BRIDGING DEEP LEARNING TO POWER SYSTEM STATE ESTIMATION WITH PMUS

YI HE

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Electrical Engineering

Bridging Deep Learning to Power System State Estimation with PMUs

YI HE

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

November 2021

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

(Signed)

YI HE (Name of student)

To my parents,

Aiguo HE and Biyun YAN,

Abstract

In smart grids, the periodical magnitude-based measurement by remote terminal units (RTUs) will be gradually replaced by real-time phasor measurement unit (PMU)-based measurements in the future. The PMU measurements supply synchronous and precise measurements of voltage/current phasors, i.e., synchrophasors at a rate of up to 60 samples per second. These measurements are widely used for the state estimation (SE), power system stability analysis and fault location, thereby improving the situational awareness of the grid operators.

SE plays a vital role in contemporary energy management systems, where sufficient measurements must be provided to make the system observable. However, in a real power grid with thousands of buses, it is impossible to install PMUs wholly as they are costly. Therefore, it is necessary to investigate more economical and effective PMU placement frameworks to mitigate voltage estimation uncertainty. In addition, various incidents can result in the system unobservability, and SE cannot be implemented appropriately via conventional estimators. On the other hand, another critical issue is that classical SE methods are inapplicable in distribution systems without an ascertained network topology due to frequent reconfiguration actions and limited topology measurements. To cope with these challenges, the generative adversarial network (GAN)-based deep learning frameworks are proposed for the SE tasks of both transmission and distribution systems. The proposed framework is data-driven, model-free, and has a strong capability of handling missing data, which can result in the unobservability of the system per the classical SE method.

In this thesis, a comprehensive study is carried out to mainly investigate PMU-based SE where optimal PMU placement as well as deep learning-based SE algorithm are considered. The research background and purpose of this thesis are presented in

Chapter 1. Chapter 2 proposes a reliability-based probabilistic optimal PMU placement approach to ensure minimal voltage magnitude estimation uncertainty under various operating scenarios, with supplementary PMUs installed in the power grid equipped with the SCADA system. Chapter 3 proposes a data-driven deep learning approach for power system static SE based on conditional GAN. Compared with classical SE methods, the proposed method does not require any prior knowledge of the system model. Without knowing the specific model, GAN can learn the inherent physics of underlying state variables purely with historical samples. Once the GAN model has been trained, it can estimate the corresponding system state accurately given the system raw measurements even with incompletions and corruptions. Chapter 4 proposes a novel data-driven deep learning approach for distribution system SE based on the topology-aware GAN (TAGAN). Compared to conventional methods, the new method can effectively estimate system states given contaminated or even missing measurements under varying network topology, representing the first effort of applying one integral deep learning framework for SE to address the uncertainties involved in both measured states and distribution grid topology simultaneously. Chapter 5 summarizes the whole thesis with some valuable conclusions drawn.

List of Publications Arisen from the Thesis

Journal Papers

1 **Yi He**, Songjian Chai, Zhao Xu, Chun Sing Lai, "Power System State Estimation using Conditional Generative Adversarial Network," *IET Generation, Transmission & Distribution*, vol. 14, no. 24, pp. 5823-5833, 2020.

2 **Yi He**, Zhao Xu, Songjian Chai, "Distribution System State Estimation using Generative Adversarial Network with Spectral Normalization Considering Network Topology Changes," *IET Generation, Transmission & Distribution*, under review.

Conference Papers

1 **Yi He**, Zhao Xu, Songjian Chai, "A Novel Approach for State Estimation Using Generative Adversarial Network," *2019 IEEE International Conference on System, Man, Cybernetics (SMC)*, Bari, Italy, 2019, pp. 2248-2253.

2 **Yi He**, Zhao Xu, Jiayong Li, Yufei He, Loong Chan, "Distributed Optimization-based Voltage Regulation in Distribution Network via Battery Energy Storage Units," *24th International Conference on Electrical Engineering (ICEE2018)*, Seoul, Korea, 2018.

