated as:

$$\hat{x}_{k|k} = E\left(x_{k|k}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(\tilde{x}_{k|k}^{j}\right) \tag{3.11}$$

Therefore, the expectation of the variance $P[k]$ can be calculated as:

$$\hat{S}_{k|k} = E\left(P[k]\right) = \frac{1}{N}\sum_{j=1}^{N}\left(S_{k|k}^{j}\right) \tag{3.12}$$

After each local estimator estimates its state independently, Step 2 (explained in the parallel processing algorithm proposed in [74]) is used to synchronize all zones.

3.4.1   Square Root Cubature Kalman Filter

As stated in the introduction, we leverage the results in [5] to provide the proposal distribution for PF. In this section, we provide the SCKF algorithm details [5] which used as a sub-algorithm of the SCPF.

First, we compute the predicted state vector, using the transition model presented in [119]. The steps of calculating the predicted state vector is as follows:

1. For each cubature point $i \in \{1, 2, \cdots, m\}$, calculate the following:

$$X_{i,k-1|k-1} = \hat{S}_{k-1|k-1}\xi_i + \tilde{x}_{k-1|k-1}, \tag{3.13}$$

where, $\xi_i = \sqrt{\frac{m}{2}}e_i{}^4$ is the cubature point-weight set. $\hat{S}_{k-1|k-1} \in \mathbb{R}^{n_x \times n_x}$ is the previous estimated covariance matrix. $m = 2nx$ is the number of cubature points.

---

[4] $e_i - i$th natural basis vector for $\mathbb{R}^{n_x}$

2. compute the predicted state vector for each propagated cubature point i.

$$\hat{X}_{i,k|k-1} = X_{i,k-1|k-1} - H^{-1}(X_{i,k-1|k-1})$$
$$\left(Y_{k|k-1} - Y_{k-1|k-1}\right) \tag{3.14}$$

We use the transition model presented in [119]. The matrix $H^{-1} \in \mathbb{R}^{n_x \times n_z} = \partial h(x)/\partial x$ is the Jacobin matrix evaluated at $X_{i,k-1|k-1}$. $\hat{X}_{i,k|k-1} \in \mathbb{R}^{n_x}$ is the predicted state vector at cubature point i. $Y_{k-1|k-1}$ and $Y_{k|k-1}$ are respectively the forecast injected power at time $k-1$ and k.

3. The predicted state vector can be calculated as follows:

$$\hat{x}_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} \hat{X}_{i,k|k-1}, \tag{3.15}$$

4. Once the predicted state vector is calculated, the square root of the predicted error covariance can be estimated as follows:

$$S_{k|k-1} = \mathbf{Tria}([\hat{\varkappa}_{k|k-1} \ \ S_{Q,k-1}]), \tag{3.16}$$

where $S_{Q,k-1}$ denotes a square root of $Q \in \mathbb{R}^{n_x \times n_x}$ such that $Q = S_{Q,k-1}S_{Q,k-1}^{\top}$. **Tria** denotes any one of the tringularization algorithms.

$$\hat{\varkappa}_{k|k-1} = \frac{1}{\sqrt{m}}[\hat{X}_{1,k|k-1} - \hat{x}_{k|k-1} \ \ \hat{X}_{2,k|k-1} - \hat{x}_{k|k-1}$$
$$\hat{X}_{m,k|k-1} - \hat{x}_{k|k-1}] \tag{3.17}$$

Upon receiving the measurements, the predicted state vector can be further filtered using the following:

1. Let $P_{\mathcal{K}} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x^p}$ be the linear map which takes a vector $y \in \mathbb{R}^{n_x}$ and removes the rows corresponding to the measured states.

2. Let $\hat{x}_{k|k-1,p} = P_{\mathcal{K}}\hat{x}_{k|k-1}$.

3. Upon receiving the voltage magnitude and angle measurement $(\hat{x}_{k|k-1,r})$, let $\dot{x}_{k|k-1} = \left[\hat{x}_{k|k-1,r}^{\top} \ \ \hat{x}_{k|k-1,p}^{\top}\right]^{\top} \in \mathbb{R}^{n_x}$.

30

4. For each cubature point $i \in \{1, 2, \cdots, m\}$, compute the following.

$$X_{i,k|k-1} = S_{k|k-1}\xi_i + \dot{x}_{k|k-1}, \tag{3.18}$$

5. Compute the predicted measurement for each $X_{i,k|k-1}$ computed using 3.18:

$$\hat{Z}_{i,k|k-1} = h(X_{i,k|k-1}), \tag{3.19}$$

6. Square root of the innovations covariance matrix is calculated as follows:

$$S_{zz,k|k-1} = \mathbf{Tria}([\hat{\Upsilon}_{k|k-1} \ \ S_{R,k}]), \tag{3.20}$$

where, $S_{R,k}$ denotes a square root of $R \in \mathbb{R}^{n_z \times n_z}$, such that $R = S_{R,k}S_{R,k}^\top$.

$$\hat{\Upsilon}_{k|k-1} = \frac{1}{\sqrt{m}}[\hat{Z}_{1,k|k-1} - \hat{z}_{k|k-1} \ \ \hat{Z}_{2,k|k-1} - \hat{z}_{k|k-1}$$
$$\hat{Z}_{m,k|k-1} - \hat{z}_{k|k-1}] \tag{3.21}$$

7. Square-root cubature Kalman gain is computed as follows[5]:

$$K_k = (P_{xz,k|k-1}/S_{zz,k|k-1}^\top)/S_{zz,k|k-1} \tag{3.22}$$

where, $P_{xz,k|k-1}$ is the cross-covariance matrix (see [5] for more details in calculating $P_{xz,k|k-1}$)

8. The corresponding error covariance is calculated as follows:

$$S_{k|k} = \mathbf{Tria}([\varkappa_{k|k-1} - K_k\hat{\Upsilon}_{k|k-1} \ \ K_kS_{R,k}]), \tag{3.23}$$

Fig. 3.3 reflects the real scenario in the terms of the load having a high probability of being negative (i.e., the demand is greater than the generation of the RES available on the customer side), but, conversely, it also has a possibility of being positive because of the unpredictable nature of the RES. Clearly, as described in Fig. 3.3, the distribution of

---

[5]/ is used to denote the matrix right division operator, which applies the back substitution algorithm for an upper triangular matrix S and the forward substitution algorithm for a lower triangular matrix S
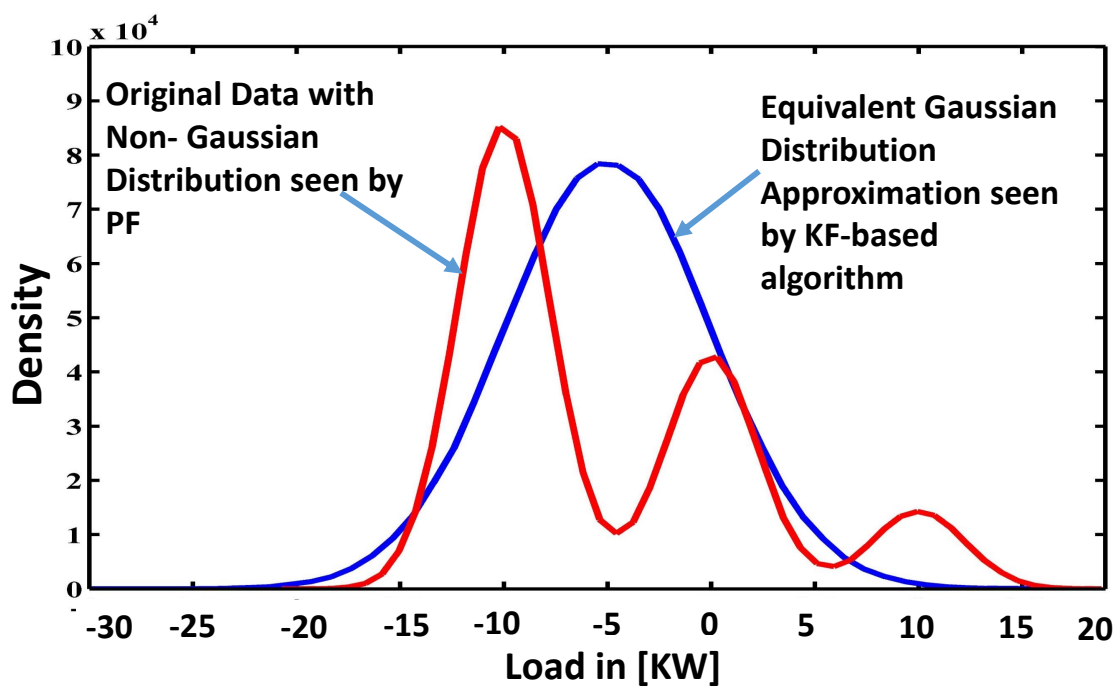
Figure 3.3: Measurement distribution [67]

the original data does not follow a Gaussian distribution. The original data distribution and the best-fit Gaussian distribution (obtained by Expected Maximization (EM) [115]) are put side by side to show the significant difference. In fact, the highest intensity of load, according to the Gaussian distribution occurs around -5 KW, while the intensity of the original data distribution occurs around -10 KW. The intensity of the original data distribution around the -5 KW is exceedingly low.

3.5 Simulation Results and Discussion

3.5.1 Generation of Bus Profiles

The IEEE 123-bus system is upgraded by including two DGs located at buses 17, 79 and 97. Moreover, a high degree of variations in the load profiles with uncertainty are considered to simulate the effect of RESs and EVs on the customer side, as shown in Fig. 3.4. In this chapter, the authors used extreme examples to represent the dynamic behavior of the system in order to fully test the proposed algorithms. For example, in Fig. 3.4(a) and (b), the load curves have high-frequency behavior as a result of the RESs and EVs available on the customer side. Moreover, the value of the active load at several instants in Fig. 3.4(a) is negative, which simulates the case when customers are supplying a power to the grid, as the power provided by the RESs and EVs and/or the DG becomes greater than the demand for that particular bus. From the above discussion the load for the buses 17, 79 and 97 is assumed to be sampled from the set [-0.04 0.045]. Moreover, all buses except buses 17, 79 and 97 is assumed to be sampled from the set [-0.003 0.008]. These load profile generated from these sets give highest load with of the distribution system with value 1.2 *p.u.* and lowest value -0.6 *p.u.*. These design assumptions are justifiable as the reverse power is expected to be 60% of the load in the presence of the high penetration of the DGs either in customer side or as an aggregated one. Moreover, a high fluctuation can be obtained from these design assumptions which simulate the expected behavior of the DGs and EVs. The reactive power for each bus is generated by assuming that the power factor varies between 0.85 to 0.99. To evaluate the accuracy of the proposed SEs, several tests were performed using Monte Carlo (MC) simulations. The number of MC simulations is

10,000, and for each simulation, a set of 96 sequential true data samples is generated for each bus. The 10,000 scenarios generated can cover all the possible circumstances that the distribution system may experience. The non-Gaussian pseudo-measurements (the true representation) are generated using GMM (3 Gaussian components) with $(\sigma_1 + \sigma_2 + \sigma_3)$ equivalent to 50% error. We choose 3 Gaussian components following the result of the work presented in [67].

As stated in [70], the relationship between the states and the measurement becomes more nonlinear in the heavy loading condition. Therefore, to test the performance of the proposed SEs in different extreme cases, the authors created heavy loading conditions between samples 80 and 90 ($k = 80$ to $k = 90$). In this particular case, the demand became slightly larger than the generation. The authors used the interval between 80 and 90, as peak consumption occurs between 6 pm (80) and 8:30 pm (90) for most countries. Fig. 3.4 shows 96 sequential samples of the true values and pseudo-measurements of the load profile of bus 97. The 96 samples were chosen to highlight the fact that the load might have negative values for active and reactive power.

3.5.2 Performance Evaluation

The IEEE 123-bus system is used to evaluate the performance of the proposed SCKF, CPF, and SCPF under normal and abnormal operations, and the performance of the proposed algorithms is compared against the latest versions of the least square family [52] with different combinations of filtering schemes. The underlying network topology is assumed to be known a priori and fixed over the time duration of state estimation. The load flow results at each time instant ($k = 1$ to 96) using true values are calculated and used as a benchmark to evaluate the accuracy of the proposed algorithms. The maximum values of the relative errors in voltage magnitudes (MRE) and angles (ARE) are used to assess the performance of the proposed filter:

$$
\begin{aligned}
\mathrm{MRE}^i &= \frac{|V_i - \hat{V}_i|}{|V_i|} \text{ and;} \\
\mathrm{ARE}^i &= \frac{|\theta_i - \hat{\theta}_i|}{|\theta_i|},
\end{aligned}
\tag{3.24}
$$

34
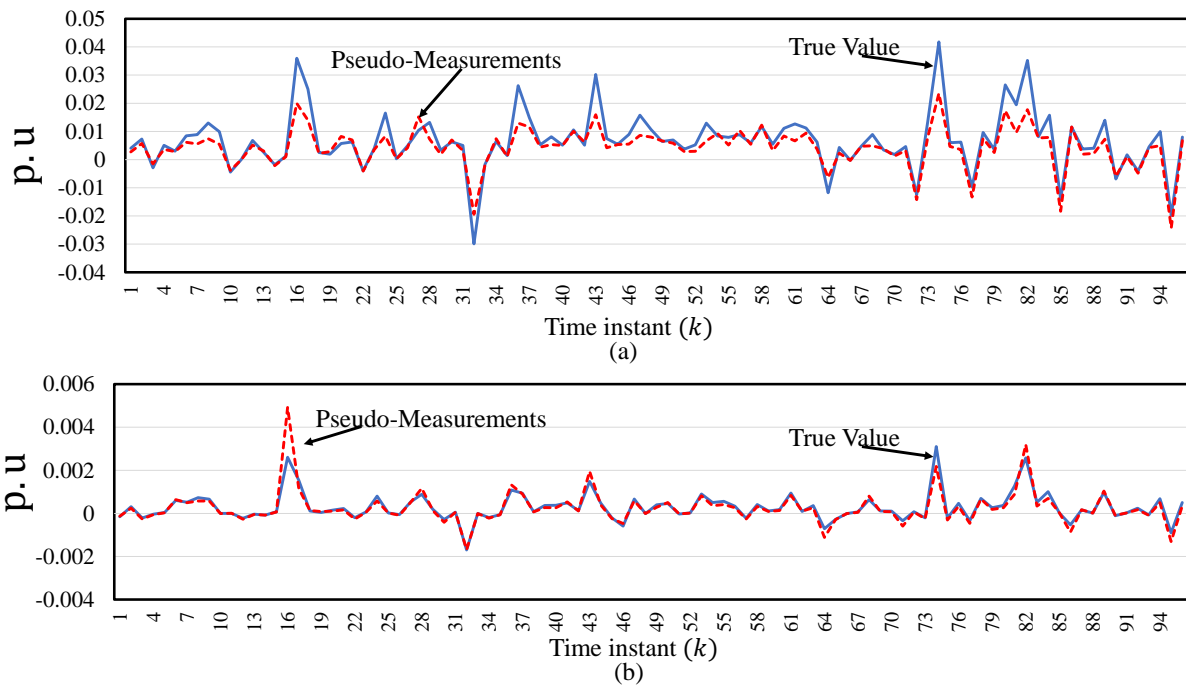
Figure 3.4: Samples of true and pseudo-measurements data at bus 97: a- active power and b- reactive power.

where, $\hat{V}_i$ is the estimated voltage magnitude at bus i, V_i is the true value, $\hat{\theta}_i$ is the estimated angle, and θ_i is the true value. The estimator result is considered satisfactory if the $\mathrm{MRE_{max}}$ is smaller than 1% and the $\mathrm{ARE_{max}}$ is smaller than 5% [98] in all system states.

Fig. 3.5, Fig. 3.6, Fig. 3.7 and Fig. 3.8 show the results of the performance of the proposed state estimators and the latest version in the KF family CKF using (MRE) and (ARE) metrics. The x-axis represents bus numbers included in zone 1 and zone 2 in ascending order. The y-axis in Fig. 3.5, Fig. 3.6, Fig. 3.7 and Fig. 3.8, represents either the maximum MRE or ARE as follows:

$$
\begin{aligned}
\mathrm{MRE}^i_{\max} &= \max_{l \leq M_c} (\max_{k \leq O} \mathrm{MRE}^l_k) \\
\mathrm{ARE}^i_{\max} &= \max_{l \leq M_c} (\max_{k \leq O} \mathrm{ARE}^l_k)
\end{aligned}
\tag{3.25}
$$

The $\mathrm{MRE_{max}}$ and $\mathrm{ARE_{max}}$[6] for all buses in zone 1 and zone 2 are for $O = 96$ time instants for $M_c = 10{,}000$ Monte Carlo simulation. Fig. 3.5, Fig. 3.6, Fig. 3.7 and Fig. 3.8 show the results when we draw the pseudo-measurement from the non-Gaussian distribution, i.e., when we assume that the pseudo-measurement has a non-Gaussian distribution. For zone 1, using the CPF, the $\mathrm{ARE_{max}}$ is $15\%(> 5.0\%)$ and $\mathrm{MRE_{max}}$ is $1.05\%(< 1\%)$, while using the CKF, the $\mathrm{ARE_{max}}$ is $47.5\%(> 5.0\%)$ and $\mathrm{MRE_{max}}$ is $7.9\%(> 1\%)$. Moreover, using the SCPF for zone 1, the $\mathrm{ARE_{max}}$ is $4\%(< 5.0\%)$ and $\mathrm{MRE_{max}}$ is $0.9\%(< 1\%)$, while using the SCKF, the $\mathrm{ARE_{max}}$ is $27.1\%(> 5.0\%)$ and $\mathrm{MRE_{max}}$ is $4.3\%(> 1\%)$. The metrics $\mathrm{MRE_{max}}$ and $\mathrm{ARE_{max}}$ are computed as follows:

$$
\begin{aligned}
\mathrm{MRE_{max}} &= \max_{i \in \overline{ZO}_l} \mathrm{MRE}^i_{\max} \\
\mathrm{ARE_{max}} &= \max_{i \in \overline{ZO}_l} \mathrm{ARE}^i_{\max}
\end{aligned}
\tag{3.26}
$$

One can notice that, in terms of accuracy the best performance is obtained by the SCPF and, then, the CPF. Moreover, the SCKF show better performance compared to the the CKF in terms of accuracy, while the SCKF shows less performance compared to

---

[6]We ignore the superscript i in the figures since the y-axis represent the number of buses i

Figure 3.5: (a) Relative voltage angle error for zone 1, CKF, and CPF, (b) Relative voltage angle error for zone 1, SCKF, and SCPF.

Table 3.2: SCPF performance summary for non-Gaussian pseudo-measurement (uncertainty 50%)

Zone.	$\mathrm{MRE_{max}}$	$\mathrm{ARE_{max}}$	$\mathrm{MRE_{ave}}$	$\mathrm{ARE_{ave}}$
1	9.0×10^{-3}	40.0×10^{-3}	6.0×10^{-3}	18.1×10^{-3}
2	3.8×10^{-3}	38.1×10^{-3}	2.8×10^{-3}	13.5×10^{-3}
3	4.2×10^{-3}	38.4×10^{-3}	2.7×10^{-3}	14.4×10^{-3}
4	8.8×10^{-3}	39.3×10^{-3}	5.3×10^{-3}	17.3×10^{-3}

the CPF and SCPF in terms of accuracy. The evaluation results of applying the SCPF technique on the whole IEEE 123-bus system are summarized in Table 3.2. $\mathrm{MRE_{ave}}$ and $\mathrm{ARE_{ave}}$ used in Table 3.2 and 3.3 are defined as follows:

$$\mathrm{MRE_{ave}} = \frac{1}{|\overline{ZO_l}|} \sum_{i \in \overline{ZO_l}} \mathrm{MRE}^i_{\mathrm{max}}$$

$$\mathrm{ARE_{ave}} = \frac{1}{|\overline{ZO_l}|} \sum_{i \in \overline{ZO_l}} \mathrm{ARE}^i_{\mathrm{max}}$$

(3.27)

where, $|\overline{ZO_l}|$ represent the cardinality of the zone l (i.e., the number of buses in the zone l). Table 3.2 shows the results for the case of non-Gaussian measurements with an uncertainty equal to 50% for all zones. From Table 3.2 one can notice the performance difference in terms of accuracy between the zones. One reason for this is that we use less number of particles ($N = 65$) to estimate the states of $\overline{ZO}_1$ and $\overline{ZO}_4$, compared to ($N = 90$) to estimate the states of $\overline{ZO}_2$ and $\overline{ZO}_3$. The reason for this is that we want the computational time of each zone to be approximately equal (1.9 s). We can conclude that, as the size of the zone increases, the computational time increases. Another reason for the difference in terms of accuracy between the zones is that the R/X ratios in the $\overline{ZO}_1$ is higher than R/X ratios in the $\overline{ZO}_2$.

In this subsection, the proposed algorithm is tested for one expected case in order to show the difference in its performance against the basic case, the results for which are

Figure 3.6: (a) Relative voltage magnitude error for zone 1, CKF, and CPF, (b) relative voltage angle error for zone 1, SCKF, and SCPF.
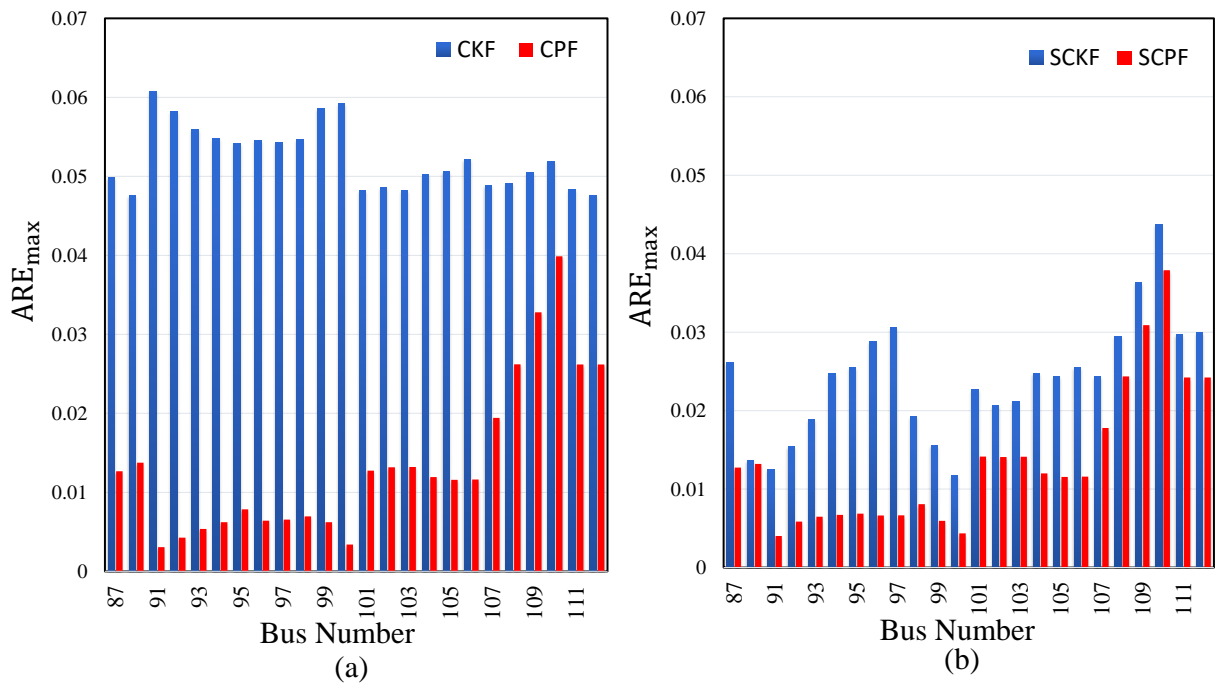
Figure 3.7: (a) Relative voltage angle error for zone 2, CKF, and CPF, (b) Relative voltage magnitude error for zone 2, SCKF, and SCPF.
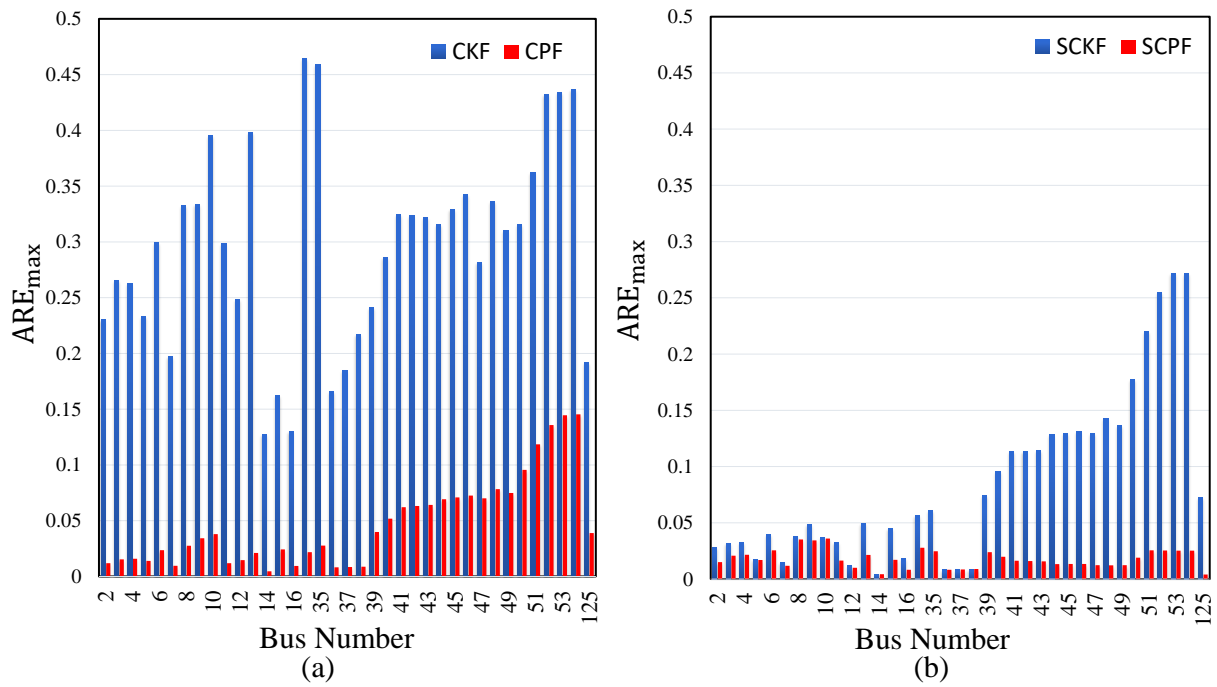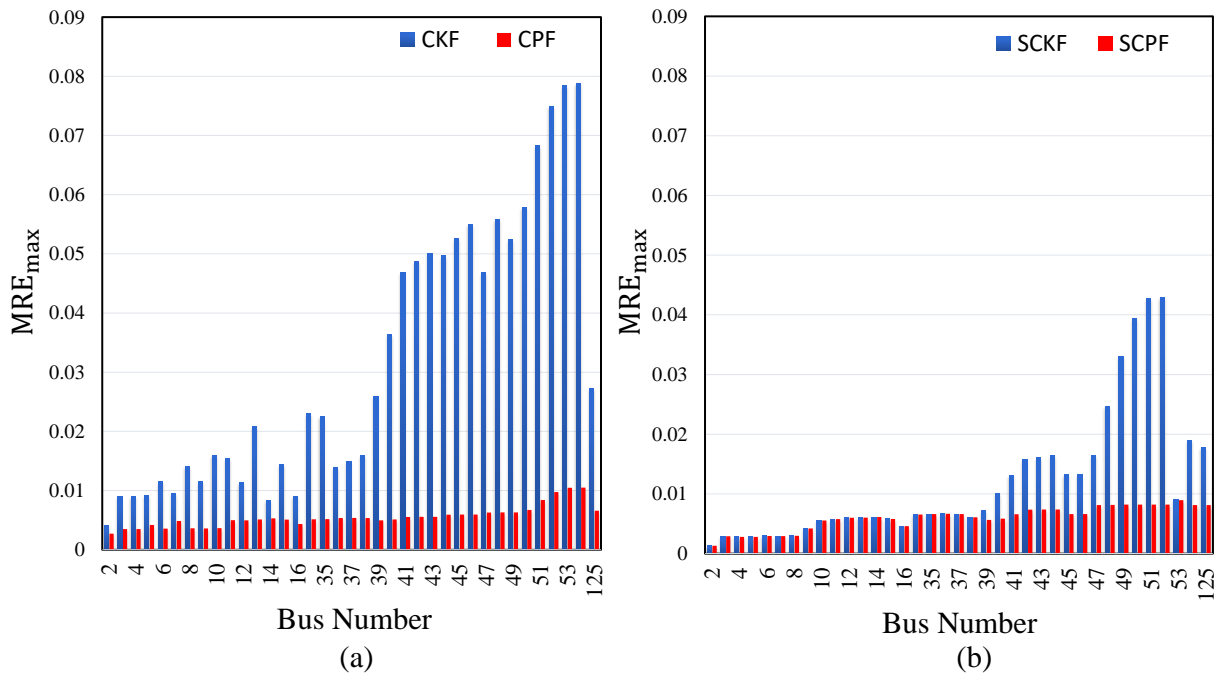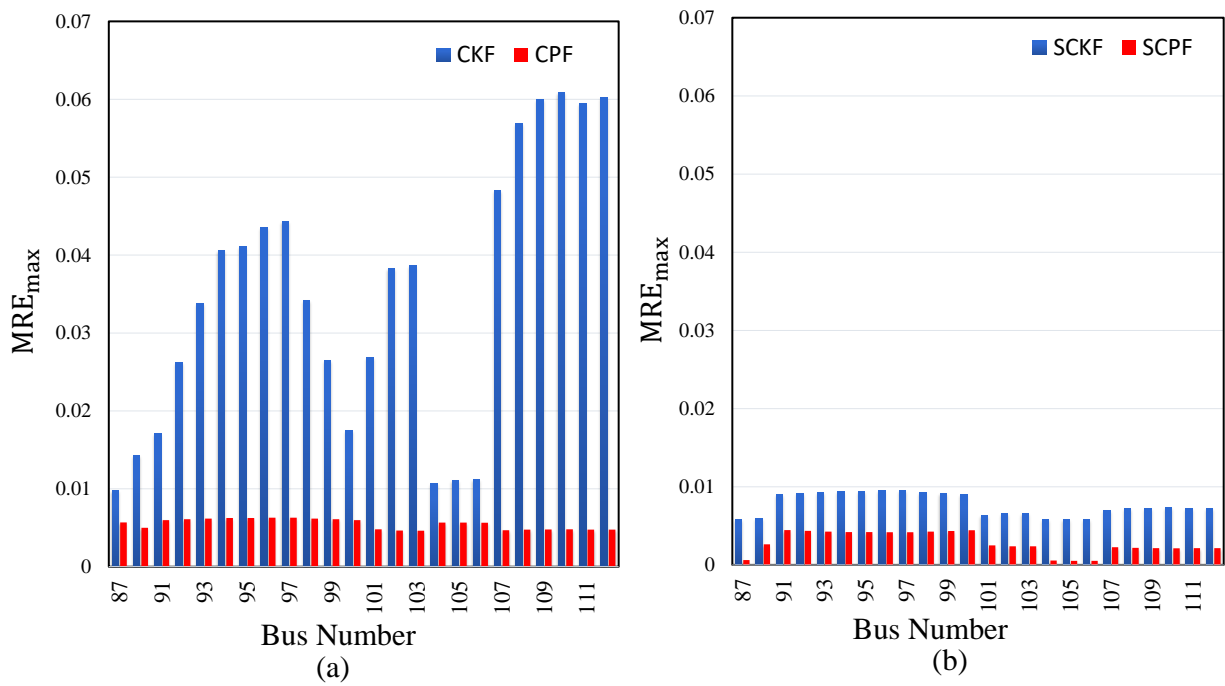
Figure 3.8: (a) Relative voltage magnitude error for zone 2, CKF, and CPF, (b) Relative voltage angle error for zone 2, SCKF, and SCPF.

Table 3.3: SCPF performance summary for non-Gaussian pseudo-measurement (uncertainty 30%)

Zone.	MRE_{max}	ARE_{max}	MRE_{ave}	ARE_{ave}
1	8.4×10^{-3}	37.6×10^{-3}	5.6×10^{-3}	17.1×10^{-3}
2	3.6×10^{-3}	35.8×10^{-3}	2.6×10^{-3}	12.7×10^{-3}
3	3.9×10^{-3}	36.1×10^{-3}	2.5×10^{-3}	13.7×10^{-3}
4	8.3×10^{-3}	36.9×10^{-3}	5.0×10^{-3}	16.6×10^{-3}

shown in Table 3.2. The expected case that was tested is the effect of the quality of the pseudo-measurements.

First, the authors want to show how the performance of the proposed algorithm is sensitive to uncertainty of the pseudo-measurements. In this test, there was a simulation of the case where the available power measurements from smart meters are used to enhance the accuracy of the load forecaster. This case was simulated by adding an error of 30% to the true value instead of 50%, which is used in the basic case. The results show that performance increased for all zones. When the results in Table 3.2 are compared with the results in Table 3.3, we can see that the performance of the estimator was enhanced by approximately 4% to 8% due to the enhancement of the pseudo-measurement accuracy.

To ensure the robustness of the proposed algorithm, the SCPF algorithm is tested under abnormal operations. A sudden change in the voltage magnitude and the angle of bus 90 is simulated for Sample 15 (Fig. 3.9). In this case, the demand became much larger than the generation for one time instant. Correctly estimating the states in the case of sudden changes is very important, especially when the frequency sensor of the automatic generation controller (AGC) is attacked. The attacker can compromise the sensor data to give the false impression to the operator that no change in the load is occurring. Using the estimated states provided by the proposed estimator allows the operator to figure out the situation in the system from different sources of data, and then take action. The proposed SCPF algorithm was capable of capturing these sudden changes accurately. Furthermore,
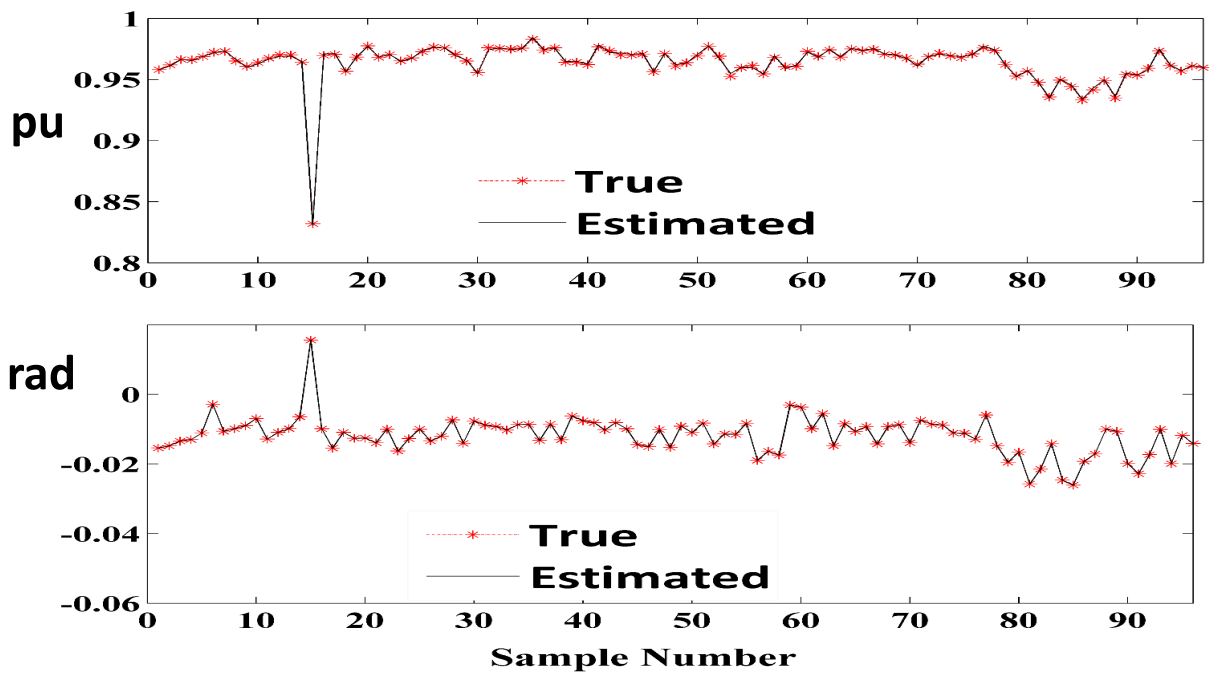
Figure 3.9: True and estimated values of the voltage magnitude and angle results during sudden change in an IEEE 123-bus system: Zone 3, bus 90.
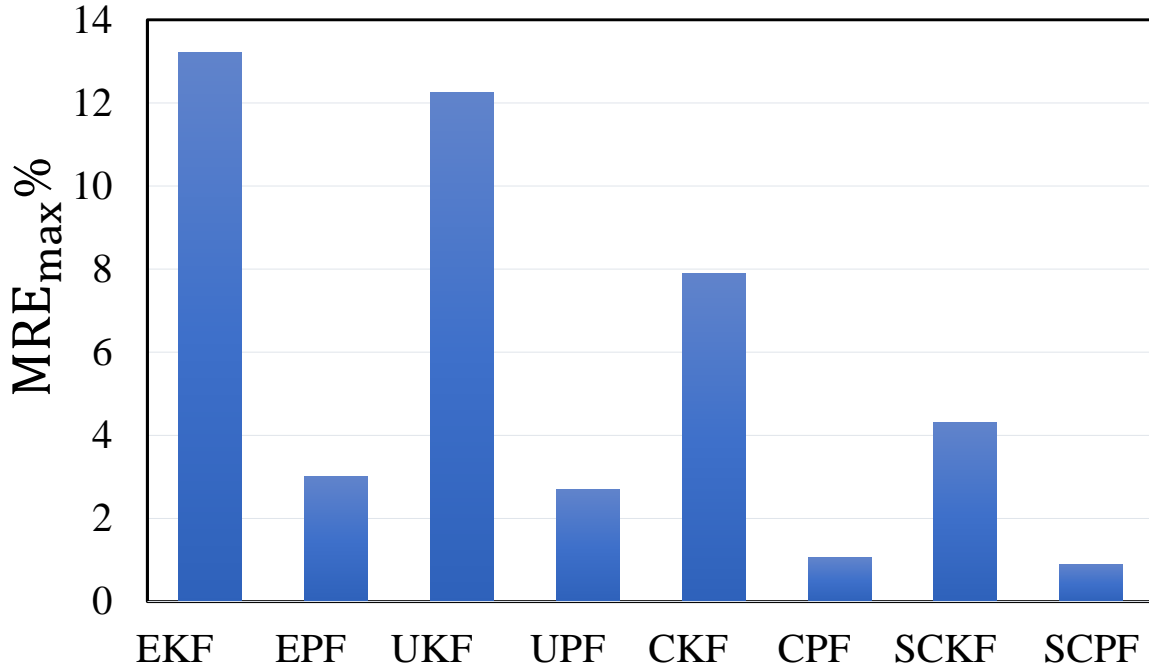
Figure 3.10: Relative Voltage Magnitude Error for the proposed algorithms and the most common versions of the KF-based algorithm.

the proposed SCPF algorithm was tested with different r/x ratios between 0.5 and 3.5, and showed satisfactory results even with greater ratios.

To make this paper self-standing, the latest versions of the KF-based algorithm have been simulated. Fig. 3.10 shows the $\text{MRE}_{\text{max}}\%$ of the proposed algorithms and the most common versions of the KF-based algorithm. Fig. 3.11 shows the $\text{ARE}_{\text{max}}\%$ of the proposed algorithms and the most common versions of the KF-based algorithm. The EKF-Case shows the algorithm's good performance in estimating the states of voltage magnitudes ($\text{MRE}_{\text{max}}\% = 13.22$) and its poor performance in estimating voltage angles ($\text{ARE}_{\text{max}}\% = 70.90$). A noticeable enhancement in accuracy is achieved ($\text{MRE}_{\text{max}}\% = 3.02$ and $\text{ARE}_{\text{max}}\% = 26.12$) when combining the EKF and PF (EPF-Case). The UKF-Case provided additional enhancement in accuracy as compared with the EPF-Case. While the estimated magnitudes are ($\text{MRE}_{\text{max}}\% = 12.25 > 1.0\%$), the estimated angles have a max-
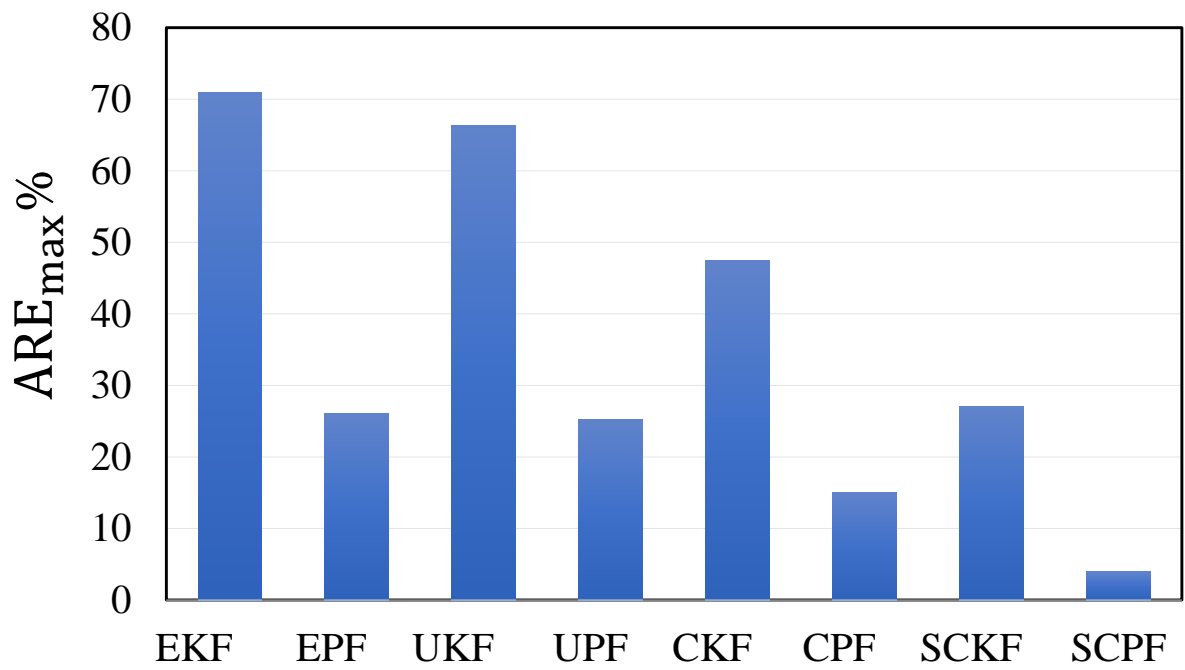
Figure 3.11: Relative Voltage Angle Error for the proposed algorithms and the most common versions of the KF-based algorithm.

imum ARE equal to 66.30%(> 5%). A noticeable enhancement in accuracy is achieved ($\text{MRE}_{\text{max}}\% = 2.70$ and $\text{ARE}_{\text{max}}\% = 25.19$) when combining the UKF and PF (UPF-Case). It is worth noting that the results based on the first step described in [74] are less accurate than than the two-step one presented above by 8% to 10%.

The simulations suggest that, for applications in which the measurements have a Gaussian distribution, the best choice is the introduced SCKF. This is because when we test the SCKF, CPF, and SCPF with measurement with Gaussian distribution, the performance in term of accuracy was almost the same with slightly better performance for the SCPF and CPF. However, the computational time recorded for the SCKF is $0.01\ s$, while for the CPF and SCPF it is $1.9\ s$. The proposed SCPF and CPF can be applied in several applications such as voltage regulation, where the accuracy of the estimation is more important than the computational time. The SCKF can be used in the application where the computational time is more important. We believe that the supercomputer of the power system operator (PSO) can estimate the system states using SCPF or CPF in much faster time. The performance of all the simulated algorithms in descending (i.e., from the highest to lowest performance) order in terms of accuracy is as follows: 1. SCPF, 2. CPF, 3. UPF, 4. EPF, 5. SCKF, 6. CKF, 7. UKF, 8. , and EKF. The performance of all the simulated algorithms in descending (i.e., from the highest to lowest performance) order in terms of computational time is as follows: 1. EKF, 2. CKF, 3. SCKF, 4. UKF, 5. EPF, 6. CPF, 7. SCPF, and 8. UPF. The previous performance comparisons are done based on non-Gaussian measurement distribution.