3 Yufei He, Ming-Hao Wang, Zhao Xu, **Yi He**, "Advanced Intelligent Micro Inverter Control in the Distributed Solar Generation System," *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, Changsha, China, 2019.

4 Yufei He, Zhao Xu, Youwei Jia, **Yi He**, "Coordinated low voltage ride through control for the hybrid solar-wind micro-grid system," *24th International Conference on Electrical Engineering (ICEE2018)*, Seoul, Korea, 2018.

Acknowledgements

This work is accomplished at the Department of Electrical Engineering, The Hong Kong Polytechnic University (HKPolyU) during my PhD candidature period from 2016 to 2021.

I am very grateful to all the people who helped and supported me during my Ph.D. study at The Hong Kong Polytechnic University. First and foremost, I would like to express my deepest appreciation to my chief supervisor, Prof. Zhao Xu, for his continuous guidance, encouragement, support, and patience throughout my entire PhD study. He is always very kind and patient in offering insightful suggestions to improve my research works. I will remember his kindness and patience during my lifetime.

I would also like to express special thanks to my parents Mr. Aiguo He, Mrs. Biyun Yan for their unconditional love, support, and encouragement. My thanks also go to my research colleagues who make progress together with me, inspire each other and become better than ever before during my precious lifetime in Hong Kong. Especially, I would thank the seniors, Dr. Songjian Chai, Dr. Jiayong Li, Dr. Youwei Jia, Dr. Minghao Wang, and Dr. ChunSing Lai, for their help and suggestions on my research.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BCSE	Branch-current state estimation
CT	Current transformer
DCNN	Deep convolutional neural network
DFT	Discrete Fourier transform
EMSs	Energy management systems
GA	Genetic algorithm
GAN	Generative adversarial network
GPS	Genetic algorithm Global Positioning System
LAV	Least absolute value
LMS	Least median of squares
LTS	Least trimmed squares
MAE	Mean average error
MCC	Main Control Center
MLP	Multilayer perceptron
OPF	Optimal power flow
OPP	Optimal PMU placement
PDC	Phasor data concentrator
PMU	Phasor measurement unit
PLF	Probabilistic load flow
PT	Potential transformer
SE	State estimation
SN	Spectral normalization
SCADA	Supervisory control and data acquisition
UTC	Coordinated Universal Time
WAMS	Wide area measurement system
WLAV	Weighted least absolute value
WLS	Weighted least-squares

Chapter 1 Introduction

1.1 Research Background

Among smart grid monitoring and surveillance technologies, phasor measurement units (PMUs) are becoming the most widely used advanced measuring equipment for real-time monitoring and control [1, 2]. They are the fundamental components of the wide-area measurement system (WAMS). With the development of the global positioning system (GPS), it becomes possible to directly measure the synchronized and real-time voltage and current phasors at widely dispersed locations of smart grids with PMUs [3]. Compared to the conventional supervisory control and data acquisition (SCADA) measurements, PMUs have the advantages of synchronization, higher measurement precision, and higher sampling rates [4, 5]. Considering their outstanding advantages, PMUs have been used to provide critical measurement data for the power system state estimation (SE), data-driven power system stability assessment, fault detection, and system protection, thereby improving the situational awareness of the grid operators [6, 7].