3.6   Conclusion

This chapter presents SE based on combining the PF and SCKF. We also leverage the SCKF and CPF algorithms for the first time in the electrical distribution system state estimation. The proposed methods transforms a large distribution network into a number of smaller, manageable-size systems (zones). The proposed SCPF and CPF algorithms are capable of processing the non-Gaussian representation of pseudo-measurements without a

Gaussian distribution approximation. The proposed SEs were applied to an IEEE 123-bus system, and the simulation results demonstrated a superior performance in terms of accuracy compared to other estimation algorithms (EKF, UKF, and CKF). Simulations with varying R/X ratios and over a range of different loading conditions show that the proposed SE algorithms is a robust estimator that offers accurate, reliable, and fast estimation.

Chapter 4

Forecast Aided State Estimator for an Electrical Distribution System

4.1 Introduction

This chapter introduces a new forecasting aided state estimation algorithm for a distribution system wide area monitoring. The proposed algorithm composed of two sub-algorithms. The first sub-algorithm is developed to forecast the load of each bus. The second sub-algorithm is developed to estimate the states of each bus using the results of the first sub-algorithm. For the first sub-algorithm, this chapter introduces a wavelet recurrent neural network (WRNN) to model the load at each bus in the distribution system. For training the WRNN, this chapter proposes a new training technique based on square root cubature kalman filter (SCKF). This chapter uses the proposed forecasting algorithm, to estimate the distribution system states. For the second sub-algorithm, this chapter uses an efficient state estimator based on SCKF to estimate the voltage magnitude and angle on each bus. This chapter integrates the VSTLF and the state estimator to design a robust state estimator. Based on the results of the proposed state estimator a voltage situation assessment can be designed to help the operator make the adequate decision.

4.2   Nomenclature

Variables

$\hat{z}_k$ RNN output.

β_k RNN weights.

u_k RNN inputs

$S_{k|k-1}$ Predicted error covariance

$S_{zz,k|k-1}$ Innovations covariance matrix

$P_{xz,k|k-1}$ Cross-covariance matrix

K_k Square-root cubature Kalman gain

$S_{k|k}$ Corresponding error covariance

$x_{k|k}$ Estimated state at time instant k

4.3   Pseudo-Measurements Generator

Fig. 4.1, shows the main stages for the proposed work which are represented by the orange blocks. The black blocks are drawn just for the sake of a complete illustration of the distribution system supervisory control and data acquisition (DSCADA). A link will be made through which the output of the state estimator is fed back to the pseudo measurements generator. The proposed VSTLF will increase the immunization of the proposed state estimator against the delay and loss of the real measurements due to the communication failure. The approach shown in Fig. 4.1 is designed so that the SE can operate in either real-time mode, applying any real-time measurements available from the network along with load estimates, or in forecasting mode to forecast future network states for hours/days-ahead analysis of the system.
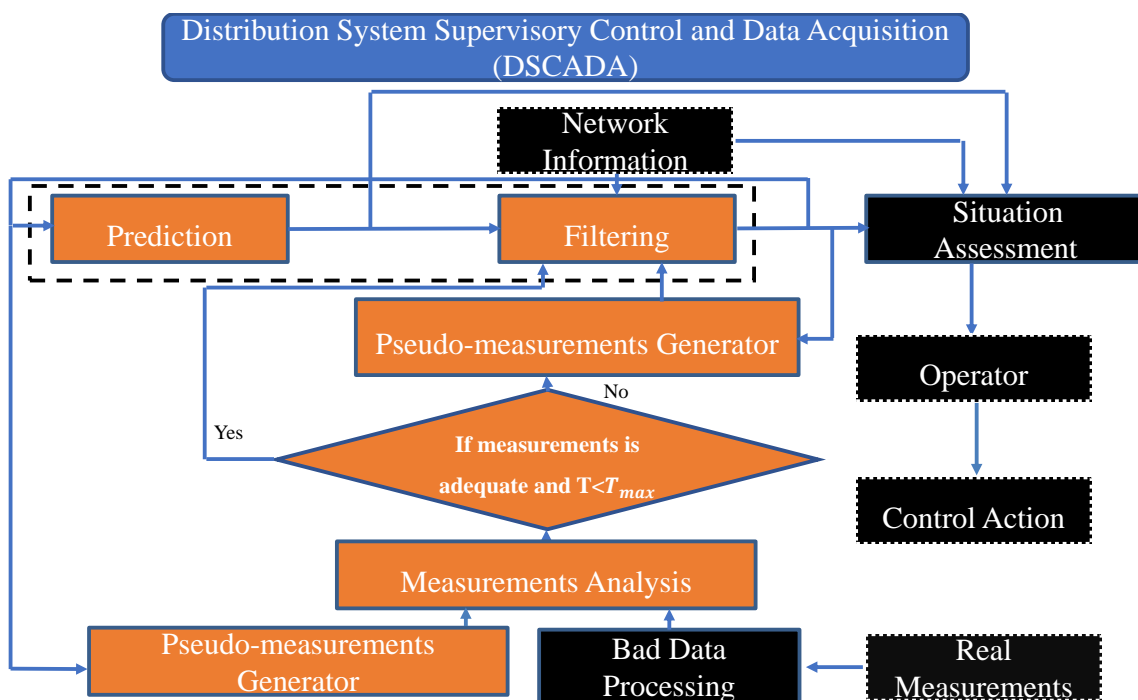
Figure 4.1: DSCADA system

4.3.1 Pseudo-Measurements Generator Design

1. A recurrent neural network (RNN) was considered to model the load dynamic. The model of the RNN is explained as follows:

$$\hat{z}_k = \beta_k [f(\theta_k, u_k, r_i, o_{k-1})^\top \; g_i]^\top \tag{4.1}$$

where, $\hat{z}_k = [\hat{z}_1 \; \hat{z}_2 \cdots \hat{z}_{n_z}]^\top \in \mathbb{R}^{n_z}$ is the RNN output. $g_i := \{1 : \forall \; i \leq n_z\}$ is the output layer bias input. $r_i := \{1 : \forall \; i \leq n_l\}$ is the hidden layer bias input. $\beta_k \in \mathbb{R}^{n_z \times (n_l+1)}$ is the weights. $u_k = [u_{k,1}^\top \; u_{k,2}^\top]^\top \in \mathbb{R}^{n_u+n_b}$ is the input to the RNN. $u_{k,1} = [u_1 \; u_2 \cdots u_{n_u}]^\top \in \mathbb{R}^{n_u}$. $u_{k,2} = [u_{n_u+1} \; u_{n_u+2} \cdots u_{n_u+n_b}]^\top \in \mathbb{R}^{n_b}$. $\theta_k \in \mathbb{R}^{n_l \times (n_u+n_l+n_b+1)}$. n_z is the number of output. n_l is the number of hidden neurons. n_u is the number of the input loads. n_b is the number of the input time index. $f(\theta_k, u_k, r_i, o_{k-1}) = [f_1 \; f_2 \cdots f_{n_l}]^\top \in \mathbb{R}^{n_l}$ is the hidden neuron output as shown in Fig. 4.2 and illustrated in great details in training algorithm steps. $o_{k-1} \in \mathbb{R}^{n_l}$ is the input to the RNN which is also the hidden layer output at $k-1$. The difference between the RNN and FNN is that the output of the hidden layer is fed-back to the Neural Network.

2. Before the training algorithm of the presented models of the RNNs is going to be presented , the preparation of the data is firstly explained. Following [41] The data set from ISO New England is decomposed into three data sets using the wavelet decomposer. It resulted in three components which are low frequency component (L), low-high frequency component (LH), and high frequency component (H) as shown in Fig. 4.3. This chapter uses u_k^L to denote the low frequent component of the data set, u_k^{LH} to denote the low high frequency component of the data set, and u_k^H to denote the high frequency component of the data set. Based on the analysis performed in [41] the low frequency component of the data set has a nearly linear relationship, while the low high and high frequency component have non-linear relationships. Therefore, three different RNN are used to model the behavior of the three different components of the data set as shown in Fig. 4.3. EKFRNN denotes the RNN trained by EKF for low frequency component of the data set. SCKRNN1 denotes the RNN trained by SCKF for high frequency component of the data set . SCKRNN2 denotes the RNN trained
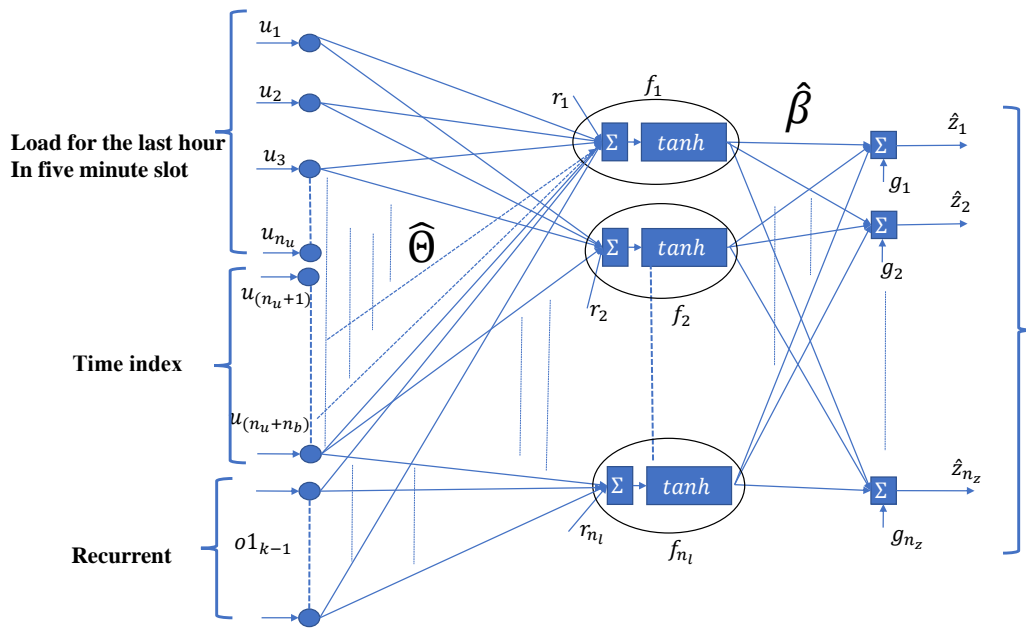
Figure 4.2: Recurrent neural network configuration

by SCKF for low-high frequency component of the data set. The low frequency data set is further manipulated to be ready for training. The low frequency component of the data set is transformed using the relative increment transformation (RI). The reason for using the RI transformation is to make the low frequency component of the data set stationary [41]. The transformed low frequency component of the data set is further normalized using 4.2 to satisfy require of the NN.

The low-high, and high frequency component of the data set are only normalized using the following,

$$\hat{u}_k^{\text{G}} = \frac{u_{k,1}^{\text{G}} - u_{\min}^{\text{G}}}{u_{\max}^{\text{G}} - u_{\min}^{\text{G}}} \tag{4.2}$$

where, $\text{G} \in \{LH \ H\}$. $u_{\max}^{\text{G}}$ and $u_{\min}^{\text{G}}$ respectively denote the maximum and minimum value of the G component of the data set. In the following remaining part of Section 4.3.1, the training algorithm is explained using the high and the low-high frequency components of the data set ($\hat{u}_k^{\text{H}}$, and $\hat{u}_k^{\text{LH}}$).

3. Now, the proposed training algorithm is going to be presented. Unlike the training algorithm proposed in [41], the proposed algorithm in this chapter is based on the SCKF, which is proven to work well with highly nonlinear systems, and has more numerical accuracy [5]. Fig. 4.4 shows the schematic diagram of the training model [1] which is described in details as follows:

   (a) Evaluate the cubature points $i \in \{1, 2, \cdots, m\}$, where $m = 2n_x$. $n_x = n_l(n_u + n_b + n_l + 1) + n_z(n_l + 1)$ is the number of states ($x = [\hat{\theta}^\top \ \hat{\beta}^\top]^\top$) to be estimated.

   (b) Estimate the square root of the predicted error covariance

   $$S_{k|k-1} = \textbf{Tria}([S_{k-1|k-1} \ S_{Q,k-1}]), \tag{4.3}$$

   where $S_{Q,k-1}$ denotes a square root of $Q \in \mathbb{R}^{n_x \times n_x}$ such that $Q = S_{Q,k-1}S_{Q,k-1}^\top$. **Tria** denotes any given triangularization algorithm.

   $$X_{i,k|k-1} = S_{k|k-1}\xi_i + x_{k-1|k-1}, \tag{4.4}$$

   where, $\xi_i = \sqrt{\frac{m}{2}}e_i^{\,2}$ is the cubature point-weight set. $S_{k-1|k-1} \in \mathbb{R}^{n_x \times n_x}$ is the

---

[1]Fig. 4.4 shows the detailed components of the SCKRNN1 and SCKRNN2

[2]$e_i - i$th natural basis vector for $\mathbb{R}^{n_x}$
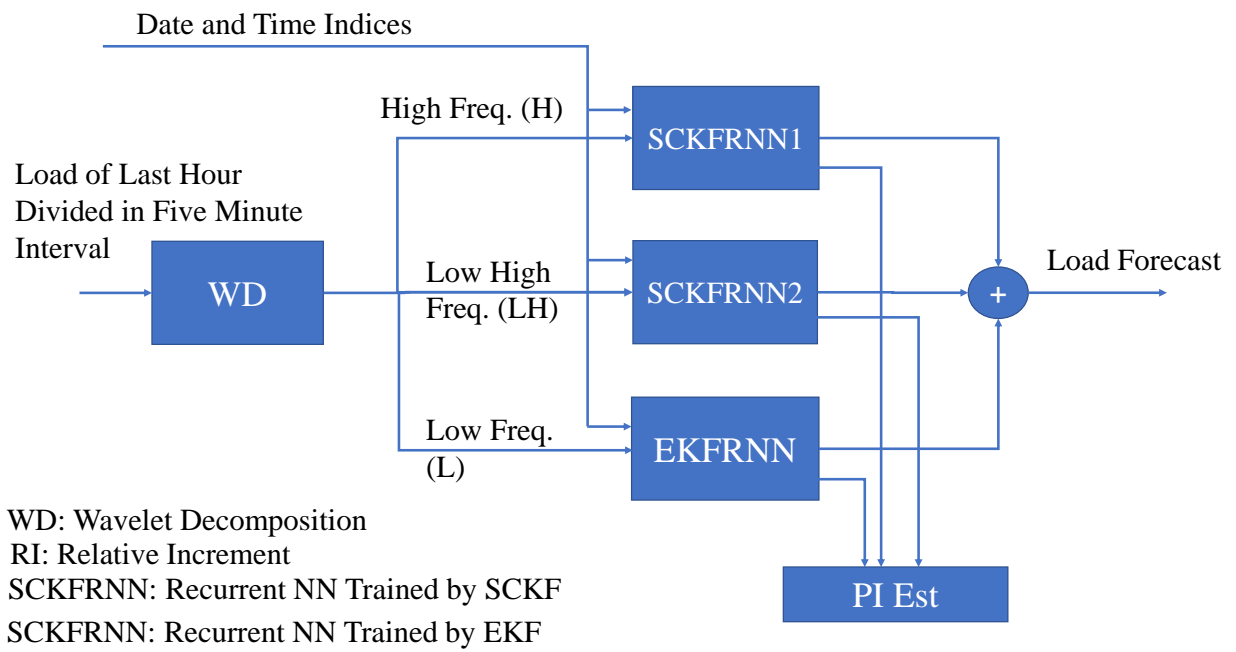
Figure 4.3: Load forecaster configuration
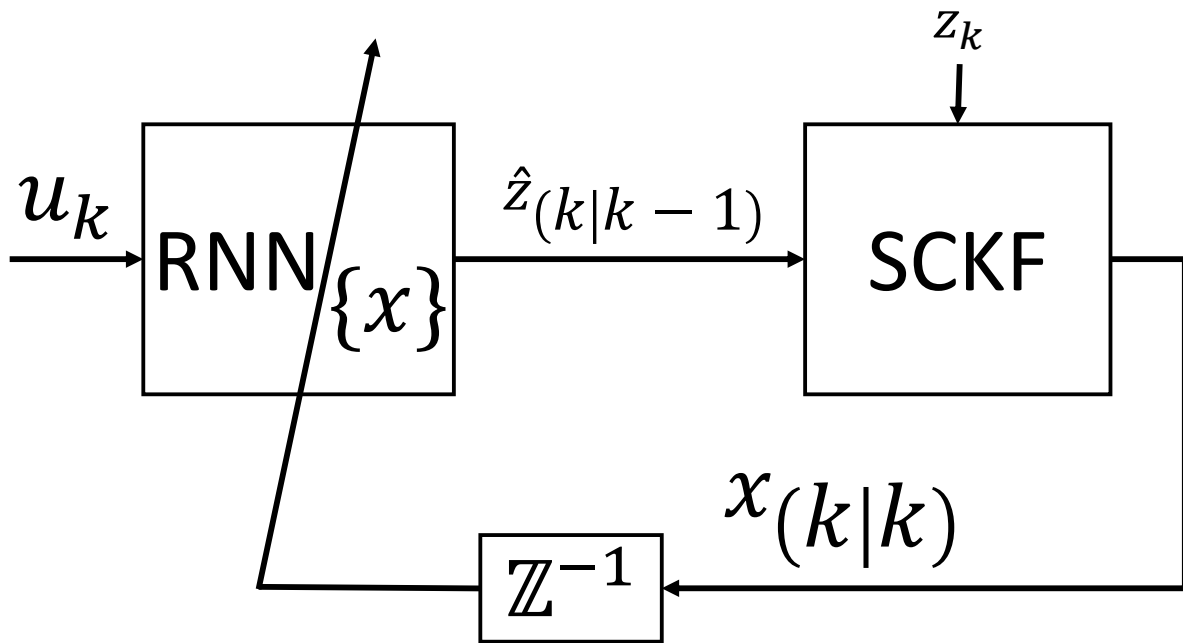
Figure 4.4: Block diagram of the training model

previous estimated covariance matrix.

(c) For each cubature point i calculate the predicted measurement as follows,

$$U1_i = \hat{\theta}_{i,k-1}[u_k^\top \; 1 \; o1_{i,k-1}^\top]^\top \tag{4.5}$$

where, $\hat{\theta}_{i,k-1}$ is the first n_l rows of the $X_{i,k-1|k-1}$ [3].

(d) Calculate the hidden layer as follows,

$$o1_{i,k} = \frac{1 - exp(-2U1_i)}{1 + exp(-2U1_i)} \tag{4.6}$$

(e) The final output is calculated, for each cubature point i as follows,

$$Z_i = \hat{\beta}_{i,k-1}[o1_{i,k}^\top \; 1]^\top \tag{4.7}$$

(f) The following steps are for calculating the Kalman gain and the weight covariance matrix,

$$\hat{z}_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} Z_i, \tag{4.8}$$

$$x_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} X_{i,k|k-1}, \tag{4.9}$$

$$\hat{\Upsilon}_{k|k-1} = \frac{1}{\sqrt{m}}[Z_1 - \hat{z}_{k|k-1} \; Z_2 - \hat{z}_{k|k-1}$$
$$Z_m - \hat{z}_{k|k-1}] \tag{4.10}$$

$$\varkappa_{k|k-1} = \frac{1}{\sqrt{m}}[X_{1,k|k-1} - x_{k|k-1} \; X_{2,k|k-1} - x_{k|k-1}$$
$$X_{m,k|k-1} - x_{k|k-1}] \tag{4.11}$$

(g) Estimate the square root of the innovations covariance matrix

$$S_{zz,k|k-1} = \mathbf{Tria}([\hat{\Upsilon}_{k|k-1} \; S_{R,k}]), \tag{4.12}$$

[3]Please note the u_k used in 4.5 is equal to $u_k^G = \left[\hat{u}_k^{G\top} \; u_{k,2}^\top\right]^\top$. The superscript G is dropped to avoid overloading the notations.

(h) Estimate the cross-covariance matrix

$$P_{xz,k|k-1} = \varkappa_{k|k-1} \hat{\Upsilon}^{\top}_{k|k-1} \tag{4.13}$$

(i) Estimate the square-root cubature Kalman gain [4]

$$K_k = (P_{xz,k|k-1}/S^{\top}_{zz,k|k-1})/S_{zz,k|k-1} \tag{4.14}$$

(j) Estimate the updated state $x_{k|k}$

$$x_{k|k} = x_{k|k-1} + K_k \left(z_k - \hat{z}_{k|k-1} \right), \tag{4.15}$$

(k) Estimate the corresponding error covariance

$$S_{k|k} = \mathbf{Tria}([\varkappa_{k|k-1} - K_k \hat{\Upsilon}_{k|k-1} \quad K_k S_{R,k}]), \tag{4.16}$$

After calculating the normalized predicted measurement for low high, and high frequent data set ($\hat{z}_{k|k-1} \in \mathbb{R}^{n_z}$), This chapter uses the equation 4.17 to denormalize z_1.

$$y_k^{\mathrm{G}} = \hat{z}_{k|k-1}^{\mathrm{G}}(u_{\max}^{\mathrm{G}} - u_{\min}^{\mathrm{G}}) + u_{\min}^{\mathrm{G}} \tag{4.17}$$

In the training algorithm, the superscript G is dropped to avoid overloading the notations. However, the training algorithm described above are used to train the two RNNs for the LH and H component data set. For the prediction interval, this chapter uses $S_{zz,k|k-1}$ as we will see in the following subsection.

4.3.2 Prediction Interval Estimation

The overall variance estimate is derived in this subsection, to estimate prediction intervals online for the PG. As shown in Fig. 4.3, the key idea is to use an overall estimate of the variance obtained by adding together the three estimates from EKFNN , SCKFNN1 , and SCKFNN2 . This is because these components are orthogonal based on the wavelet

---

[4]/ is used to denote the matrix right division operator, which applies the back substitution algorithm for an upper triangular matrix S and the forward substitution algorithm for a lower triangular matrix S

theory. To obtain the individual variance estimates, the diagonal elements of the innovation covariance $(S_{zz,k|k-1})$ for H, LH, and LL components should be de-normalized individually.