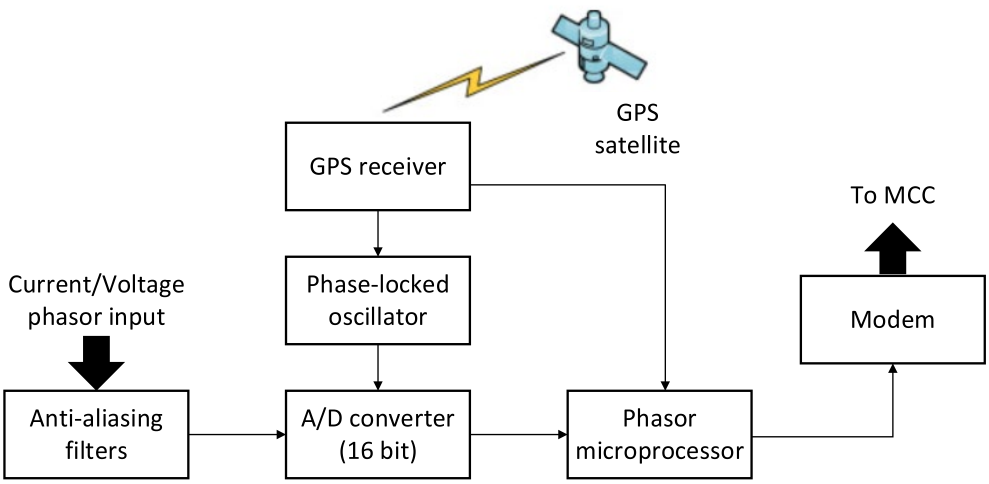


Fig. 1-1. Functional block diagram of PMU.

Fig 1-1 depicts the functional block diagram of a PMU, and procedures of the measurement are described in the following [8]. First, the GPS receiver synchronized by the GPS satellite provides the one pulse-per-second (PPS) signal with a time tag. Then, to take the sample of the analog signals, the l-PPS signal is divided into the required number of pulses by the phase-locked oscillator. The analog signals are obtained from the secondary side of the voltage transformer (PT) and the current transformer (CT) filtered out through anti-aliasing filters [9]. This will restrict the bandwidth of the signal to satisfy the Nyquist–Shannon sampling theorem. The filtered signals are then converted to digital with A/D converters. By using Discrete Fourier Transform (DFT), phasor values computed from digital signals are fed to phasor microprocessor [10] and finally, to the Main Control Center (MCC) by modem.

A PMU deployed on the system bus is able to measure this bus's voltage phasor and the neighboring branches' current phasor. With the increase in the availability of substation's PMUs, the performance of various essential functions such as monitoring, protection, and control of the associated system has been enhanced [11, 12, 13, 14]. Therefore, the PMU-based wide-area measurement system (WAMS) has become important to guarantee safety and stability in the power system [15]. In this sense, if each bus of the network is installed with a PMU, the voltage phasors for the entire system can be directly and fully acquired [16]. Nevertheless, it is impossible for a practical power system with thousands of buses to cover every bus with a PUM due to the heavy cost of PMUs and their networking communication system, which makes the current penetration of PMUs far from the desired level [17, 18]. Therefore, how to realize economical and effective placement of supplementary PMUs considering the existing SCADA system while ensuring minimal voltage magnitude estimation uncertainty deserves careful investigations. Meanwhile, the reliability of measurement devices and the uncertainty of system operation should also be considered.

As the main application of PMUs, SE plays a vital role in contemporary energy management systems (EMSs) [19, 20]. The accurate SE is essential for power system control, optimization, and security analysis [21]. The classical SE solves an optimization formulation with raw measurements and a network model, of which the purpose is to identify the most likely estimate of the system state, i.e., the estimated state [22, 23]. SE assesses whether load-flow constraints are met with measurements computed by the network model [24]. While the system observability is the prerequisite to the classical SE methods, the network is, however, not always fully observable due to malfunction of measurement devices, miss of measurement data, or interference by malicious attacks [25]. Thus, it is highly desirable to effectively restore all system states with corrupted measurements or even missing measurements under contingencies when the system is not fully observable.

The topology of the network can be defined directly by the status of switching devices [26]. Nevertheless, the fundamental topology could be altered partially by the cause of local events like line outages, switching events, and faults [27]. The classical SE involves complicated and separated steps to deal with network topology and SE calculation, etc. [28]. In addition, in transmission systems, the system topology is generally deemed to be very constant [29]. Therefore, many methods premise that the system topology is unchanged and completely known based on monitoring devices. However, this assumption is invalid for the distribution systems because it is nearly impossible to monitor distribution grids topology of considerable size [30]. Therefore, how to deal with distribution system SE with topology changes with little or very limited topology monitoring is worthy of investigation.

1.2 Literature Review

1.2.1 OPP Methods Review

Actually, the optimal PMU placement (OPP) is essentially regarded as an NP-complete problem. In an N-bus power system, the possible solution combinations of OPP are 2^N [33]. Thus, it is deemed a combinatorial optimization problem, and relevant works have been developed within this field [34]. The OPP problems can be categorized into two classes according to their objectives. The first class aims to deploy a minimum number of PMUs on strategic buses to achieve an observable system [21]. The second class is widely formulated for specific applications [22]. More details are reviewed in the following.

For the first class OPP problems, conventionally, the basic constraint of surveillance infrastructure was the observability of the overall network and components, and most OPP methods adopt an optimization constraint based on the observability of basic scenarios on networks. These works may adopt a dual search algorithm as well as the immunity GA [35], weighted least squares algorithm [36], genetic algorithms (GA) [37], recursive Tabu search method [38], integer linear programming [39], and simulated annealing [40]. Further research suggests that the number of optimal solutions for the OPP problem may be more than one, and the optimal solution with the largest redundancy of measurement can be more desirable because this solution leads to a more accurate SE as well as greater robustness in opposition to component failure [41]. Other works emphasized the increasing reliance on WAMS applications for power system operation, as well as making WAMS infrastructure more stable and reliable. These efforts solve the OPP problem with groups of constraints to ensure that the system is fully observable when the transmission line and/or PMU is interrupted [42].

For the second class OPP problems, they are formulated to achieve specific applications or functions. For example, defense against data injection attacks [43], parameter error identification [44], minimized SE errors [45], fault location observability [46], improved topology error handling [47], reduction of SE error variance and improvement local redundancy [48], bad data detection in SE [12] and optimizing useful indicators of the power system state estimation are three important requirements: observability, performance, and convergence [49]. PMUs can also be critically placed in the power grids to identify parameter errors on unilateral cut sets (critical branches) or bilateral cut sets (critical branch pairs) [11, 50, 51].

Among these published works, [36], [39], [41], [44], [45] and [48] incorporated conventional SCADA measurements. In these works, conventional SCADA measurements have been deployed in the power grid, and new PMUs have to be placed in the presence of these SCADA measurements. These existing SCADA measurements could be beneficial to the system observability or specific applications by decreasing the number of needed PMUs [52]. The feasible practice is to deploy PMUs incrementally in conjunction with the SCADA measurements [53].

SE utilizes the redundant measurements acquired from power grids to filter out the measurement errors and supply the most likely estimation of the operating conditions of the system [58, 59]. In [59], it is presented that the accuracy of SE can be significantly enhanced by deploying PMUs incrementally in conjunction with the SCADA measurements, through which the associated uncertainties of voltage magnitude estimation can be quantitively measured based on branch-current SE (BCSE) algorithms. Thus, in order to improve the accuracy of SE in the system with SCADA measurements and the limited number of PMUs, it is highly desired to find an OPP solution considering voltage magnitude estimation uncertainty.

Besides, most of the previous OPP methods have considered the losses of the PMU to design a reliable and robust WAMS [54, 55]. However, these methodologies do not take into account the random outages of the PMU measurement system, including the PMU, the phasor data concentrator (PDC), the communication system, and other components that can affect the power system observability [56]. The random occurrence of these events may cause part of the network to be unobservable, thereby endangering the security of the power system [57]. In addition, under varying operating scenarios, the optimal PMU placement solution may be different. However, conventional OPP methods only concentrate on one scenario to develop the OPP plan that may not be fit for other operation scenarios, which might cause biased solutions. Thus, it is necessary to comprehensively consider the operating uncertainties to obtain an unbiased solution suitable for different operating conditions.