The overall variance estimate is obtained by combining the three estimated variances as follows:

$$\sigma_k^2 = \sigma_{k,L}^2 + \sigma_{k,LH}^2 + \sigma_{k,H}^2 \tag{4.18}$$

σ_k^2 is the overall variance. $\sigma_{k,L}^2$, $\sigma_{k,LH}^2$, and $\sigma_{k,H}^2$ are the variant of the L, LH and H component data set. $\sigma_{k,LH}^2$, and $\sigma_{k,H}^2$ are calculated as follows,

$$\sigma_{k,G}^2 = (u_{\max}^G - u_{\min}^G) \, diag(S_{zz,k|k-1}) \tag{4.19}$$

The predictive data base $(z_k = y_k^L + y_k^H + y_k^{LH}$ and $\sigma_k^2)$ obtained from the proposed VSTLF is used by the proposed SE, explained in Section 4.4, to forecast the distribution system states. Forecasting the distribution system states can be used by the operator for assessing the situation of the system or/and short term planing.

4.4   The Proposed Filter Scheme

The objective is to estimate the distribution system bus voltage magnitudes and angles of the all buses of the distribution system network. The proposed SCKF technique recursively updates the estimates using the previous estimated value from time $k-1$ and the pseudo measurements at time k to provide the final estimation of the bus voltage magnitudes and angles at time instant k.

At time instant k , a measurement vector $z \in \mathbb{R}^{n_z}$ is given. $z \in \mathbb{R}^{n_z}$ is the vector of pseudo-measurements provided by the proposed PG. n_z is the total number of power measurements. In compact notation, the set of measurement equations that include the injected active and reactive power can be written as follows:

$$z = h\left(x\right) + v \tag{4.20}$$

where, $h(x)$ is a non-linear function (power flow equation relating the injected power to the voltage magnitudes and angles). v is the measurement noise, with a Gaussian distribution.

$x \in \mathbb{R}^{n_x}$ is is the vector of the state variables. n_x is the total number of state variables. The states that going to be estimated are the voltage magnitude and angle at each bus $x := \left[|\mathbf{v}_1| \cdots |\mathbf{v}_{\frac{n_x}{2}}| \quad \delta_1 \cdots \delta_{\frac{n_x}{2}} \right]^\top$.

The SCKF is developed using the least-squares method for the cubature Kalman gain and matrix triangular factorization or triangularization for the covariance updates. The least-squares method avoids the explicit computation of matrix inversion, whereas triangularization essentially computes a triangular square-root factor of the covariance without square rooting a squared-matrix form of the covariance. The SCKF essentially propagates the square roots of the predictive and posterior error covariances and offers the following benefits as compared to the CKF introduced in [92]:

1. Preservation of symmetry and positive (semi) definiteness [5]; and

2. Numerical accuracy enhancment [5].

Please note that to estimate the distribution system states x, this chapter uses the SCKF algorithm described in Section 4.3, *mutatis mutandis*, using distribution system network information (i.e,. instead of using the SCKF to estimate the weights of the RNN utilizing the training data, here this chapter uses the pseudo measurement to estimate the distribution system network states). The following points (1, 2, 3, and 4) will highlight the changes made to the algorithm described in Section 4.3 due to the change of the model (distribution system network rather than RNN).

At each time instant k, we perform the following:
First, the predicted state vector is computed, using the transition model presented in [119]. The steps for calculating the predicted state vector are as follows:

1. For each cubature point $i \in \{1, 2, \cdots, m\}$, calculate the following:

$$X_{i,k-1|k-1} = \hat{S}_{k-1|k-1}\xi_i + \tilde{x}_{k-1|k-1}, \tag{4.21}$$

where, $\hat{S}_{k-1|k-1} \in \mathbb{R}^{n_x \times n_x}$ is the previous estimated covariance matrix. Here, n_x is the number of distribution system network states, which were defined previously. As one may notice from 4.21 that the predation model is highly dependent on the the forecasting data. This is usually the case in all the prediction model available in the literature since modeling the distribution system dynamic is very hard to obtain. therefore, the enhancement of the accuracy of the forecast data can enhance the state prediction accuracy.

2. compute the predicted state vector for each propagated cubature point i.

$$\hat{X}_{i,k|k-1} = X_{i,k-1|k-1} - H^{-1}(X_{i,k-1|k-1})$$
$$\left(Y_{k|k-1} - Y_{k-1|k-1}\right) \tag{4.22}$$

This chapter uses the transition model presented in [119]. The matrix $H^{-1} \in \mathbb{R}^{n_x \times n_z} = \partial h(x)/\partial x$ is the Jacobin matrix evaluated at $X_{i,k-1|k-1}$. $\hat{X}_{i,k|k-1} \in \mathbb{R}^{n_x}$ is the predicted state vector at cubature point i. $Y_{k-1|k-1}$ and $Y_{k|k-1}$ are respectively the forecast injected power at time $k-1$ and k.

3. Using $\hat{X}_{i,k|k-1}$, we compute,

   a The predicted measurement as follows,

   $$\hat{Z}_{i,k|k-1} = h(\hat{X}_{i,k|k-1}), \tag{4.23}$$

   b The predicted states using 4.9.

4. Using $\hat{X}_{i,k|k-1}$ instead of $X_{i,k|k-1}$, and $\hat{Z}_{i,k|k-1}$ instead of Z_i, the same steps described in Section 4.3 are followed to obtain the estimated states.

4.4.1 Integration of the Proposed PG and SE

In this subsection, the integration (closed loop) between the proposed PG and the proposed SE is explained. Two values are fed from the proposed PG to the proposed SE. One value is the forecast value of the load which is calculate as follows,

$$y_k = y_k^{\mathrm{L}} + y_k^{\mathrm{LH}} + y_k^{\mathrm{H}} \tag{4.24}$$

The y_k is measurement point for the proposed SE. It also will be used as $Y_{k|k-1}$, and the previous forecast value (y_{k-1}) will be used as $Y_{k-1|k-1}$ in 4.22. The second value is the variance which is calculated by 4.18. The variance of each measurement point at time k is used to build the measurement covariance matrix R of the proposed SE.

4.5   Simulation Results and Discussion

In this section, the results of the proposed forecasting aided state estimation algorithm is demonstrated. First, the test system (IEEE 123-bus system) shown in Fig. 4.5, which is used to evaluate the performance of the proposed SE, is described. Moreover, the load characteristics used is also described. Second, the performance of the PG is compared against the most commonly used VSTLF. Finally, the results of the proposed SE when used with the proposed PG, against the results of the proposed SE when used with most commonly used VSTLF, is compared. The underlying network topology is assumed to be known a priori and fixed over the time duration of state estimation.

4.5.1   Test System and Load Characteristics

The IEEE 123-bus test system shown in Fig. 4.5 is used to evaluate the proposed forecasting aided state estimation algorithm, as we will see in Section 4.5.3. Each bus is assumed to be equipped with the proposed VSTLF to generate the measurement. In case of presence of the real data the configuration used in the previous chapter can be used. The test network is a 10 kV (where each MV node corresponds to a secondary transformer 10:0.4 kV) system with a weakly-meshed structure. This network has a peak demand of 13.8 MW, which is primarily made up of suburban/rural residential customers. The proposed algorithm were implemented in MATLAB and time the execution of each iteration on a 3.6 GHz i7-7700 CPU. The best results were obtained using a feed forward neural network (FNN) comprised of an input layer with 57 neurons ($n_u = 12$ and $n_b = 45$). n_u inputs represent the last hour load in five minute intervals (one for each five minute load value). One hidden layer with ten neurons ($n_l = 10$), and an output layer with 12 neurons ($n_z = 12$). n_z outputs represent
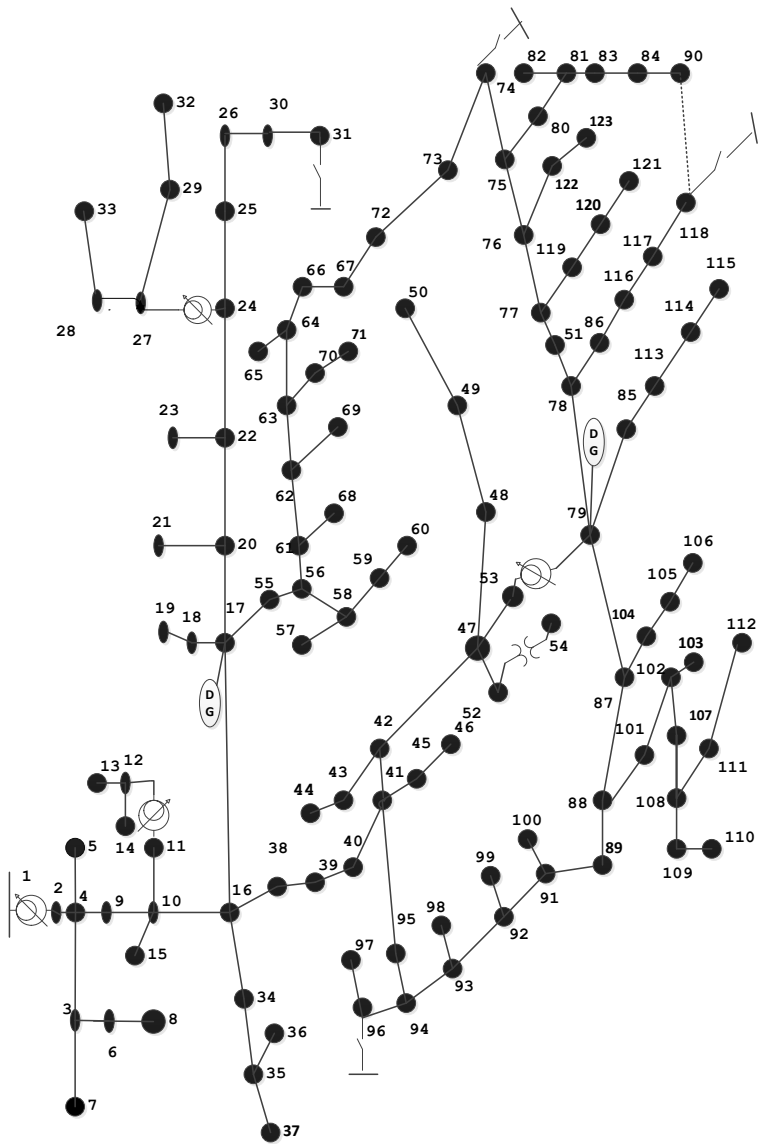
Figure 4.5: IEEE 123-bus system

the forecast hour load in five minute intervals (one for each five minute load value). The best results were obtained using a RNN comprised the same structure of FNN except that the RNN has an extra n_l input neurons in the input layer as explained in Section 4.3.

4.5.2 Proposed Pseudo-measurement Generator Results and Evaluation

The training period is from January 1, 2007 to December 31, 2007, the validation is from January 1, 2008 to June 30, 2008, and the test is from July 1, 2008 to December 31, 2008. The data can be obtained from http://github.com/ldmbouge/vstlf. The mean average percentage error (MAPE), mean absolute error (MAE), and estimated standard deviation (ESD) are used to assess the performance of the proposed PG. Table 4.1, and 4.2 respectively show the results of validation, and testing data for the proposed PG based FNN. Table 4.1, and 4.2 compare the performance of the proposed PG based FNN (WNNHKF2) with one proposed in [41] (WNNHKF1) based FNN. The proposed method in [41] used the UKF as a training algorithm for both LH and H component data sets. Unlike our proposed PG, the proposed algorithm in [41] uses the FNN. One can notice that, in terms of accuracy, the best performance is obtained by the proposed PG. The proposed PG can enhance the accuracy of forecasting by approximately 22% to 42%, compared to the proposed VSTLF proposed in [41]. Accuracy enhancement in the forecasting data can potentially enhance the accuracy of the estimated states as it will shown in 4.5.3.

Please note that the training, validation, and testing data sets used to perform the above comparison are based on an aggregated load data available in http://github.com/ldmbouge/vstlf. Table 4.3, and 4.4 respectively show the results of validation, and testing data for the proposed PG based RNN. Table 4.3, and 4.4 compare the performance of the proposed PG based RNN (WNNHKF4) with one proposed in [41] (WNNHKF3) based RNN. One can notice that even if RNN with proposed training algorithm in [41] is used, our proposed algorithm still show a better performance enhancement by approximately 22% to 42%. Updating the type of NN from FNN to RNN can enhance the accuracy by 8 to 20 %.

From Table 4.1, and 4.2, one can notice that as the forecasting step increases (from 5- to 60-min-ahead), the three metric values gradually increase as shown in Fig. 4.8. This

Table 4.1: MAPES (%), MAES (MW), and ESD (MW) for WNNHKF1 and WNNHKF2
methods (based on validation data set) based FNN

	WNNHKF1			WNNHKF2		
Min.	MAPE	MAE	ESD	MAPE	MAE	ESD
5	0.12	17.22	22.79	0.09	13.90	18.55
10	0.17	25.48	35.60	0.13	18.62	29.41
15	0.21	31.64	49.29	0.16	24.36	37.50
20	0.26	37.70	55.62	0.20	28.03	43.33
25	0.29	42.98	61.19	0.22	32.10	46.12
30	0.33	50.12	75.70	0.25	39.60	56.30
35	0.36	54.16	81.43	0.28	40.70	61.70
40	0.40	60.38	97.32	0.31	47.50	75.44
45	0.44	65.98	102.60	0.34	49.81	77.30
50	0.48	72.12	107.60	0.37	53.53	83.10
55	0.51	76.52	112.53	0.39	58.92	85.64
60	0.55	82.76	130.08	0.42	62.73	101.16

Table 4.2: MAPES (%), MAES (MW), and ESD (MW) for WNNHKF1 and WNNHKF2 methods (based on test data set) based FNN

	WNNHKF1			WNNHKF2		
Min.	MAPE	MAE	ESD	MAPE	MAE	ESD
5	0.13	19.39	25.68	0.10	15.93	20.77
10	0.18	27.33	38.28	0.14	20.04	28.48
15	0.22	32.86	51.60	0.17	26.30	39.73
20	0.26	39.01	57.40	0.21	30.04	46.20
25	0.30	44.87	62.04	0.23	36.55	46.77
30	0.34	50.97	76.11	0.26	36.25	59.61
35	0.38	56.93	81.14	0.29	43.84	61.48
40	0.42	63.30	96.58	0.32	47.74	75.37
45	0.46	69.01	100.99	0.35	54.14	76.76
50	0.50	75.46	105.52	0.39	58.90	83.25
55	0.54	81.09	110.29	0.43	62.44	84.92
60	0.58	87.43	128.36	0.46	66.32	98.84

Table 4.3: MAPES (%), MAES (MW), and ESD (MW) for WNNHKF3 and WNNHKF4 methods (based on validation data set) based RNN

	WNNHKF3			WNNHKF4		
Min.	MAPE	MAE	ESD	MAPE	MAE	ESD
5	0.11	16.79	21.90	0.08	12.54	17.27
10	0.16	24.37	32.65	0.12	17.97	23.85
15	0.19	29.01	43.20	0.14	20.20	33.92
20	0.24	34.57	50.00	0.17	24.26	38.27
25	0.27	36.41	55.11	0.19	29.80	40.03
30	0.31	44.96	68.42	0.22	31.58	51.72
35	0.33	47.67	75.67	0.24	37.29	52.57
40	0.37	53.37	88.24	0.27	41.96	66.20
45	0.40	61.50	93.08	0.30	45.21	67.74
50	0.44	65.13	98.67	0.32	49.32	72.10
55	0.47	71.17	104.19	0.34	50.27	72.40
60	0.50	73.89	117.28	0.37	55.45	88.16

Table 4.4: MAPES (%), MAES (MW), and ESD (MW) for WNNHKF3 and WNNHKF4 methods (based on test data set) based RNN

	WNNHKF3			WNNHKF4		
Min.	MAPE	MAE	ESD	MAPE	MAE	ESD
5	0.11	17.80	22.55	0.09	14.43	18.21
10	0.17	26.06	34.10	0.12	19.31	26.65
15	0.20	31.13	45.32	0.15	22.01	34.57
20	0.24	33.77	52.64	0.17	27.14	38.96
25	0.28	40.15	55.89	0.20	31.06	42.57
30	0.31	44.74	70.79	0.23	34.15	51.90
35	0.35	51.21	75.41	0.26	39.14	55.36
40	0.39	57.05	87.56	0.28	43.41	66.71
45	0.42	61.28	93.61	0.31	45.24	68.66
50	0.46	66.20	96.76	0.34	51.56	70.70
55	0.50	72.36	101.14	0.36	53.33	73.90
60	0.53	83.17	116.91	0.39	59.58	86.00
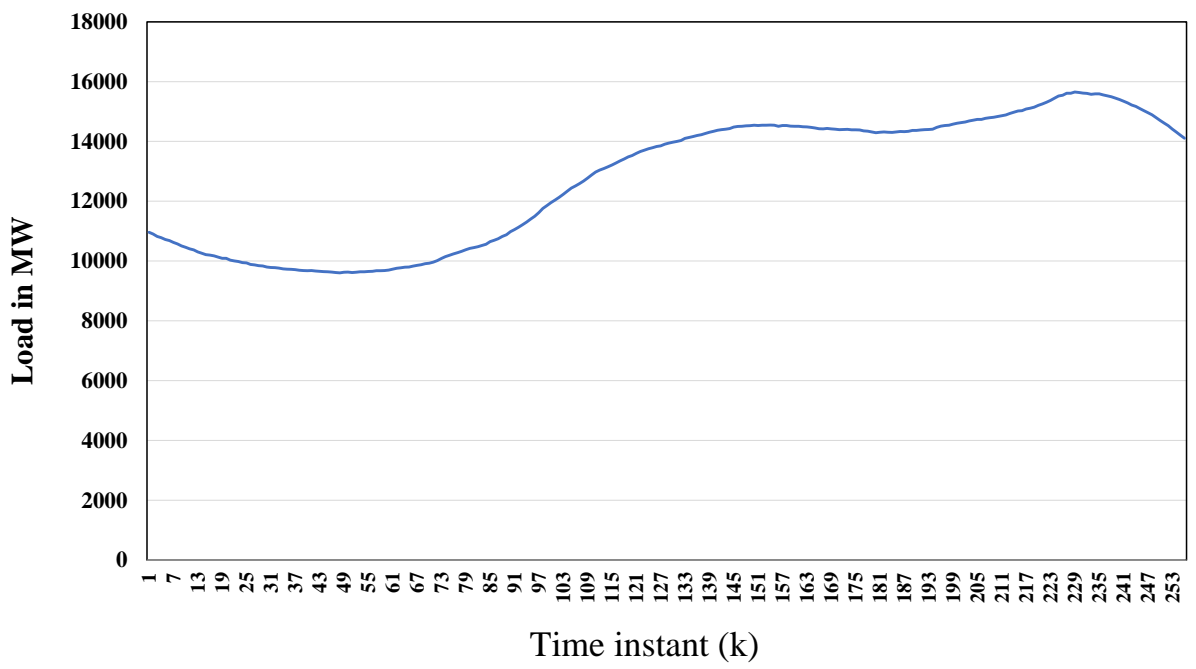
Figure 4.6: Samples from the aggregated load
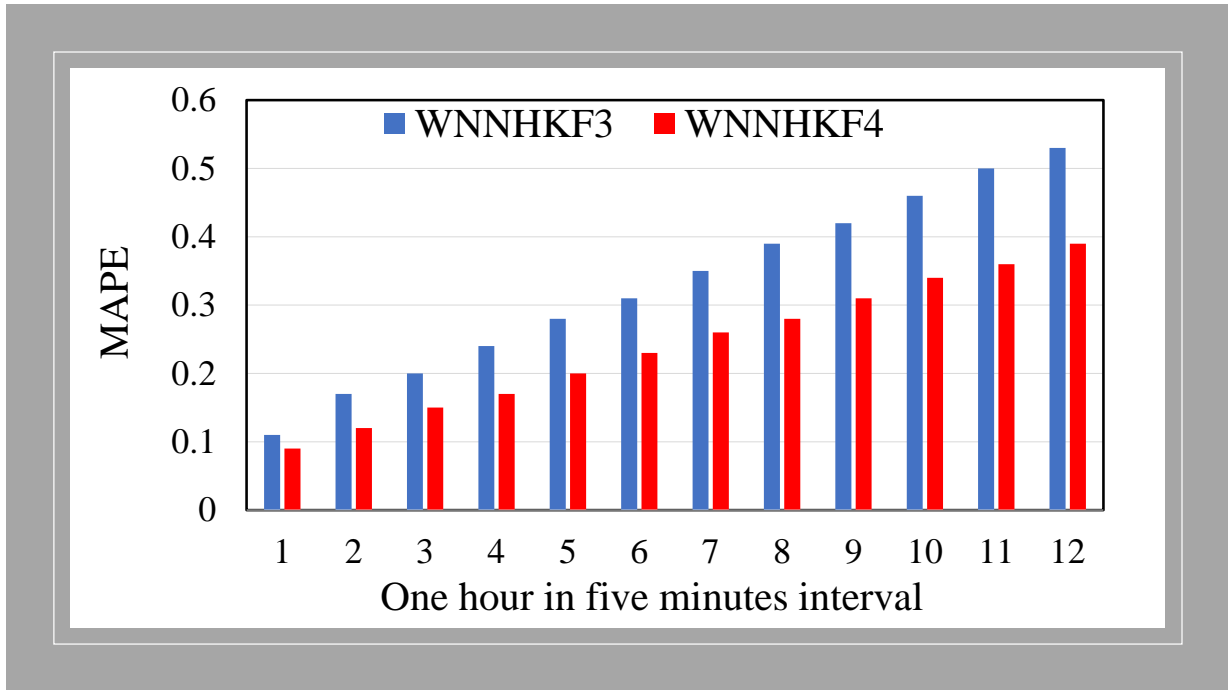
Figure 4.7: MAPE values for the testing data

is because the uncertainty expands as the forecasting step increases. The metric values in Table 4.1 (using validation data), are very close to the ones in 4.2 (using test data). This indicates that WNNHKF1 and WNNHKF2 parameters are properly selected. All the metric values quantify forecasting accuracy in a certain way, with the last one directly related to PIs. One can notice from Table 4.1, and 4.2, that the ESDs have values from 22 MW to 131 MW, and ISO-NEs system load data have values around 15000 MW as shown in Fig. 4.6.