1.2.2 SE Methods Review

The studies of SE with PMU measurements have already been extensively carried out, and the approaches are either based on the combination of SCADA and PMU measurements [26, 66, 67, 68, 69, 70, 71] or purely PMU measurements [64, 72, 73, 74]. The former ones can be further divided into two categories: namely hybrid state estimator method and the multi-stage method. The hybrid state estimator combines PMU measurements and conventional SCADA measurements by using a non-linear transformation to connect the traditional state vector in polar form with the voltage phasors in the rectangular form [66, 67, 68, 69]. It is proved that incorporation of PMU measurements can significantly improve the SE performance as compared to that with only SCADA measurements under steady-state conditions. Nevertheless, the accuracy of those approaches can be compromised due to different time scales between PMUs and SCADA systems [26]. To address the issues of time scale inconsistency, the multi-stage approach is adopted by processing PMU measurements or SCADA measurements in independent stages. A Bar-Shalom-Campo data fusion technique is applied to combine the results of different PMU and SCADA stages in [70, 71]. All those methods based on the combination of SCADA and PMU measurements adopt various principles to enhance the robustness against gross errors. On the other hand, the purely PMU-based SE method has shown various benefits compared with purely conventional SCADA based or PMU-SCADA-based SE methods [73, 75, 76]: (i) The measurement function is linear as only current or voltage phasors are measured by PMUs, which gets rid of the computationally expensive iterative process in traditional SE with SCADA measurements. (ii)

No reference bus is needed since the voltage/current phasor phase angle can be directly acquired at the same time with time-stamped from the GPS [77]. (iii) Real-time SE can be realized due to high sampling rates and low latency of PMUs.

The aforementioned SE methods are all based on weighted least squares (WLSs) or their variants. These methods are well-known non-robust, and a single outlier can severely alter the estimation results. To overcome this drawback, a separate post-estimation bad data processing function is needed to detect and eliminate gross errors [23]. In addition, robust SE methods are recently proposed to further enhance the robustness against different bad data situations by introducing, e.g., adaptive pre-processing steps, etc. [27]. The least absolute value (LAV) estimator is an alternative technique. By minimizing the L1 norm (rather than L2 in WLS) of measurement residuals, the LAV estimator can be executed via linear programming (LP) solvers, and it will detect and eliminate gross errors accordingly [78]. The authors in [79] formulated a robust LAV estimator by using PMUs to improve the computational performance of the LAV estimator. A hybrid state estimator was proposed in [80] with the coexistence of PMU measurements and SCADA measurements, to determine the states based on the weighted LAV (WLAV). Besides, WLAV is also one of the most common robust estimators with high performance in the aspect of robustness [81, 82]. Other enhanced estimators for robustness include the method of least median of squares (LMS) and least trimmed squares (LTS), and they aim to restrain the effects of bad data by alternatively using measurements [83, 84]. The LMS method is proposed depending on the notion that the median of a set of values can be more effective than the mean in the estimation process though it has got the particular disadvantage of rejecting several normal measurements along with the outliers. The LTS method calculates the sum of squared errors for the smallest residuals only.