4.5.3 The Proposed State Estimator Results and Evaluation

This chapter assumes that the sampling period T = 5 min. The load data for each bus in medium voltage distribution system used to assess the proposed SE was obtained by modifying the aggregated load used in the previous section. Since the load data used in the previous section is for aggregated load at transmission level, the data is

scaled down by dividing by $160,000$ to be appropriate for the distribution system in the medium voltage level. This modification is reasonable as when the modified load curve is compared with typical load of point in medium voltage distribution system available in http://smarthg.di.uniroma1.it/test-beds-data-anlytics-charts/, one can notice that the two aforementioned curves have the same shape features [5]. Since no recorded data were available for the reactive power at each MV node, the measured reactive power demand at the primary transformer was used to scale reactive power demands at each MV node so that the total reactive power demand matches the recorded value at the primary. As would be expected for an MV network serving mainly residential load, power factors are generally very close to unity, but there was some variation between 0.8 and 1.0, particularly during the early morning hours. The load flow results at each time instant using true testing data values are calculated and used as a benchmark to evaluate the accuracy of the proposed algorithms. To evaluate the accuracy of the proposed SEs, several tests were performed using the test data. For each bus, a set of sequential true data samples is generated using the test data as follows,

$$z^i \sim \mathcal{N}(\mu, \sigma_1) \tag{4.25}$$

where z_i is the true value of the load point at bus i. μ is the true value obtained from the test data set and σ_1 equal to 0.001.

It is worth noting that in practical each bus is equipped with the proposed VSTLF to generate the measurements. The RNN used in the proposed VSTLF is trained using the historical data available at that particular bus. The training stage should be done every two or three months as the load behavior is expected to change after that period. The main advantage of the RNN is that can be trained using less data compared to FNN.

The Gaussian pseudo-measurements with mean z_i and σ equivalent to the corresponding error of the proposed forecasting algorithm and the forecasting algorithm proposed in [41].

---

[5]We were not able to get real typical data for a point in the distribution system, this is why that mild modification is made. The main contribution in this chapter is proposing the new forecasting algorithm and SE algorithm and integrating the two proposed algorithms. However, the proposed forecasting algorithm can work more efficiently for data for a point in the distribution system, as this load data has more higher frequency component

The maximum values of the relative errors in voltage magnitudes (MRE) is used to assess the performance of the proposed SE:

$$\text{MRE}^i = \frac{|V_i - \hat{V}_i|}{|V_i|} \text{ and} \tag{4.26}$$

where, $\hat{V}_i$ is the estimated voltage magnitude at bus i, V_i is the true value.

Table 4.5 shows the comparison between the proposed SE when used with the proposed PG (Case 2), and the results of the proposed SE when used with one proposed in [41] (Case 1). The metrics MRE_{max} is computed as follows:

$$\text{MRE}_{\text{max}} = \max_{i \in \overline{ZO}_l} \text{MRE}^i_{\text{max}} \tag{4.27}$$

$$\text{MRE}_{\text{ave}} = \frac{1}{|\overline{ZO}_l|} \sum_{i \in \overline{ZO}_l} \text{MRE}^i_{\text{max}} \tag{4.28}$$

where, $|\overline{ZO}_l|$ represents the cardinality of the IEEE 123-bus system (i.e., the number of buses in the IEEE 123-bus system). The $\text{MRE}^i_{\text{max}}$ is computed as following,

$$\text{MRE}^i_{\text{max}} = \max_{l \leq M_c} (\text{MRE}^i_l) \tag{4.29}$$

where M_c is the number of the test data.

Table 4.5 shows the comparison between case 1 (the proposed SE when used with one proposed method in [41]) and case 2 (the proposed SE, when used with the proposed PG). Case 2 has a performance enhancement of approximately 4% to 7%, compared to the results of the case 1.

4.6   Conclusion

This chapter presents a forecasting aided state estimation algorithm which allows the operator to forecast the system states for a one hour horizon in five minute intervals.

Table 4.5: Performance comparison between WNNHKF1 and WNNHKF4 method (based on test data set)

Min.	CASE1		CASE2	
	$\mathrm{MRE_{max}}$	$\mathrm{MRE_{ave}}$	$\mathrm{MRE_{max}}$	$\mathrm{MRE_{ave}}$
5	0.52×10^{-3}	0.45×10^{-3}	0.34×10^{-3}	0.29×10^{-3}
10	0.72×10^{-3}	0.62×10^{-3}	0.47×10^{-3}	0.38×10^{-3}
15	0.87×10^{-3}	0.75×10^{-3}	0.58×10^{-3}	0.44×10^{-3}
20	1.03×10^{-3}	0.93×10^{-3}	0.69×10^{-3}	0.57×10^{-3}
25	1.19×10^{-3}	1.05×10^{-3}	0.79×10^{-3}	0.68×10^{-3}
30	1.35×10^{-3}	1.18×10^{-3}	0.90×10^{-3}	0.82×10^{-3}
35	1.51×10^{-3}	1.34×10^{-3}	1.01×10^{-3}	0.94×10^{-3}
40	1.66×10^{-3}	1.53×10^{-3}	1.11×10^{-3}	1.01×10^{-3}
45	1.82×10^{-3}	1.71×10^{-3}	1.22×10^{-3}	1.07×10^{-3}
50	2.00×10^{-3}	1.92×10^{-3}	1.32×10^{-3}	1.13×10^{-3}
55	2.14×10^{-3}	2.05×10^{-3}	1.43×10^{-3}	1.23×10^{-3}
60	2.30×10^{-3}	2.11×10^{-3}	1.54×10^{-3}	1.39×10^{-3}
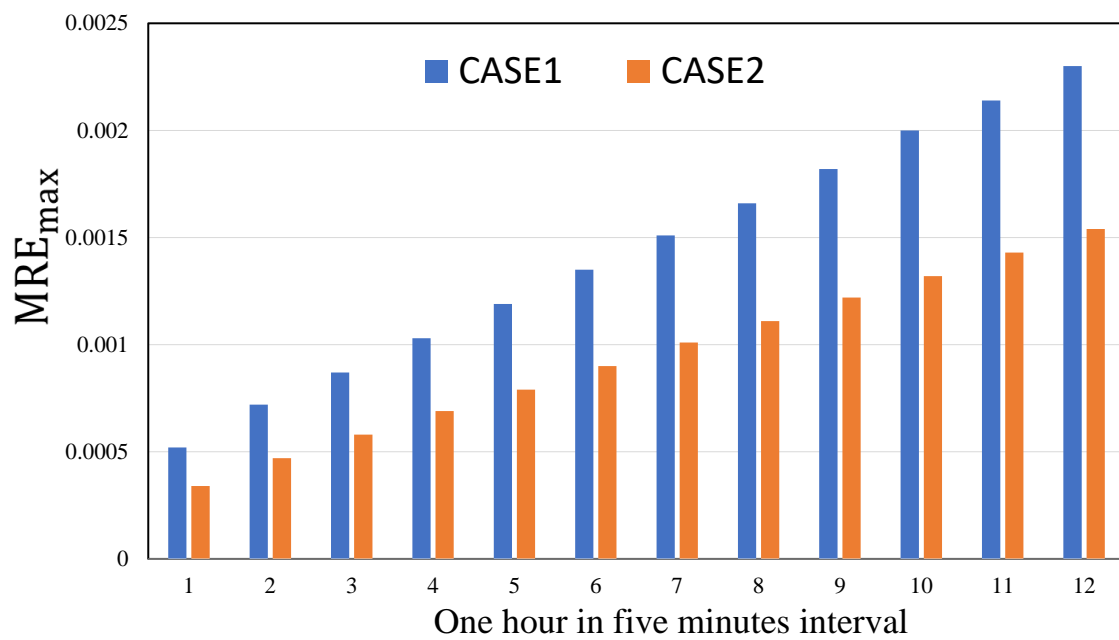
Figure 4.8: MRE_{max} values for the testing data. CASE1 is based on load forecaster trained by UKF, while CASE2 is load forecaster trained by SCKF (the proposed algorithm) .

The PG is designed utilizing the strength of the RNN. Three RNNs are trained using three components of the data set which are low frequency, low-high frequency and high frequency data set. The three components of the data set are obtained using the wavelet decomposer. The training algorithm for the RNN with low frequency data set is performed using the EKF as the low frequency data set shows a nearly linear relationship. The training algorithm for the RNN with low-high and high frequency data set is performed using the SCKF as the low-high, and high frequent data set shows a non-linear relationship. This chapter compares the results of the proposed SE when used with the proposed PG, against the results of the proposed SE when used with other commonly used VSTLF. The proposed SE when used with the proposed PG were applied to an IEEE 123-bus system, and the simulation results demonstrated a superior performance in terms of accuracy compared to the proposed SE when used with other commonly used VSTLF.

Chapter 5

Mitigation of Cyber-Physical Attacks in Multi-Area Automatic Generation Control

5.1 Introduction

In this chapter new attack resilient scheme for single and multi-area AGC systems is developed. The proposed approach detects and identifies the sensors under attack on single area AGC systems leveraging the results presented in [29] and [78]. Once an attack has been identified, a simple switching scheme is employed to ensure that the AGC feedback loop continues to make its control decisions using uncompromised sensor data. We also characterize the degree of sensor redundancy needed in order to implement a robust solution and ensure uninterrupted service. The above ideas is extended to a multi-area AGC framework where the frequency and tie-line sensors may be attacked. s-sparse observability analysis is provided for AGC in single area and multi-area power networks. A numerical simulations demonstrating the effectiveness of the proposed approach in detecting and mitigating attacks as well as the feasibility of real-time implementation are provided.

5.2 Nomenclature

Variables

x_1	Steam valve position command.
x_2	Change in mechanical power.
x_3	Frequency change
$d(t)$	Load change
K_P	Proportional gain
K_I	Integral gain
w	Measurement noise
$e[k]$	Attack on the sensors at time $t = kT$
$x(kT)$	Value of the state at time kT
$x_c[k]$	Controller state at time kT
T	Sampling period
$\mathbf{y}$	Information available for feedback
$\mathbf{E}^*[k]$	Actual attack vector
ΔP_{ij}	Deviation from the scheduled exchanges between areas i and j
$\Delta \omega_i$	Deviation from the from the nominal frequency value for area i

5.3 Problem Formulation for a Single AGC

We start by deriving a continuous-time state-space representation of the linearised model of a single area AGC system. The conventional control architecture based on the linearised model is used. The state-space variable $x_1 := \Delta P_v$ equals the steam valve position command. The variable x_2 equals ΔP_m, the change in mechanical power. Finally, the state x_3 equals the frequency change $\Delta \omega$ of the power system network. The state vector of the AGC system is $x := (x_1, x_2, x_3) \in \mathbb{R}^3$. The exogenous signals are the control input $u(t)$ which physically represents the generation difference and the disturbance $d(t)$ which physically represents load change $\Delta P_L(t)$. With these definitions the linearised AGC dynamics can be expressed in state-space form as

$$\dot{x}(t) = Ax(t) + Bu(t) + Ed(t). \tag{5.1}$$

Following [86], the matrices in (5.1) are given by

$$A = \begin{bmatrix} -1/\tau_g & 0 & 0 \\ 1/\tau_T & -1/\tau_T & 0 \\ 0 & 1/2H & -D/2H \end{bmatrix}, \quad B = \begin{bmatrix} 1/\tau_g \\ 0 \\ 0 \end{bmatrix} \tag{5.2}$$

and $E^\top = \begin{bmatrix} 0 & 0 & 1/2H \end{bmatrix}$. The physical constant H is the total generator inertia, τ_g is the governor time constant, τ_T is the turbine time constant and D is the load damping constant. The signal available for feedback is $x_3(t)$ which is also the variable that must be regulated. Therefore, the system output is taken to be

$$y(t) = Cx(t) + w(t) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} x(t) + w(t) \tag{5.3}$$

where $w \in \mathbb{R}$ represents measurement noise and is assumed to take values in a *known* compact subset of the output space $\mathbb{R}$. In other words, there exists a known constant $\delta_w > 0$ such that for all $t \geq 0$, $|w(t)| \leq \delta_w$. Upon receiving the frequency measurement, the controller adjusts the valve position to maintain the network frequency at its nominal value. A commonly used control law is proportional-integral (PI) control [86] given by $u(t) = K_P y(t) + K_I \int_0^t y(\tau) \, d\tau$, where K_P, K_I are, respectively, the proportional and integral gains. It is common to express the proportional gain as $K_P = 1/R$, where R represents the droop constant.

5.3.1 Discrete-time Model

In practice, the controller is often designed for the continuous-time dynamics (5.1), (5.3) and then descritized for implementation. Assuming an ideal sample and an ideal zero-order hold at a sampling period of $T > 0$ seconds, and further assuming that $d(t) = \Delta P_L(t)$, and $w(t)$ are constant over a sampling period, the AGC evolves in discrete time according to the dynamics,

$$x[k+1] = A_d x[k] + B_d u[k] + E_d d[k]$$
$$y[k] = C_d x[k] + w[k] \tag{5.4}$$

where $x[k] := x(kT)$ is the value of the state at time kT, $k \in \mathbb{Z}$. Similarly $y[k] := y(kT)$, $d[k] := d(kT)$, $w[k] := w(kT)$. The matrices in (5.4) are given by $A_d = \mathrm{e}^{AT}$, $B_d = \int_0^T \mathrm{e}^{A\tau} B \, \mathrm{d}\tau$, $E_d = \int_0^T e^{A\tau} E \, \mathrm{d}\tau$ and $C_d = C$. We assume that the sampling isn't pathological so that the pairs (A_d, B_d), (C_d, A_d) remain, respectively, controllable and observable. The discretized PI controller in state-space form is given by

$$x_c[k+1] = x_c[k] + Ty[k]$$
$$u[k] = K_P y[k] + K_I x_c[k] \tag{5.5}$$

where $x_c[k]$ is the controller state at time kT. During normal operation, thanks to the integral action in the controller, the single area AGC system is able to reject piecewise constant disturbances while keeping the system frequency at its nominal value.

5.3.2 Sensor Redundancy and Attack Model

As mentioned in the introduction, in this chapter the AGC is assumed to be equipped with more than one frequency sensor. This redundancy will be used to detect cyber-physical attacks and take appropriate control action. As such, we no longer assume that the measured output is 1-dimensional as in (5.4), but rather that there are p frequency sensors. Let $\mathbf{y} \in \mathbb{R}^p$ denote the information available for feedback using redundant sensors. As in the single output case (5.4), each sensor is assumed to be corrupted by measurement noise which, with mild abuse of notation, we denote as a vector $w \in \mathbb{R}^p$. We continue to

assume that there is a known constant $\delta_w > 0$ such that[1] for all $k \in \mathbb{Z}$, $\|w[k]\|_\infty \leq \delta_w$. To model a cyber-physical attack, we let $e \in \mathbb{R}^p$ denote an *attack vector* so that $e[k]$ represents an attack on the sensors at time $t = kT$. This attack vector appears as an additive disturbance to the measured variables so that the AGC model, with redundant measurements and attacks, is given by

$$
\begin{aligned}
x[k+1] &= A_d x[k] + B_d u[k] + E_d d[k] \\
\mathbf{y}[k] &= \mathbf{C} x[k] + w[k] + e[k]
\end{aligned}
\tag{5.6}
$$

where $\mathbf{C} := (1_p \otimes C_d)$, $\otimes$ denotes the Kronecker product and $1_p \in \mathbb{R}^p$ is the vector of all ones. The first problem considered in this chapter, under the assumption that we have sufficient redundancy in the sensors, is to identify which sensors have been corrupted by an attack and then use the non-attacked measurements as inputs to the control law (5.5) in order to ensure reliable operation of the AGC. In Section 5.6 we extend this idea to multi-area automatic generation control.

5.4   Attack Detection

As stated in the introduction, we leverage the results in [29] and [78] to detect attacks on AGC systems. In this section, we summarize how the attack detection scheme works in the context of an AGC. We start by determining the maximum number of sensors that can be simultaneously attacked while still retaining the ability to detect the attacks when there is no noise and no load change. To do this we need some notation. If $\mathbb{N}_p := \{1, \ldots, p\}$ and $\mathcal{K} \subseteq \mathbb{N}_p$, then $\mathcal{K}^c := \mathbb{N}_p \backslash \mathcal{K}$ denotes the complement of $\mathcal{K}$ in $\mathbb{N}_p$. Let $P_\mathcal{K} : \mathbb{R}^p \to \mathbb{R}^{|\mathcal{K}|}$ be the linear map which takes a vector $y \in \mathbb{R}^p$ and removes the rows in $\mathcal{K}^c$.

Definition 5.4.1 ([96]). *A pair* $(\mathbf{C}, A)$ *is s-**sparse observable** if, for every set* $\mathcal{K} \subset \mathbb{N}_p$ *of cardinality s, the pair* $(P_{\mathcal{K}^c} \mathbf{C}, A)$ *is observable.*

We assume that the number of compromised sensors in (5.6) is less than or equal to $q_{\max}$, the maximum number of sensors for which the system's state can be recovered when there is no noise and no load changes.

---

[1] If $x \in \mathbb{R}^n$, then $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$.

Table 5.1: Number of frequency sensors p and maximum number of attacked sensors $q_{\max}$.

p	1	2	3	4	5	6	7	8	9	10	11
$q_{\max}$	0	0	1	1	2	2	3	3	4	4	5

Lemma 5.4.2 ([29])**.** *The maximum number $q_{\max}$ of attacked sensors for which the state of (5.6) can be recovered when there is no noise and no disturbance equals the largest $s \in \mathbb{N}_p$ for which the pair $(\mathbf{C}, A_d)$ is $2s$-sparse observable.*

From this lemma we can immediately quantify the minimum degree of redundancy needed for frequency control of an AGC.

Proposition 5.4.3. *The maximum number of sensors which can be attacked while retaining the ability to estimate the state of the single AGC system when there is no noise and no load changes equals $\lfloor (p-1)/2 \rfloor$.*

Proof. Observe that the pair (C, A_d) is observable. Therefore the pair $(P_{\mathcal{K}^c}\mathbf{C}, A_d)$ is observable as long as $P_{\mathcal{K}^c}$ isn't equal to the zero map, i.e., $\mathcal{K}^c \neq \varnothing$. This in turn implies that $|\mathcal{K}| \leq p - 1$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result. $\qquad\square$

Table 5.1 gives an interpretation of Proposition 5.4.3. Given an expected maximal number of sensors $q_{\max}$ that can be attacked, we can choose the number of frequency sensors that the AGC should be equipped with. The criteria to define the redundancy of sensors is based on an optimal sensor cost that gives a highest possible robustness.

5.4.1 Identifying Attacked Sensors

The support of the attack vector e at time k is $\operatorname{supp}(e[k]) = \{i \in \mathbb{N}_p : e_i[k] \neq 0\}$, where e_i is the ith component of e. The symbol $\|e[k]\|_0 = |\operatorname{supp}(e[k])|$ denotes the number of non-zero entries in the vector $e[k]$. Let $\mathcal{K} := \bigcup_{k \in \mathbb{Z}} \operatorname{supp}(e[k])$ and, in light of Proposition 5.4.3,

we assume throughout that[2] $|\mathcal{K}| \leq q_{\max} = \lfloor (p-1)/2 \rfloor$. To identify the sensors that have been attacked we collect the last N output measurements and solve an MILP at every time step. The value used for N is naturally taken to be equal to the dimension of the system's state space. Let $\tilde{\mathbf{y}}_i[k] := \left[P_i \tilde{\mathbf{y}}[k-N+1]^\top \quad \cdots \quad P_i \tilde{\mathbf{y}}[k]^\top \right]^\top \in \mathbb{R}^N$ be the vector which maintains the last N sensor measurements from sensor $i \in \mathbb{N}_p$, compensated for the inputs applied during this interval, i.e,

$$\tilde{\mathbf{y}}[k-N+1] := \mathbf{y}[k-N+1]$$

$$\tilde{\mathbf{y}}[n] := \mathbf{y}[n] - \sum_{j=0}^{n-k+N-2} \mathbf{C}A_d^j B_d u[n-j-1],$$

where $n \in \{k-N+2, \ldots, k\}$. Define, for output $i \in \mathbb{N}_p$, $O_i := \left[P_i \mathbf{C} \quad P_i \mathbf{C}A \quad \cdots \quad P_i \mathbf{C}A^{N-1} \right]^\top \in \mathbb{R}^{N \times n}$ and let $\mathbf{e}_i[k] := \left[e_i[k-N+1]^\top \quad \cdots \quad e_i[k]^\top \right]^\top \in \mathbb{R}^N$. The vector $\mathbf{e}_i$ represents the attack values injected into sensor i over the last N time steps. Similarly, define $\mathbf{w}_i[k] := \left[w_i[k-N+1]^\top \quad \cdots \quad w_i[k]^\top \right]^\top \in \mathbb{R}^N$ as the list of measurement noise over the last N time steps at sensor i. Then we can express the last N measurements obtained from output $i \in \mathbb{N}_p$ as

$$\tilde{\mathbf{y}}_i[k] = O_i x[k-N] + \mathbf{e}_i[k] + \mathbf{w}_i[k]. \tag{5.7}$$

Following [79], in this expression we have absorbed the effect of the disturbance d into the measurement noise terms and therefore, if necessary, increasing δ_w so that $\|w[k]\|_\infty \leq \delta_w$ continues to hold. Since the open-loop matrix (5.2) is Hurwitz, this approach isn't overly conservative. Next we stack each of vectors, corresponding to each output, in (5.7) to define $\mathbf{Y}[k] := \left[\tilde{\mathbf{y}}_1[k]^\top \quad \cdots \tilde{\mathbf{y}}_p[k]^\top \right]^\top, \mathbf{E}[k] := \left[\mathbf{e}_1[k]^\top \quad \cdots \mathbf{e}_p[k]^\top \right]^\top, \mathbf{W}[k] := \left[\mathbf{w}_1[k]^\top \quad \cdots \mathbf{w}_p[k]^\top \right]^\top$. Each of these are pN-dimensional real vectors. Finally, define $O := \left[O_1^\top \quad \cdots \quad O_p^\top \right]^\top \in \mathbb{R}^{pN \times n}$ so we can compactly write

$$\mathbf{Y}[k] = Ox[k-N] + \mathbf{E}[k] + \mathbf{W}[k]. \tag{5.8}$$

---

[2]This assumption can be weakened to state that, over any time interval of length N, the number of attacked sensors is less than $q_{\max}$.

By assumption, the measurement noise is uniformly bounded $\|w[k]\|_\infty \leq \delta_w$ and so the vector $\mathbf{W}[k]$ is restricted to the compact 'box' $\Omega := [-\delta_w, \delta_w]^{pN} \subset \mathbb{R}^{pN}$.

Following the notation from [24], define

$$\|\mathbf{E}[k]\|_{2,0} := \sum_{i=1}^{p} \mathbb{I}(\|\mathbf{e}_i[k]\|_2 > 0),$$

$$\|\mathbf{E}[k]\|_{2,1} := \sum_{i=1}^{p} \|\mathbf{e}_i[k]\|_2, \tag{5.9}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. The number $\|\mathbf{E}[k]\|_{2,0}$ equals the number of sensors that have been attacked over the last N time steps while $\|\mathbf{E}[k]\|_{2,1}$ represents the cumulative size of the attacks. Consider the following optimization problem:

Problem 1.

$$\underset{\mathbf{E}, x[k-N]}{\text{minimize:}} \; \|\mathbf{E}[k]\|_{2,0}$$

$$\text{subject to:} \; \mathbf{Y}[k] - Ox[k-N] + \mathbf{E}[k] = \mathbf{W}[k]$$

$$\mathbf{W} \in \Omega.$$

$\triangle$

In the absence of noise, i.e., when $\Omega = 0 \in \mathbb{R}^{pN}$, if the number of attacked sensors $|\mathcal{K}|$ is less than or equal to $q_{\max} = \lfloor (p-1)/2 \rfloor$, then the values of the minimizing decision variables for Problem 1 are the state $x[k-N]$ and the actual attack vector $\mathbf{E}^*[k]$ [29]. In the presence of noise, the minimizing variables aren't necessarily $(x[k-N], \mathbf{E}^*[k])$.

Problem 1 involves combinatorial optimization and can be solved using MILP solvers. However, solving this problem is NP–hard in the general case which limits its use to smaller size systems. We will see that, since our AGC system with $n = 3$ state, $N = 3$ time steps and $p = 3$ frequency sensors, solving Problem 1 is feasible in real-time applications. Another approach is to convexify Problem 1 and solve the following instead:

Problem 2.

$$\underset{\mathbf{E}, x[k-N]}{\text{minimize:}} \; \|\mathbf{E}[k]\|_{2,1}$$

$$\text{subject to:} \; \mathbf{Y}[k] - Ox[k-N] - \mathbf{E}[k] = \mathbf{W}[k]$$

$$\mathbf{W} \in \Omega.$$

$\triangle$

This problem can be solved efficiently but is only effective at detecting relatively large attacks.

5.4.2 Performance

In [78], the authors provide a bound on the error between the true value of $x[k-N]$ and its estimated value for both of the aforementioned optimization problems. Let $(\tilde{x}_{2,0}, \tilde{\mathbf{E}}_{2,0})$ be the minimizing values for the decision variables for Problem 1 and let $(\tilde{x}_{2,1}, \tilde{\mathbf{E}}_{2,1})$ be the minimizing values for the decision variables for Problem 2. Let $\mathbf{E}^*$ denote the true attack vector over the last N time steps. Define the errors for Problem 1 $\Delta x_{2,0} := \tilde{x}_{2,0} - x[k-N]$, $\Delta \mathbf{E}_{2,0} := \tilde{\mathbf{E}}_{2,0} - \mathbf{E}^*$ as well as the errors $\Delta x_{2,1} := \tilde{x}_{2,1} - x[k-N]$, $\Delta \mathbf{E}_{2,1} := \tilde{\mathbf{E}}_{2,1} - \mathbf{E}^*$ the errors for Problem 2. By first computing an error bound on $\Delta x_{2,0}$ and $\Delta x_{2,1}$, one can prove the existence of constants D_j^i, $i \in \{0, 1\}$, $j \in \mathbb{N}_p$ such that if

$$\mathbb{I}\big(\|\tilde{\mathbf{e}}_j[k]\| > D_j^i\big), \tag{5.10}$$

where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in either $\tilde{\mathbf{E}}_{2,0}$ (when $i = 0$) or $\tilde{\mathbf{E}}_{2,1}$ (when $i = 1$), then sensor j has been attacked over the last N time steps. In this chapter we use the integer decision variable $\mathbf{E}$ in Problem 1 as our estimate of which sensors were attacked over the last N time steps. In other words, we will say that sensor j has been attacked over the last N times steps if

$$\mathbb{I}(\|\tilde{\mathbf{e}}_j[k]\|_2 > 0) \tag{5.11}$$

While there is no guarantee that the policy (5.11) will correctly detect the attacked sensor, simulations suggest that this approach is able to identify sensors under attack even when the attacks are small in magnitude (stealthy). In the case where, due to computational efficiency considerations, one instead solves Problem 2, the policy (5.10) (using D_j^1) guarantees no false positive attack detections if the sufficient condition [78, Eqn. (29)]] is satisfied. Unfortunately, it can be shown using the parameters in Table 5.2 that the single AGC system does not satisfy the aforementioned sufficient condition. Therefore, the pol-

Table 5.2: AGC parameters in single area system [101]

D	R	K_I	$H(s)$	$\tau_g(s)$	$\tau_T(s)$
0.8	0.05	7	5	0.2	0.5

icy (5.10) is neither reliable nor effective for stealthy attack[3]. Since the bound computed in [78] is very conservative, due to several applications of the triangular inequality in its derivation, a huge range of stealthy attacks elude the proposed attack detector.

5.5 Attack Mitigation

At each time step k, the proposed attack mitigation strategy for a single AGC system with $p \geq 3$ frequency sensors is the as follows.

1. Fix $N \geq 3$ to be the number of steps over which we aim to detect attacks. The lower bound of 3 comes from the dimension of the AGC's state-space.

2. If $k \geq N - 1$, solve the mixed integer linear program Problem 1. Let $\tilde{\mathbf{E}}_{2,0}[k]$ denote the value of the decision variable returned by the solver.

3. Set
$$s[k] := 1_p - \left[\mathbb{I}(\|\tilde{\mathbf{e}}_1[k]\|_2 > 0) \quad \ldots \quad \mathbb{I}(\|\tilde{\mathbf{e}}_p[k]\|_2 > 0) \right]^\top$$
where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in $\tilde{\mathbf{E}}_{2,0}[k]$. The ith component of $s[k]$ is 1 if we haven't detected an attack on sensor i over the last N time steps. Otherwise the ith component equals zero.

4. Take the average of the un-attacked sensor readings as the information available for feedback
$$y_{\text{FB}}[k] := \frac{s^\top[k]}{s^\top[k]1_p} \mathbf{y}[k]. \tag{5.12}$$

---

[3]Based on the parameters of the single AGC system shown in Table 5.2 $D_i^{\tilde{e}^{l_0}} = 1.2077$, and $D_i^{\tilde{e}^{l_1}} = 0.0049$.

5. Update the discretized PI control signal

$$x_c[k+1] = x_c[k] + Ty_{\text{FB}}[k]$$
$$u[k] = K_P y_{\text{FB}}[k] + K_I x_c[k]$$

(5.13)

and return to Step 2 at the next sample instant.

5.6   Extension to Multi-area AGC Systems

Conventional multi-area AGC is based upon tie-line bias control where each area tends to reduce the area control error (ACE) to zero [86].

Consider a power network with n areas represented as a set of vertices $\mathbb{N}_n = \{1, \ldots, n\}$ and overhead or underground lines (tie- lines) represented by a set of edges $\mathsf{E} \subseteq \mathsf{V} \times \mathsf{V}$. The neighbors of area i are defined as $\mathcal{N}_i := \{j \in \mathbb{N}_n : (i,j) \in \mathsf{E}\}$. The neighbours of area i are simply the areas connected to it via tie-lines. If $(i,j) \in \mathsf{E}$, then ΔP_{ij} represents a deviation from the scheduled exchanges between areas i and j. The variable $\Delta\omega_i$ represents the deviation from the nominal frequency value for area i. With this notation, the area control error for area i consists of a linear combination of frequency and its neighboring tie-line error $\text{ACE}_i := \beta_i \Delta\omega_i + \sum_{j \in \mathcal{N}_i} \Delta P_{ij}$. The area bias β_i determines the amount of interaction during a disturbance in the neighboring areas. To model the interconnection with its neighbors, we modify the single AGC continuous-time model (5.1) as follows. For simplicity assume that all the AGCs have the same physical constants. Let $r_i := |\mathcal{N}_i|$ and $n_i := 3 + r_i$. The three comes from the original state variables in (5.1) and the $|\mathcal{N}_i|$ extra states come from the interconnections. The model of AGC i in the multi-area AGC setup is then given by

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + E_i d_i(t)$$

(5.14)

with $A_i \in \mathbb{R}^{n_i \times n_i}$, $B_i \in \mathbb{R}^{n_i}$, $E_i \in \mathbb{R}^{n_i \times (1+r_i)}$. Specifically, letting $N_i := e_3 1_{r_i}^\top \in \mathbb{R}^{3 \times r_i}$, where $e_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\top$ we have that

$$A_i = \begin{bmatrix} A & \frac{1}{2H} N_i \\ P_s N_i^\top & 0_{r_i \times r_i} \end{bmatrix}, B_i = \begin{bmatrix} B \\ 0_{r_i \times 1} \end{bmatrix}, E_i = \begin{bmatrix} E & 0_{3 \times r_i} \\ 0_{1 \times r_i} & -P_s I_{r_i} \end{bmatrix}$$

(5.15)

where A, B, E are the same matrices as in the single AGC system (5.1). The constant P_s is the synchronizing power coefficient, and equals the slope of the power angle curve at the initial operating angle. The first three state variables are the same as before while the last $r_i = |\mathcal{N}_i|$ state variables are ΔP_{ij}, i.e., the deviation from the scheduled exchanges between areas i and j where $j \in \mathcal{N}_i$. The control signal $u_i(t)$ is the same as in the single AGC model and the disturbance vector $d_i(t) = (\Delta P_L(t), \Delta\omega_{j_1}(t), \ldots, \Delta\omega_{j_{r_i}}(t))$ is the load change on AGC i followed by the frequency change of each of the AGC's neighbors.

In the conventional setup, AGC i has access to its own frequency measurement $\Delta\omega_i$ as well as ΔP_{ij} for each $j \in \mathcal{N}_i$. Therefore the information available for feedback is given by

$$y_i(t) = C_i x_i(t) + w_i(t) \tag{5.16}$$

where

$$C_i = \begin{bmatrix} C & 0_{1\times r_i} \\ 0_{r_i\times 3} & I_{r_i} \end{bmatrix} \in \mathbb{R}^{(1+r_i)\times n}. \tag{5.17}$$

Once again the measurement noise $w_i(t)$ is assumed to be bounded in a known compact set. The conventional control scheme [86] is to use the PI controller $u_i(t) = K_{P,i}y_{i,1}(t) + K_{I,i}\int_0^t \begin{bmatrix} \beta_i & 1 & \cdots & 1 \end{bmatrix} y_i(\tau)\mathrm{d}\tau$ where $y_{i,1}$ is the first component of y_i, i.e., AGC i's own frequency measurement. The scalers $K_{P,i}$, $K_{I,i}$ are, respectively, the proportional and integral gains of AGC i. Satisfactory performance is achieved by setting $\beta_i = K_{P,i} + D_i$ [86]

5.6.1 Proposed Attack-resilient State Estimation for the Multi Areas AGC System

Descretize the continuous time model (5.14), (5.16) of AGC i in the same manner as (5.6). The attack vector appears as an additive disturbance to the measured variables so that the AGC i model with redundant measurements and attacks is given by

$$\begin{aligned} x_i[k+1] &= A_{i_d}x_i[k] + B_{i_d}u_i[k] + E_{i_d}d_i[k] \\ \mathbf{y}_i[k] &= \mathbf{C_i}x_i[k] + w_i[k] + e_i[k] \end{aligned} \tag{5.18}$$

where $\mathbf{C_i}^\top := \begin{bmatrix} 1_{p_1} \otimes C_{i,1} & 1_{p_2} \otimes C_{i,2} & \cdots & 1_{p_{r_i+1}} \otimes C_{i,r_i+1} \end{bmatrix}$ The first p_1 rows in $\mathbf{C_i}$ represent the redundant frequency measurements, and $1_{p_{j+1}} \otimes C_{i,j+1}$, where $j \in \{1, \cdots, r_i\}$, represent

jth tie-line redundant measurements. Once again the sampling is assumed to be non-pathological. We make the mild assumption that all interconnected AGCs are sampled at the same rate $T > 0$ and that all sample and hold operators in the network are synchronised. The discretized PI controller for AGC i in state-space form is

$$x_{c,i}[k+1] = x_{c,i}[k] + Ty_i[k]$$
$$u_i[k] = K_{P,i}y_{i,1}[k] + K_{I,i}\begin{bmatrix} \beta_i & 1 & \cdots & 1 \end{bmatrix} x_{c,i}[k]. \tag{5.19}$$

The first issue considered in this section, under the assumption that we have sufficient redundancy in the sensors, is to identify which sensors have been corrupted by an attack and then use the non-attacked measurements as inputs to the control law (5.19). To understand how many attacks can tolerated, we need to understand the s-sparse observability of (5.14), (5.16).

Proposition 5.6.1. *The pair $(P_\mathcal{K}C_i, A_i)$, where A_i is given in (5.15), C_i is given by (5.17) and $\mathcal{K} \subseteq \mathbb{N}_{r_i+1}$, is observable if, and only if*

(i) $\mathcal{K} \neq \varnothing$, and

(ii) $|\mathcal{K} \cap \{2, \ldots, r_i + 1\}| \geq r_i - 1$.

The conditions of this proposition can be intuitively interpreted as saying that AGC i is observable so long as it's only missing information from *at most* one of its neighbours.

Proof. We start by understanding the structure of the unobservable subspace of the pair (C_i, A_i). To simplify notation let c_j denote the jth row of C_i. Then $\operatorname{Ker} c_1 \subseteq \operatorname{Ker} c_j A_i$ for $j \geq 2$. This can be shown by noting that $c_j A_i = c_1$ It follows that $\operatorname{Ker}(c_1 A_i^{k-1}) \subseteq \operatorname{Ker}(c_j A_i^k)$ for $k \geq 1$, $2 \leq j \leq r_i + 1$. Therefore the unobservable subspace of the pair (C_i, A_i) can be written

$$\mathcal{N}_i = \bigcap_{k=1}^{n_i} \operatorname{Ker}(C_i A_i^k) = \bigcap_{k=1}^{n_i} \left(\bigcap_{j=1}^{r_i+1} \operatorname{Ker}(c_j A_i^k) \right)$$
$$= \operatorname{Ker} C_i \cap \operatorname{Ker}(c_1 A_i) \cap \cdots \cap \operatorname{Ker}(c_1 A_i^{n_i-1})$$

where we've used $\operatorname{Ker}(c_1 A_i^{k-1}) \subseteq \operatorname{Ker}(c_j A_i^k)$ for $k \geq 1$ and $2 \leq j \leq r_i + 1$ to obtain the last equality. We can re-write the last line in the form $\mathcal{N}_i = \left(\bigcap_{j=2}^{r_i+1} \operatorname{Ker} c_j \right) \cap \mathcal{N}_{(c_1, A_i)}$, where

$\mathcal{N}_{(c_1,A_i)}$ is the unobservable subspace of (c_1, A_i). We claim that the dimension of $\mathcal{N}_{(c_1,A_i)}$ equals $n_i - 4 = r_i - 1$. To show this, compute the first 5 rows of the observability matrix of (c_1, A_i)

$$\begin{bmatrix} c & 0 \\ cA & \gamma_1 1_{r_i}^\top \\ cA^2 + \gamma_2 c & \gamma_3 1_{r_i}^\top \\ cA^3 + \gamma_2 cA + \gamma_4 c & \gamma_5 1_{r_i}^\top \\ cA^4 + \gamma_2 cA^2 + \gamma_6 cA + \gamma_7 c & \gamma_8 1_{r_i}^\top \end{bmatrix}$$

The various constants γ_i in this matrix are $\gamma_1 = 1/(2H)$, $\gamma_2 = r_i P_s \gamma_1$, $\gamma_3 = \gamma_1 cAc^\top$, $\gamma_4 = r_i \gamma_3$, $\gamma_5 = \gamma_1 c \left(A^2 + \gamma_2 I\right) c^\top$, $\gamma_6 = r_i P_s \gamma_3$, $\gamma_7 = r_i P_s \gamma_5$, $\gamma_8 = \gamma_1 c \left(A^3 + \gamma_2 A \gamma_2 I\right) c^\top$.

$$\begin{aligned} \gamma_1 &= 1/(2H), & \gamma_2 &= r_i P_s \gamma_1 \\ \gamma_3 &= \gamma_1 cAc^\top, & \gamma_4 &= r_i \gamma_3 \\ \gamma_5 &= \gamma_1 c \left(A^2 + \gamma_2 I\right) c^\top, & \gamma_6 &= r_i P_s \gamma_3 \\ \gamma_7 &= r_i P_s \gamma_5, & \gamma_8 &= \gamma_1 c \left(A^3 + \gamma_2 A \gamma_2 I\right) c^\top. \end{aligned}$$

Performing elementary row reduction on this matrix, then, using the Cayley-Hamilton theorem before doing further row reduction, we obtain

$$\begin{bmatrix} c & 0 \\ cA & \gamma_1 1_{r_i}^\top \\ cA^2 & cAc^\top/(2H) 1_{r_i}^\top \\ cA^3 & cA^2 c^\top/(2H) 1_{r_i}^\top \\ cA^4 & cA^3 c^\top/(2H) 1_{r_i}^\top \end{bmatrix}$$

Therefore, using the Cayley-Hamilton theorem and doing further row reduction, we obtain

$$\begin{bmatrix} c & 0 \\ cA & \gamma_1 1_{r_i}^\top \\ cA^2 & cAc^\top/(2H) 1_{r_i}^\top \\ 0 & \gamma_9 1_{r_i}^\top \\ 0 & \gamma_{10} 1_{r_i}^\top. \end{bmatrix}$$

The coefficient γ_9 is guaranteed to be positive because $cA^2 c^\top/(2H)$ is positive and, since A is Hurwitz, the coefficients of its characteristic polynomial are all positive. Since (c, A)

is observable (cf. Section 5.4), we conclude that this matrix has rank 4 and that the dimension of $\mathcal{N}_{(c_1, A_i)}$ equals $n_i - 4 = r_i - 1$ as claimed. We can therefore conclude that $\mathcal{N}_i = \{0\}$, i.e., (C_i, A_i) is observable, if and only if there are $r_i - 1$ tie-line measurements. This is precisely condition (ii) of the proposition statement while condition (i) ensures that the frequency sensor is measured in the case when the AGC only has one neighbor.   □

Corollary 5.6.2. *For $r_i = 1$, the maximum number of sensors which can be attacked when there is no noise and no disturbances while retaining the ability to estimate the state of the AGC i equals $\lfloor (p_1 + p_2 - 1)/2 \rfloor$.*

Proof. Based on Proposition 5.6.1, for any sensor $j \in \mathbb{N}_{p_1 + p_2}$ observe that the pair $(P_j C_i, A_{i_d})$ is observable. Therefore the pair $(P_{\mathcal{K}^c} \mathbf{C_i}, A)$ is observable as long as $P_{\mathcal{K}^c}$ isn't equal to the zero map, i.e., $\mathcal{K}^c \neq \varnothing$. This in turn implies that $|\mathcal{K}| \leq p_1 + p_2 - 1$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result.   □

Before, introducing Corollary 5.6.3, for sake of clarity and without loss of generality, assume that the tie-line measurements p_{j+1} are ordered in increasing manner, where $p_2 \leq p_3 \leq \cdots \leq p_{r_i + 1}$.

Corollary 5.6.3. *For $r_i \geq 2$, the maximum number of sensors which can be attacked when there is no noise and no disturbances while retaining the ability to estimate the state of the AGC i equals $\lfloor (p_2 + p_3 - 1)/2 \rfloor$.*

Proof. Based on Proposition 5.6.1, observe that the pair (C_i, A_i) is observable as long as the number of available tie-line measurements is at least equal to $r_i - 1$. Therefore the pair $(P_{\mathcal{K}^c} \mathbf{C_i}, A)$ is observable as long as the cardinality of $\mathcal{K}$ is less than $p_2 + p_3$. Setting $2q_{\max} = |\mathcal{K}|$ and solving for $q_{\max}$ gives the desired result.   □

We now present the attack mitigation strategy for AGC in multi-area power network. To understand the strategy, note that each AGC in the network solves it's own version of Problem 1, *mutatis mutandis*, using its own model information. At each time step k, the proposed attack mitigation strategy for AGC i is as follows.
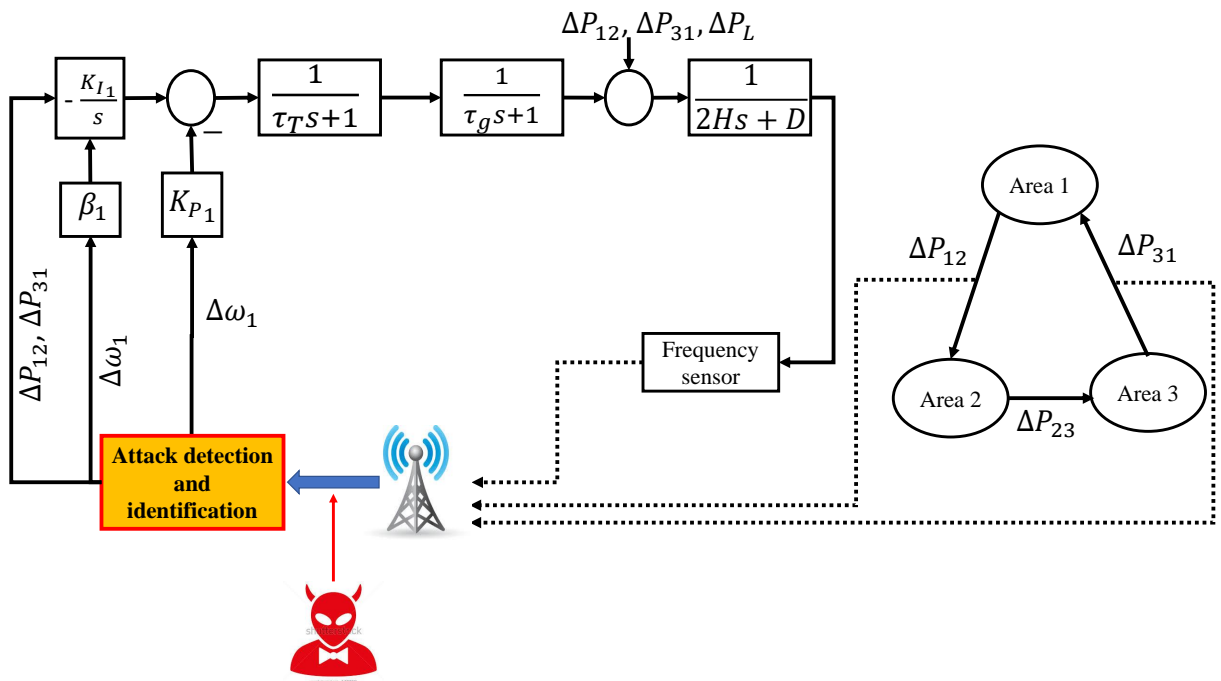
Figure 5.1: The scheme of the multi-area AGC system with the proposed algorithm.