The main issue of the above-mentioned studies, including WLS-based SE methods and robust SE methods, is that the system must be pre-assumed to be fully observable, and they cannot work effectively when the system is unobservable. However, in reality, many contingencies can result in system unobservability, including measurement loss, line outage, failure of the data concentrator, or failure of the local communication system [24]. The network is unobservable if any state variables cannot be uniquely computed for a given set of measurements and network topology [25]. In unobservable networks, SE cannot be implemented expectedly via WLS-based estimators or robust estimators. Hence the system operator cannot monitor any violations or events in these unobservable buses, which may lead to catastrophic outcomes. The traditional solution for unobservability is to use pseudo-measurements to replace the missing measurement. The pseudo-measurements are generated based on external processes such as historical data, prediction procedures, load curve assessment, or derived from interpolated observations [28]. Due to the nature of the pseudo-measurements, the accuracy of SE cannot be guaranteed, thus failing to satisfy the SE requirement. Besides, it is difficult to implement real-time SE as a result of the sparse data rate of pseudo-measurements with finite source data. Other studies are conducted to handle this issue. In [109], auto-associative neural networks (NNs) or autoencoders are used to reconstruct missing data in SE. In [98] and [106], multilayer perceptron (MLP) based NNs are applied for SE. Once trained offline using historical data and/or simulated samples, NNs can be implemented for real-time SE. Limited by the development of AI technology during that period, the accuracy of SE was not outstanding, especially for large-scale systems.

1.2.3 DSSE Methods Review

When distribution system state estimation (DSSE) was firstly studied in [112], a weighted least-squares (WLS)-based DSSE solver was proposed using a three-phase nodal voltage formulation. To handle topology issues in DSSE, some methods have been proposed to identify the correct topology before performing SE. [99] proposes a computationally efficient approach to diagnose the lines' statuses, and the fastest change detection method is applied to identify network topology in [100]. These methods are developed to decouple the problem into topology identification and state estimation. Nevertheless, such implementations might lose the essential information about the system topology hidden in the raw measurements [101]. Therefore, the integral approaches of topology processing and SE have been developed. To handle this problem in decoupled methods, a generalized state estimation (GSE) model was proposed in [102], where the status of lines can be determined through diagnosing the open breakers and topology error (measurement error of switches status) [103]. In [104], the authors proposed the algorithm that can deal with two different types of criteria composing the objective function of estimation, namely: weighted least absolute value (WLAV) criteria for topologically relevant relations and WLS criteria for estimating system states. Nevertheless, the computational cost of these methods is high due to many iterations involved [105]. Besides, the GSE model is usually required to include the extra branches' status variables on the SE formulation to identify and fix topology errors [15].

To resolve the aforementioned issues existing in GSE, a number of data-driven and probabilistic approaches have been developed in DSSE, such as auto-encoders [106], probabilistic recursive Bayesian approach [108], correlation analysis [109], fuzzy-based pattern recognition [110]. However, these approaches generally involve a data-driven search process within a finite topology space (i.e., topology library). Due to the various operation needs, numerous possible configurations of distribution grid topology exist given certain combinations of line switch statues (2N, N is the number of buses). Once the input topology is out of the dataset for training, i.e., an out-of-sample topology, these DSSE methods perform poorly.

1.3 Purpose of the Thesis

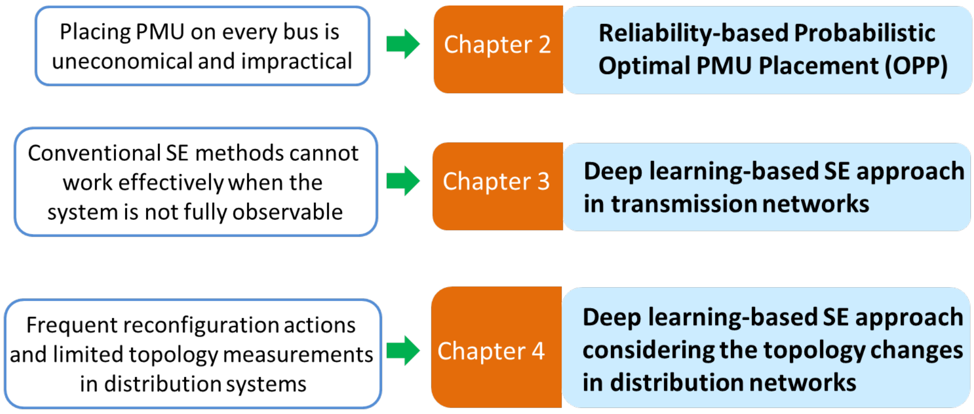


Fig. 1-2. Schematic diagram of the thesis.