1. Fix $N \geq 3 + r_i$ to be the number of steps over which we aim to detect attacks. The lower bound of $3 + r_i$ comes from the dimension of the AGC's state-space.

2. If $k \geq N - 1$, solve the mixed integer linear program Problem 1. Let $\tilde{\mathbf{E}}_{2,0}[k]$ denote the value of the decision variable returned by the solver.

3. For the redundant measurements, we recursively define $\rho_0 := 0$, $\rho_j := \rho_{j-1} + p_j$ for $j \geq 1$ and then set

$$
s_j[k] \quad := 1_{p_j}^\top - \tag{5.20}
$$
$$
\left[\mathbb{I}(\|\tilde{\mathbf{e}}_{\rho_{j-1}+1}[k]\|_2 > 0) \quad \cdots \quad \mathbb{I}(\|\tilde{\mathbf{e}}_{\rho_j}[k]\|_2 > 0) \right]^\top.
$$

where $\tilde{\mathbf{e}}_j[k]$ is the jth block vector in $\tilde{\mathbf{E}}_{2,0}[k]$. The ith component of $s_j[k]$ is 1 if we haven't detected an attack on sensor i over the last N time steps, where $j \in \{1, \cdots, r_i + 1\}$. Otherwise the ith component equals zero.

4. Take the average of the un-attacked sensor readings as the information available for feedback
$$
y_{\mathrm{FB}_i}[k] := \mathbf{M} y_i[k]. \tag{5.21}
$$
where, $\mathbf{M} := \mathrm{diag}\left(\frac{s_1^\top[k]}{s_1^\top[k]1_{p_1}}, \cdots, \frac{s_{r_i+1}^\top[k]}{s_{r_i+1}^\top[k]1_{p_{r_i+1}}} \right)$.

5. Update the discretized PI control signal

$$
x_{c,i}[k+1] = x_{c,i}[k] + T y_{\mathrm{FB}_i}[k] \tag{5.22}
$$
$$
u_i[k] = K_{P,i} y_{\mathrm{FB}_{i,1}}[k] + \tag{5.23}
$$
$$
K_{I,i} \begin{bmatrix} \beta_i & 1 & \cdots & 1 \end{bmatrix} x_{c,i}[k]
$$

where, $y_{\mathrm{FB}_{i,1}}[k]$ is the first element in $y_{\mathrm{FB}_i}[k]$ and return to Step 2 at the next sample instant.

Of course, for this strategy to be feasible for real-time control of the AGC i, one must be able to solve Problem 1 within the scheduling constraints of the embedded controller. The real-time constraint becomes too restrictive as the number of neighbours increases.

However, we show that for the use of a fully connected, 3 AGC system, Problem 1 can be solved in real-time. Most existing interconnected systems [48, 112] consist of at most three tie lines. This is due to economic and technical constraints discussed in [1, 31]. Therefore, from a practical point of view, the proposed algorithm can be applied to most multi-AGC systems in existence.

5.7 Simulation Results

We present and illustrate the simulation results of the proposed algorithm compared with legacy AGC in the single and three-area system. The parameters used in simulation are given, respectively, in Tables 5.2 and 5.3. The parameters are in per unit on a common $1000\ MVA$ base. We implement the proposed algorithm in MATLAB and time the execution of each iteration on a 3.6 GHz i7-7700 CPU. We use CVX solver (MOSEK) for solving Problem 1. We begin with the AGC in the single area with sampling period $T = 0.1\ s$. For simulation purposes, we consider the following attack signal on sensor i

$$e_i(k) = \begin{cases} \mathbf{a}(k - k_r^1 + 1), & \text{for}, k_r^1 \leq k \leq k_r^2 \quad \text{(ramp)} \\ \mathbf{a}, & \text{for}, k_c^1 \leq k \leq k_c^2 \quad \text{(constant)} \\ \mathbf{a}f(k), & \text{for}, k_p^1 \leq k \leq k_p^2 \quad \text{(pulse)} \\ \mathbf{a} \sim \mathcal{N}(\mu, \sigma), & \text{for}, k_k^1 \leq k \leq k_k^2 \quad \text{(random)} \\ 0 & \text{otherwise} \end{cases} \tag{5.24}$$

where, $f(k)$, is a pulse wave with a 50 % duty cycle and period $6T$. The parameters μ, σ are, respectively, the mean and the standard deviation of the Gaussian distribution. The parameter $\mathbf{a}$ controls the "size" of the attack signal. Informally, when $\mathbf{a}$ is small we call the attacks "stealthy". The load change ΔP_L is generated by scaling the steady state load by zero-mean Gaussian random variable of standard deviation 0.05 per unit (p.u.). We assume that the number of sensors available is equal to 3 ($p = 3$), so based on Proposition 5.4.3, $q_{\max}$ is equal to one. Fig. 5.2 shows the frequency deviation of the grid in Hz against time (kT) in seconds. Fig. 5.2 shows the frequency deviation of the grid with and without the proposed algorithm, and compared with the original frequency

deviation. Several kinds of attacks are applied during the first and the second shaded area to simulate the unpredictable behaviour of the attacker. For the stealthy attacks during the first shaded area, ramp, pulse, constant, and random signal have been initiated at $k_r^1 = 40$, $k_p^1 = 150$, $k_c^1 = 200$, and $k_k^1 = 250$ respectively. The ramp attack with $\mathbf{a} = -0.005$ lasts for 8.6 seconds ($k_r^2 = 125$). The pulse attack with $\mathbf{a} = -0.2$ lasts for 3.6 seconds ($k_p^2 = 185$). The constant with $\mathbf{a} = -0.2$, and random with $\mu = -0.2$, $\sigma = 0.1$, attack lasts for 2.6 seconds ($k_c^2 = 235$, and $k_k^2 = 285$). When the attacker is applying stealthy attacks (first shaded area), the frequency goes slightly above the true grid frequency value. This means that the generation is unnecessarily above the demand. Consequently, the grid is operated uneconomically, causing some profit loss to the owner. In the second shaded area, the ramp attack with $\mathbf{a} = 0.05$ is launched again at 30 seconds ($k_r^1 = 300$). The aim of this attack is to cause high deviation from the nominal frequency and trigger the remedial actions. The thresholds ϵ_L and ϵ_U shown in Fig. 5.2 are set to those for triggering remedial actions. We use $\epsilon_L = -0.5$ Hz and $\epsilon_U = 0.5$ Hz [104]. As shown in Fig. 5.2, the attacker manipulates the AGC system to make the grid frequency leave the safety region at time 37.5 s, while the AGC system equipped with proposed algorithm follows the not attacked frequency deviation. The proposed algorithm is able to detect and identify the sensors under attack in 0.07 s (i.e., Problem 1 is solved using MOSEK in 0.07 s). The proposed algorithm is able to detect the attack of size 0.02 Hz (i.e., $\mathbf{a} = 0.02$) once the attack is launched. Smaller sizes can be detected after 0.3 to 0.5 second. The proposed algorithm is able to detect and identify the sensors under attack even in the case of sudden load change or a high load fluctuation i.e., the proposed algorithm is able to detect and identify the sensors under attack when the load change ΔP_L is generated by scaling the highly fluctuated load by zero-mean Gaussian random variable with a standard deviation of 0.2 p.u.

Next, we test the AGC equipped with the proposed algorithm in a three-area configuration under various types of coordinated attacks with sampling period $T = 0.1 s$. For each AGC i, we assume that the number of frequency sensors available is equal to 3 ($p_1 = 3$), the number of first tie-line sensors available is equal to 3 ($p_2 = 3$), and the number of second tie-line sensors available is equal to 4 ($p_3 = 4$). Based on Corollary 5.6.3, the maximum number of sensor that can be attacked is equal to 3 ($q_{\max} = 3$), however, simulations
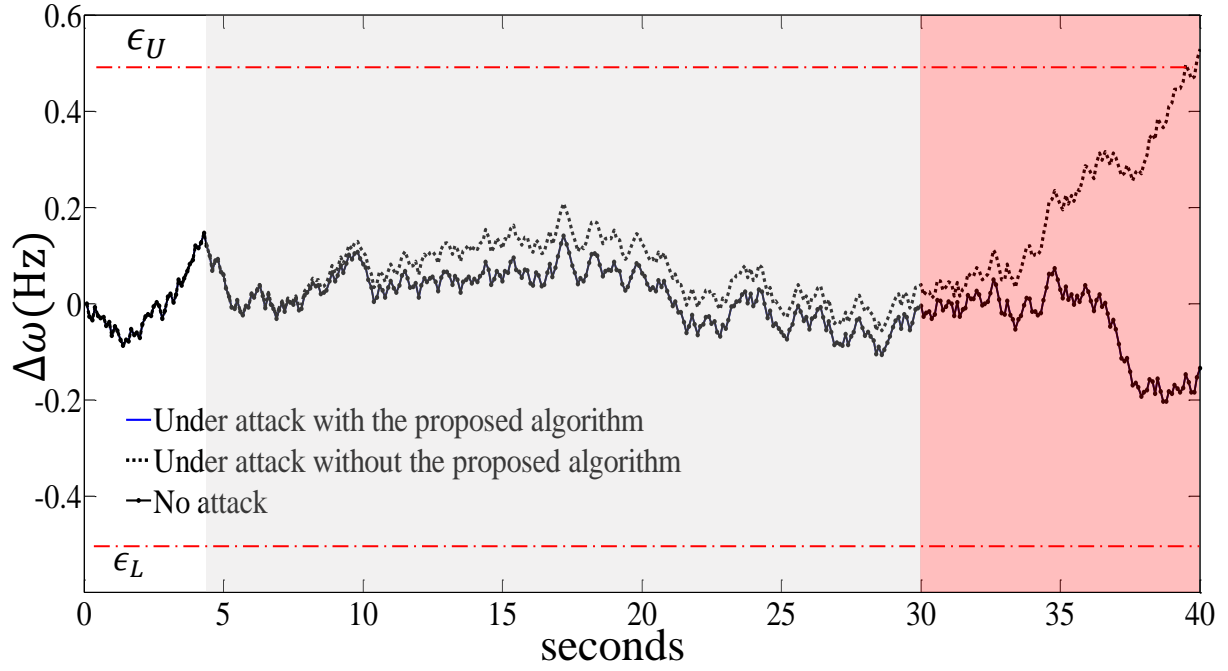
Figure 5.2: The grid frequency deviation where the system is under stealthy attack (first shaded region) and under severe attack (second shaded region)

Table 5.3: System parameters [101]

Area 'i'	D	R	β	K_i	$H(s)$	$\tau_g(s)$	$\tau_T(s)$
Area 1	0.8	0.05	20.8	0.5	5	0.2	0.5
Area 2	0.9	0.0625	16.9	0.5	4	0.3	0.6
Area 3	0.9	0.0625	16.9	0.5	4	0.3	0.6

suggest that ($q_{\max} = 2$). This is due to the presence of noise and load disturbance. The proposed algorithm is able to detect and identify the tie-line sensors under attack even in the case of sudden load change or a high load fluctuation, while frequency sensors under attack can only be detected during steady-state load conditions.

In our simulation, the attacker misleads AGC 1 by decreasing the flow measurement from Area 1 to Area 2 by 0.03 p.u., and Area 1 to Area 3 by 0.03 p.u.. At the same time, the attacker keeps normal measurements according to the scheduled values to AGC in Areas 2 and 3 in order to prevent corrective action. In this scenario Area 1 increase the generated power in the tie-line causing the frequency goes above the upper safety limit for all areas causing a remedial action to be triggered.

For the stealthy attacks during the first shaded area, pulse, constant, and random signal have been initiated at $k_p^1 = 40$, $k_c^1 = 80$, and $k_k^1 = 150$ respectively. The pulse attack with $\mathbf{a} = -0.03$ lasts for 3.6 seconds ($k_p^2 = 75$). The constant with $\mathbf{a} = -0.03$, last for 6.6 seconds ($k_c^2 = 145$). The random attacks with $\mu = -0.03$, and $\sigma = 0.01$ last for 2.6 seconds ($k_k^2 = 175$). In the second shaded area, the ramp attack with $\mathbf{a} = -0.01$ is launched at 20 s. As shown in Fig. 5.3, the attack will cause a remedial action as the frequency goes beyond the safe region at 24 s, while the AGC system equipped with proposed algorithm follows the not attacked frequency deviation. The proposed algorithm is able to detect and identify the sensors under attack in 0.09 s (i.e., Problem 1 is solved in 0.09 s). The proposed algorithm is able to detect the attack on tie-line sensors of size 0.01 p.u. (i.e., $\mathbf{a} = 0.01$) once the attack launched , while for frequency sensors the size is $0.1\,\mathrm{Hz}$ (i.e., $\mathbf{a} = 0.1$). Smaller sizes can be detected after 0.3 to 0.5 second.

The proposed algorithm was also tested on a system with twenty states (i.e., eighteen-area AGC system), and is able to detect and identify the sensors under attack in $0.51s$ (i.e., Problem 1 is solved using MOSEK in $0.51\,s$). However, since most existing multi-area AGC systems are based on three or two areas, we prefer to demonstrate the above explained example. We believe that the supercomputer of the power system operator (PSO) can solve Problem 1 in much faster time.
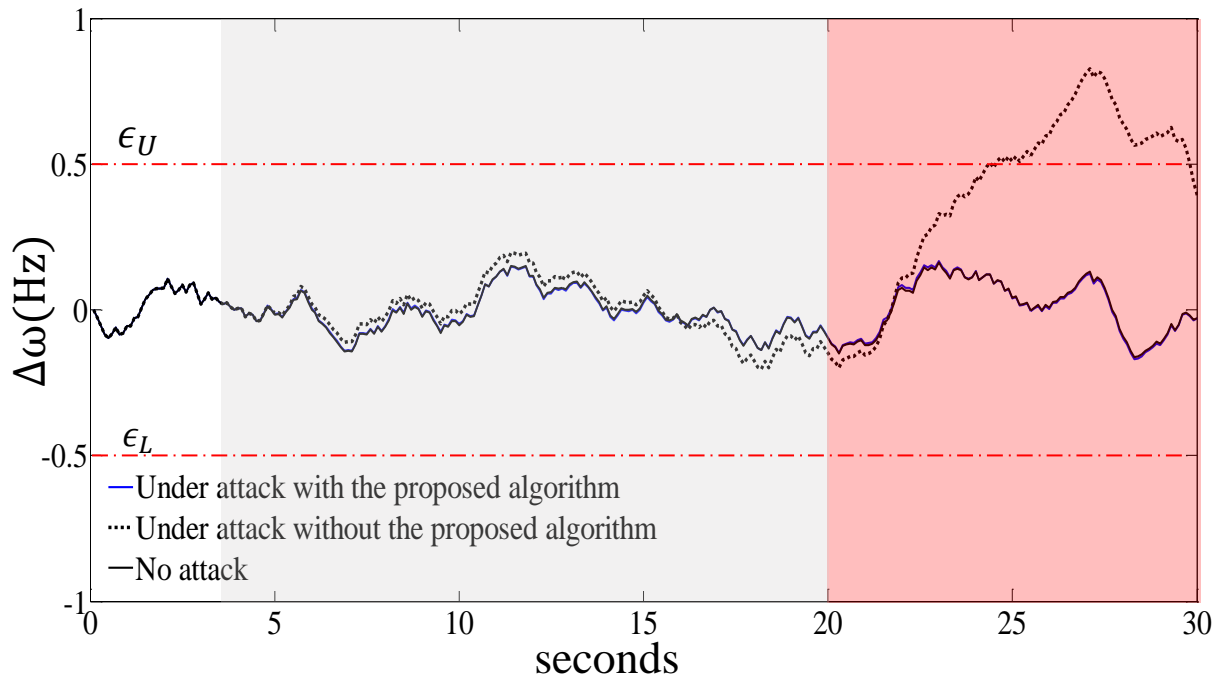
Figure 5.3: Area 1 frequency deviation where the system is under stealthy attack (first shaded region) and under severe attack (second shaded region)

5.8   Conclusion

This chapter developed an efficient algorithm for detecting, identifying and mitigating cyber-physical attacks for single and multi-area AGC systems. The proposed algorithm develop a MILP-based state estimation procedure considering uncertainty. We provide s-sparse observability analysis for AGC in single area and multi-area power networks. We derived a key formula to compute the number of sensors needed versus the number of attacks that can be tolerated. We propose a mitigation procedure based on simple switching algorithm. Our analysis and algorithms are validated by simulations for AGC in single and three-area power network. The proposed algorithm is capable of providing an accurate detection of attacks and identification of the sensors under attack even in the case of sudden load change or high load fluctuation.

Chapter 6

Synopsis and Future Research

6.1 Synopsis

The classic Kalman filter provides the optimal estimate of the state in the minimum mean-squared error sense. This optimal estimate is only obtained for linear Gaussian dynamic systems. However, real-world systems in general are nonlinear, non-Gaussian or both. In this context, finding a closed-form solution for the state estimate is very difficult, and therefore some approximations must be made. An intensive research has been carried out to find an approximate, yet more accurate nonlinear filtering solution since the original formulation of the Kalman filter in 1960. In this thesis, we introduced one nonlinear filter, which is the SCKF, and two nonlinear non-Gaussian filters, which are CPF and SCPF, to the area of the distribution system state estimation. First, the three proposed filters were applied to estimate the distribution system states utilizing on-line measurements. The sampling rate T was assumed to be equal to $5\,\mathrm{sec}$. A hypothesis load profile was generated, simulating the effect of the DGs to be used as pseudo-measurements. Following, a comprehensive comparison between the proposed algorithms and the latest methods in the literature was provided. Second, we proposed a pseudo measurement generator based wavelet RNN to allow the operator to forecast the status of the distribution system in different time horizons. The proposed PG was integrated with one of the proposed SE to estimate the states of the distribution system in a much longer time horizon (1 hour in 5 min

98

interval). The proposed SE when used with the proposed PG is based on $5\,\mathrm{min}$ sampling rate, which facilitates the monitoring and planing activities to be carried out. Finally, we proposed a cyber attack detection algorithm based state estimator. The proposed algorithm was applied to the AGC system in a single area and a multi area. The proposed algorithm can also identify the sensors under attack. We proposed a simple switching algorithm that enables the controller to make the proper decision. We provide an S sparse observability analysis to the AGC in a single area and a multi area.

In summery the contributions of this thesis can be stated as follows,

- Introducing new SEs that can handle the challenges of future distribution grids with higher accuracy.

- Providing a comparison study between the most common state estimation algorithms used to estimate the distribution system states.

- Introducing a wavelet recurrent neural network (WRNN) to model the load at each bus in the distribution system

- Developing a new attack resilient scheme for single and multi-area AGC systems.

6.2   Directions for Future Work

There is much room for further work on problems related to the three problem solved in this thesis. The proposed SE for operational purposes may provide a basis for contributions. A number of interesting problems to be treated in the years ahead include:

1. Design a VSTLF based real data that provides a load change in $5\,\mathrm{sec}$ or less horizon.

2. Build a operational situation assessment algorithm utilizing the data obtained by the proposed SE.

The proposed SE for planing purposes may provide a basis for contributions. A number of interesting problems to be treated in the years ahead include:

1. Design a VSTLF with more sophisticated NN.

2. Build a planing situation assessment algorithm utilizing the data obtained by the proposed SE based VSTLF.

The proposed algorithm for mitigation the Cyber-Physical Attacks in Multi-Area Automatic Generation may provide a basis for contributions. A number of interesting problems to be treated in the years ahead include:

1. Apply the proposed algorithm to more larger AGC system.

References

[1] P.-O. Pineau and D.J. Dupuis and T. Cenesizoglu, assessing the value of power interconnections under climate and natural gas price risks. *Elsevier, Energy*, 82:128 – 137, 2015.

[2] SM Shafiul Alam, Balasubramaniam Natarajan, and Anil Pahwa. Distribution grid state estimation from compressed measurements. *IEEE Transactions on Smart Grid*, 5(4):1631–1642, 2014.

[3] Safoan MO Alhalali and Ramadan A Elshatshat. State estimator for electrical distribution systems based on a particle filter. In *Power & Energy Society General Meeting, 2015 IEEE*, pages 1–5. IEEE, 2015.

[4] Ienkaran Arasaratnam. *Cubature Kalman filtering theory & applications*. PhD thesis, 2009.

[5] Ienkaran Arasaratnam and Simon Haykin. Cubature kalman filters. *IEEE Transactions on automatic control*, 54(6):1254–1269, 2009.

[6] A. Ashok, Pengyuan Wang, M. Brown, and M. Govindarasu. Experimental evaluation of cyber attacks on automatic generation control using a cps security testbed. In *2015 IEEE Power Energy Society General Meeting*, pages 1–5, July 2015.

[7] Amir Asif, Arash Mohammadi, and Shivam Saxena. Reduced order distributed particle filter for electric power grids. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 7609–7613. IEEE, 2014.

[8] H. M. Beides and G. T. Heydt. Dynamic state estimation of power system harmonics using kalman filter methodology. *IEEE Transactions on Power Delivery*, 6(4):1663–1670, Oct 1991.

[9] S. Bhowmik, K. Tomsovic, and A. Bose. Communication models for third party load frequency control. *IEEE Transactions on Power Systems*, 19(1):543–548, Feb 2004.

[10] EA Blood, BH Krogh, and MD Ilic. Electric power system static state estimation through kalman filtering and load forecasting. In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–6. IEEE, 2008.

[11] E Caro and AJ Conejo. State estimation via mathematical programming: a comparison of different estimation algorithms. *IET generation, transmission & distribution*, 6(6):545–553, 2012.

[12] Wiktor Charytoniuk and M-S Chen. Very short-term load forecasting using artificial neural networks. *IEEE transactions on Power Systems*, 15(1):263–268, 2000.

[13] Phani Chavali and Arye Nehorai. Distributed power system state estimation using factor graphs. *IEEE Transactions on Signal Processing*, 63(11):2864–2876, 2015.

[14] Liang Chen, Tianshu Bi, Ancheng Xue, and Qixun Yang. A novel approach for dynamic state estimation with synchrophasor measurements. In *Security in Critical Infrastructures Today, Proceedings of International ETG-Congress 2013; Symposium 1:*, pages 1–6. VDE, 2013.

[15] Antonio J Conejo, Sebastin de la Torre, and Miguel Canas. An optimization approach to multiarea state estimation. *IEEE Transactions on Power Systems*, 22(1):213–221, 2007.

[16] AM Leite Da Silva, MB Do Coutto Filho, and JF De Queiroz. State forecasting in electric power systems. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 130, pages 237–244. IET, 1983.

[17] LCM De Andrade and IN Da Silva. Using intelligent system approach for very short-term load forecasting purposes. In *Energy Conference and Exhibition (EnergyCon), 2010 IEEE International*, pages 694–699. IEEE, 2010.

[18] Luciano Carli Moreira de Andrade and Ivan Nunes da Silva. Very short-term load forecasting using a hybrid neuro-fuzzy approach. In *Neural Networks (SBRN), 2010 Eleventh Brazilian Symposium on*, pages 115–120. IEEE, 2010.

[19] Atif S Debs and Robert E Larson. A dynamic estimator for tracking the state of a power system. *IEEE Transactions on Power Apparatus and Systems*, (7):1670–1678, 1970.

[20] Milton Brown Do Coutto Filho and Julio Cesar Stacchini de Souza. Forecasting-aided state estimationpart i: Panorama. *IEEE Transactions on Power Systems*, 24(4):1667–1677, 2009.

[21] Milton Brown Do Coutto Filho, Julio Cesar Stacchini de Souza, and Ronaldo SÉrgio Freund. Forecasting-aided state estimationpart ii: Implementation. *IEEE Transactions on Power Systems*, 24(4):1678–1685, 2009.

[22] G Durgaprasad and SS Thakur. Robust dynamic state estimation of power systems based on m-estimation and realistic modeling of system dynamics. *IEEE Transactions on Power Systems*, 13(4):1331–1336, 1998.

[23] I Dzafic, S Henselmeyer, and HT Neisius. High performance state estimation for smart grid distribution network operation. In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1–6. IEEE, 2011.

[24] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Transactions on Signal Processing*, 58(6):3042–3054, June 2010.

[25] Kianoush Emami, Tyrone Fernando, Herbert Ho-Ching Iu, Hieu Trinh, and Kit Po Wong. Particle filter approach to dynamic state estimation of generators in power systems. *IEEE Trans. Power Syst.*, 30(5):2665–2675, 2015.

[26] Kianoush Emami, Tyrone Fernando, and Brett Nener. Power system dynamic state estimation using particle filter. In *Industrial Electronics Society, IECON 2014-40th Annual Conference of the IEEE*, pages 248–253. IEEE, 2014.

[27] P. Mohajerin Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Cyber attack in a two-area power system: Impact identification using reachability. In *Proceedings of the 2010 American Control Conference*, pages 962–967, June 2010.

[28] P Favre-Perrod, R Critchley, E Catz, and M Bazargan. New participants in smart-grids and associated challenges in the transition towards the grid of the future. In *PowerTech, 2009 IEEE Bucharest*, pages 1–5. IEEE, 2009.

[29] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, June 2014.

[30] Wang Feng, Xie Kai, Yu Erkeng, Li Guoqi, and Wang Manyi. A simple and effective ultrashort term load forecasting method [j]. *Power System Technology*, 3, 1996.

[31] Independent Electricity System Operator for the Deputy Minister of Energy. Ontario-quebec interconnection capability a TECHNICAL REVIEW. [Online] (2016) .Available: http://www.ieso.ca/-/media/files/ieso/document-library/power-data/supply/intertiereport-20170508.pdf?la=en.

[32] Brendan Fox, Damian Flynn, Leslie Bryans, Nick Jenkins, David Milborrow, Mark O'Malley, Richard Watson, and Olimpo Anaya-Lara. *Wind power integration: connection and system operational aspects*, volume 50. Iet, 2007.

[33] E. Ghahremani and I. Kamwa. Dynamic state estimation in power system by applying the extended kalman filter with unknown inputs to phasor measurements. *IEEE Transactions on Power Systems*, 26(4):2556–2566, Nov 2011.

[34] Esmaeil Ghahremani and Innocent Kamwa. Dynamic state estimation in power system by applying the extended kalman filter with unknown inputs to phasor measurements. *IEEE Transactions on Power Systems*, 26(4):2556–2566, 2011.

[35] PA Gnadt and JS Lawler. *Automating electric utility distribution systems*. Old Tappan, NJ (USA); Prentice Hall Inc., 1990.

[36] Antonio Gomez-Exposito and Ali Abur. *Power system state estimation: theory and implementation*. CRC press, 2004.

[37] Antonio Gomez-Exposito, Ali Abur, Antonio de la Villa Jaen, and Catalina Gomez-Quiles. A multilevel state estimation paradigm for smart grids. *Proceedings of the IEEE*, 99(6):952–976, 2011.

[38] Antonio Gómez-Expósito, Antonio de la Villa Jaén, Catalina Gómez-Quiles, Patricia Rousseaux, and Thierry Van Cutsem. A taxonomy of multi-area state estimation methods. *Electric Power Systems Research*, 81(4):1060–1069, 2011.

[39] Catalina Gomez-Quiles, Antonio de la Villa Jaen, and Antonio Gomez-Exposito. A factorized approach to wls state estimation. *IEEE Transactions on Power Systems*, 26(3):1724–1732, 2011.

[40] George Gross and Francisco D Galiana. Short-term load forecasting. *Proceedings of the IEEE*, 75(12):1558–1573, 1987.

[41] Che Guan, Peter B Luh, Laurent D Michel, and Zhiyi Chi. Hybrid kalman filters for very short-term load forecasting and prediction interval estimation. *IEEE Transactions on Power Systems*, 28(4):3806–3817, 2013.

[42] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu. Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid. *IEEE Transactions on Smart Grid*, 4(2):847–855, June 2013.

[43] Mohammad Hassanzadeh and Cansin Yaman Evrenosoglu. A regression analysis based state transition model for power system dynamic state estimation. In *North American Power Symposium (NAPS), 2011*, pages 1–5. IEEE, 2011.

[44] Mohammad Hassanzadeh and Cansın Yaman Evrenosoğlu. Power system state forecasting using regression analysis. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–6. IEEE, 2012.

[45] Barry Patrick Hayes and Milan Prodanovic. State forecasting and operational planning for distribution network energy management systems. *IEEE Transactions on Smart Grid*, 7(2):1002–1011, 2016.

[46] Simon Haykin. *Kalman filtering and neural networks*, volume 47. John Wiley & Sons, 2004.

[47] Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1):44–55, 2001.

[48] Sara Hoff. U.S. electric system is made up of interconnections and balancing authorities. [Online] (2016). Available:https://www.eia.gov/todayinenergy/detail.php?id=27152.

[49] Weiqing Jiang, Vijay Vittal, and Gerald T Heydt. A distributed state estimator utilizing synchronized phasor measurements. *IEEE Transactions on Power Systems*, 22(2):563–571, 2007.

[50] Weiqing Jiang, Vijay Vittal, and Gerald T Heydt. Diakoptic state estimation using phasor measurement units. *IEEE Transactions on Power Systems*, 23(4):1580–1589, 2008.

[51] Xie Kai, Wang Feng, Yu Erkeng, Ge Weichun, Ma Xin, and Pang Minghui. Very shortterm load forecasting by kalman filter algorithm [j]. *Proceedings of the Chinese Society of Electric Engineering*, 4:2–45, 1996.

[52] Edward W Kamen and Jonathan K Su. *Introduction to optimal estimation.* Springer Science & Business Media, 1999.

[53] S. Karnouskos. Stuxnet worm impact on industrial cyber-physical system security. In *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, pages 4490–4494, Nov 2011.

[54] M Kiani and Seid H Pourtakdoust. Adaptive square-root cubature–quadrature kalman particle filter via kld-sampling for orbit determination. *Aerospace Science and Technology*, 46:159–167, 2015.

[55] Maryam Kiani and Seid H Pourtakdoust. Adaptive square-root cubature–quadrature kalman particle filter for satellite attitude determination using vector observations. *Acta Astronautica*, 105(1):109–116, 2014.

[56] Piotr Kozierski, Marcin Lis, and Dariusz Horla. Power system state estimation using dispersed particle filter. *Journal of Automation Mobile Robotics and Intelligent Systems*, 8, 2014.

[57] Piotr Kozierski, Marcin Lis, Adam Owczarkowski, and Dariusz Horla. Dispersed filters for power system state estimation. In *Methods and Models in Automation and Robotics (MMAR), 2014 19th International Conference On*, pages 129–133. IEEE, 2014.

[58] C. Kwon, W. Liu, and I. Hwang. Security analysis for cyber-physical systems against stealthy deception attacks. In *2013 American Control Conference*, pages 3344–3349, June 2013.

[59] Robert H Lasseter. Smart distribution: Coupled microgrids. *Proceedings of the IEEE*, 99(6):1074–1082, 2011.

[60] Pei H Leong, Sanjeev Arulampalam, Tharaka A Lamahewa, and Thushara D Abhayapala. A gaussian-sum based cubature kalman filter for bearings-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 49(2):1161–1176, 2013.

[61] Junqi Liu, Andrea Benigni, Dragan Obradovic, Sandra Hirche, and Antonello Monti. State estimation and learning of unknown branch current flows using decentralized kalman filter with virtual disturbance model. In *Applied Measurements For Power Systems (AMPS), 2010 IEEE International Workshop on*, pages 31–36. IEEE, 2010.

[62] K Liu, S Subbarayan, RR Shoults, MT Manry, C Kwan, FI Lewis, and J Naccarino. Comparison of very short-term load forecasting techniques. *IEEE Transactions on power systems*, 11(2):877–882, 1996.

[63] Qian Liu, Lei Ding, Xiaohui Wang, Yadong Gao, Fei Jin, Jinyu Wu, and Xiaoliang Liu. Distribution system state estimation considering the uncertainty of dg output. In *Energy Internet and Energy System Integration (EI2), 2017 IEEE Conference on*, pages 1–6. IEEE, 2017.

[64] Hanif Livani, Saeed Jafarzadeh, Cansin Yaman Evrenosoglu, and M Sami Fadali. A unified approach for power system predictive operations using viterbi algorithm. *IEEE Transactions on Sustainable Energy*, 5(3):757–766, 2014.

[65] Dian-sheng Luo and Hong-ying He. A shape similarity criterion based curve fitting algorithm and its application in ultra-short-term load forecasting. *Power System Technology*, 31(21):81–84, 2007.

[66] M Majidi, M Etezadi-Amoli, and H Livani. Distribution system state estimation using compressive sensing. *International Journal of Electrical Power & Energy Systems*, 88:175–186, 2017.

[67] Efthymios Manitsas, Ravindra Singh, Bikash C Pal, and Goran Strbac. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems*, 27(4):1888–1896, 2012.

[68] Y. Mo, T. H. J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. Cyber-physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, Jan 2012.

[69] Y. Mo, S. Weerakkody, and B. Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1):93–109, Feb 2015.

[70] Yasser Abdel-Rady I Mohamed and Amr A Radwan. Hierarchical control system for robust microgrid operation and seamless mode transfer in active distribution systems. *IEEE Transactions on Smart Grid*, 2(2):352–362, 2011.

[71] Arash Mohammadi and Amir Asif. Distributed state estimation for large-scale non-linear systems: A reduced order particle filter implementation. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pages 249–252. IEEE, 2012.

[72] James A Momoh. Smart grid design for efficient and flexible power networks operation and control. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pages 1–8. IEEE, 2009.

[73] Alcir Monticelli. *State estimation in electric power systems: a generalized approach.* Springer Science & Business Media, 2012.

[74] C. Muscas, M. Pau, P. A. Pegoraro, S. Sulis, F. Ponci, and A. Monti. Multiarea distribution system state estimation. *IEEE Transactions on Instrumentation and Measurement*, 64(5):1140–1148, May 2015.

[75] Marcos Netto, Junbo Zhao, and Lamine Mili. A robust extended kalman filter for power system dynamic state estimation using pmu measurements. In *Power and Energy Society General Meeting (PESGM), 2016*, pages 1–5. IEEE, 2016.

[76] K Nishiya, J Hasegawa, and T Koike. Dynamic state estimation including anomaly detection and identification for power systems. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 129, pages 192–198. IET, 1982.

[77] N. Nusrat, M. Irving, and G. Taylor. Development of distributed state estimation methods to enable smart distribution management systems. In *2011 IEEE International Symposium on Industrial Electronics*, pages 1691–1696, June 2011.

[78] M. Pajic, I. Lee, and G. J. Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1):82–92, March 2017.

[79] Miroslav Pajic, James Weimer, Nicola Bezzo, Paulo Tabuada, Oleg Sokolsky, Insup Lee, and George J. Pappas. Robustness of attack-resilient state estimators. In *ACM/IEEE 5th International Conference on Cyber-Physical Systems*, pages 163–174, Berlin, Germany, 2014.

109

[80] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, Nov 2013.

[81] M. Pau, P. A. Pegoraro, and S. Sulis. Efficient branch-current-based distribution system state estimation including synchronized measurements. *IEEE Transactions on Instrumentation and Measurement*, 62(9):2419–2429, Sept 2013.

[82] Paolo Attilio Pegoraro and Sara Sulis. On the uncertainty evaluation in distribution system state estimation. In *Smart Measurements for Future Grids (SMFG), 2011 IEEE International Conference on*, pages 59–63. IEEE, 2011.

[83] Paolo Attilio Pegoraro and Sara Sulis. Robustness-oriented meter placement for distribution system state estimation in presence of network parameter uncertainty. *IEEE Transactions on Instrumentation and Measurement*, 62(5):954–962, 2013.

[84] Junjian Qi, Kai Sun, and Wei Kang. Adaptive optimal pmu placement based on empirical observability gramian. *IFAC-PapersOnLine*, 49(18):482–487, 2016.

[85] Branko Ristic, Sanjeev Arulampalam, and Neil Gordon. *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house, 2003.

[86] Hadi Saadat. *Power system analysis*. McGraw-Hill, 1999.

[87] Shivam Saxena, Amir Asif, and Hany Farag. Nonlinear, reduced order, distributed state estimation in microgrids. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 2874–2878. IEEE, 2015.

[88] F. C. Schweppe and J. Wildes. Power system static-state estimation, part i: Exact model. *IEEE Transactions on Power Apparatus and Systems*, PAS-89(1):120–125, Jan 1970.

[89] Anthony Setiawan, Irena Koprinska, and Vassilios G Agelidis. Very short-term electricity load demand forecasting using support vector regression. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 2888–2894. IEEE, 2009.

[90] Jason Sexauer, Pirooz Javanbakht, and Salman Mohagheghi. Phasor measurement units for the distribution grid: Necessity and benefits. In *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, pages 1–6. IEEE, 2013.

[91] Payman Shamsollahi, KW Cheung, Quan Chen, and Edward H Germain. A neural network based very short term load forecaster for the interim iso new england electricity market system. In *Power Industry Computer Applications, 2001. PICA 2001. Innovative Computing for Power-Electric Energy Meets the Market. 22nd IEEE Power Engineering Society International Conference on*, pages 217–222. IEEE, 2001.

[92] A Sharma, SC Srivastava, and S Chakrabarti. Testing and validation of power system dynamic state estimators using real time digital simulator (rtds). *IEEE Transactions on Power Systems*, 31(3):2338–2347, 2016.

[93] A Sharma, Suresh Chandra Srivastava, and Saikat Chakrabarti. A cubature kalman filter based power system dynamic state estimator. *IEEE Transactions on Instrumentation and Measurement*, 66(8):2036–2045, 2017.

[94] Kuang-Rong Shih and Shyh-Jier Huang. Application of a robust algorithm for dynamic state estimation of a power system. *IEEE Transactions on Power Systems*, 17(1):141–147, 2002.

[95] NR Shivakumar and Amit Jain. A review of power system dynamic state estimation techniques. In *Power System Technology and IEEE Power India Conference, 2008. POWERCON 2008. Joint International Conference on*, pages 1–6. IEEE, 2008.

[96] Y. Shoukry and P. Tabuada. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, 61(8):2079–2091, Aug 2016.

[97] R Singh, BC Pal, and RA Jabr. Choice of estimator for distribution system state estimation. *IET generation, transmission & distribution*, 3(7):666–678, 2009.

[98] Ravindra Singh, Bikash C Pal, Rabih A Jabr, and Richard B Vinter. Meter placement for distribution system state estimation: An ordinal optimization approach. *IEEE Transactions on Power Systems*, 26(4):2328–2335, 2011.

[99] AK Sinha and JK Mandal. Hierarchical dynamic state estimator using ann-based dynamic load prediction. *IEE Proceedings-Generation, Transmission and Distribution*, 146(6):541–549, 1999.

[100] AK Sinha and JK Mondal. Dynamic state estimator using ann based bus load prediction. *IEEE Transactions on Power Systems*, 14(4):1219–1225, 1999.

[101] S. Sridhar and M. Govindarasu. Model-based attack detection and mitigation for automatic generation control. *IEEE Transactions on Smart Grid*, 5(2):580–591, March 2014.

[102] S. Sridhar and G. Manimaran. Data integrity attacks and their impacts on scada control system. In *IEEE PES General Meeting*, pages 1–6, July 2010.

[103] Chun-Lien Su and Chan-Nan Lu. Interconnected network state estimation using randomly delayed measurements. *IEEE Transactions on Power Systems*, 16(4):870–878, 2001.

[104] R. Tan, H. H. Nguyen, E. Y. S. Foo, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi. Modeling and mitigating impact of false data injection attacks on automatic generation control. *IEEE Transactions on Information Forensics and Security*, 12(7):1609–1624, July 2017.

[105] James W Taylor. An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, 24(4):645–658, 2008.

[106] Kevin Tomsovic, David E Bakken, Vaithianathan Venkatasubramanian, and Anjan Bose. Designing the next generation of real-time control, communication, and computations for large power systems. *Proceedings of the IEEE*, 93(5):965–979, 2005.

[107] Dan J Trudnowski, Warren L McReynolds, and Jeffery M Johnson. Real-time very short-term load prediction for power-system automatic generation control. *IEEE Transactions on Control Systems Technology*, 9(2):254–260, 2001.

[108] Scott Victor Valentine. Canadas constitutional separation of (wind) power. *Energy Policy*, 38(4):1918–1930, 2010.

[109] Gustavo Valverde and Vladimir Terzija. Unscented kalman filter for power system dynamic state estimation. *IET generation, transmission & distribution*, 5(1):29–37, 2011.

[110] Rudolph Van Der Merwe, Arnaud Doucet, Nando De Freitas, and Eric A Wan. The unscented particle filter. In *Advances in neural information processing systems*, pages 584–590, 2001.

[111] Warit Wichakool, Zachary Remscrim, Uzoma A Orji, and Steven B Leeb. Smart metering of variable power loads. *IEEE Transactions on Smart Grid*, 6(1):189–198, 2015.

[112] Wikipedia. High-voltage direct curren.[Online] (2018). Available: https://en.wikipedia.org/wiki/high-voltagedirectcurrent/media/file:hvdceurope.svg.

[113] J. Wu, Y. He, and N. Jenkins. A robust state estimator for medium voltage distribution networks. *IEEE Transactions on Power Systems*, 28(2):1008–1016, May 2013.

[114] Yingmeng Xiang, Lingfeng Wang, and Nian Liu. Coordinated attacks on electric power systems in a cyber-physical environment. *Electric Power Systems Research*, 149:156 – 168, 2017.

[115] Guorong Xuan, Wei Zhang, and Peiqi Chai. Em algorithms of gaussian mixture model and hidden markov model. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 145–148. IEEE, 2001.

[116] Chenguang Yang, Yan Li, Bingzhen Yang, Mingqiu Du, and Shaorong Wang. Dynamic variable-weight least squares for state estimation of distribution network based on data fusion. In *Energy Internet and Energy System Integration (EI2), 2017 IEEE Conference on*, pages 1–5. IEEE, 2017.

[117] Hongying Yang, Hao Ye, Guizeng Wang, and Tongfu Hu. Fuzzy neural very-short-term load forecasting based on chaotic dynamics reconstruction. In *International Symposium on Neural Networks*, pages 622–627. Springer, 2005.

[118] Zheng-lin Yang, Guo-qing Tang, Yan-min Song, and Rong-zhang Cao. Improved cluster analysis based ultra-short term load forecasting method. *Automation of Electric Power Systems*, 24:023, 2005.

[119] J. Zhao, G. Zhang, Z. Y. Dong, and M. La Scala. Robust forecasting aided power system state estimation considering state correlations. *IEEE Transactions on Smart Grid*, pages 1–1, 2017.

[120] Liang Zhao and Ali Abur. Multi area state estimation using synchronized phasor measurements. *IEEE Transactions on Power Systems*, 20(2):611–617, 2005.

[121] Jie-Ying Zhou, Bo-Ming Zhang, Jin-Cheng Shang, Jia Yao, and Man Cheng. Very short-term load forecast based on multi-sample extrapolation and error analysis. *Electric Power Automation Equipment*, 25(2):17–21, 2005.