Fig. 1-2 illustrates the schematic diagram of the thesis. Considering the aforementioned situations, including the need for real-time monitoring of the modern power system and the advance of modern monitoring technologies in power grids, more PMUs will be deployed in the transmission system and the distribution system [31, 32]. Also, due to the aforementioned deficiencies of classical SE methods as well as the rapid development of AI technologies, this thesis intends to investigate PMUs based power system state estimation where the optimal placement of supplementary PMUs considering the existing SCADA system as well as deep learning based SE is considered. Therefore, the purpose of the thesis can be summarized in three aspects. The first one is to obtain minimal voltage estimation uncertainty with limited PMU numbers based on various operating scenarios; the second one is to accurately estimate all system states considering the corrupted measurements or even missing measurements under contingencies when the system is not fully observable in transmission networks; the third one is to apply one integral framework for SE to address the uncertainties involved in both measured states and grid topology in distribution networks. More details are illustrated as follows:

Firstly, this thesis aims at proposing a reliability-based probabilistic OPP approach to obtain minimal voltage magnitude estimation uncertainty based on various operating scenarios, with supplementary PMUs installed in the power grid, which is observable via the SCADA system. The reliability of PMU measurements should be modeled when estimating the system states. Also, their components' random outages should be considered. Meanwhile, the operating uncertainties, including the load patterns and power generations, are expected to be taken into account in a stochastic manner.

Secondly, this thesis aims at proposing a novel deep learning based SE approach in transmission networks. On the one hand, the actual correlations of system states should be well captured, and the system states should be accurately estimated with corrupted or missing measurements. On the other hand, the proposed SE method is expected to be effective even in an unobservable network.

Thirdly, this thesis aims at proposing a novel deep learning based SE approach considering the topology changes in the distribution system. On the one hand, one integral SE framework should be proposed to stay away from the complicated and separated steps in classical SE methods. On the other hand, in this framework, the uncertainties consisting of measured states and grid topology should be addressed simultaneously. Moreover, the proposed SE method should be capable of tackling a variety of out-of-sample topologies, are out of the topology library.

1.4 Primary Contributions

To achieve the research objectives, the main contributions achieved in this research are summarized as follows:

1) A comprehensive study is conducted to develop a cost-effective OPP approach that can help to mitigate voltage estimation uncertainty. The objectives include a minimal number of PMUs and minimal voltage amplitude estimation uncertainty. PMU measurement reliability is modeled when estimating the system states. In the modeling of PMU measurement reliability, PMU measurement system components' random outages are considered. These random outages may lead to the unobservability of a portion of the network, which may endanger the power system's safety. So, in the estimation process, it is important to take these random outages into account. Furthermore, probabilistic load flow (PLF)

is applied in the study to represent different operating scenarios. In this way, the load patterns and power generations are considered stochastically as the operating uncertainties such that the obtained PMU placement solution is unbiased for planning purposes. With PLF carried out, the OPP scheme is more suitable for various operating scenarios as different operating uncertainties are accordingly considered.

2) A novel data-driven and model-free deep learning approach for power system SE is proposed. By applying conditional GAN, the actual correlations of system states can be well captured, and the system states can be accurately estimated without prior knowledge of the system model. The PMU-based SE method can effectively restore all system states considering the corrupted raw measurements or even missing measurements under contingencies. Thus, the SE process can still be implemented even in an unobservable network. The influence of data contaminations is thoroughly investigated with respect to different data contamination ratios and types. The experiment is carried out on a large system, i.e., 2746-bus Polish system. The simulation results validate the effectiveness of the proposed method, and all estimated system states are close to true system states. To the best of the authors' knowledge, this is the first work using deep learning models for power system SE processes on a large-scale system.

3) A novel data-driven TAGAN model is proposed for DSSE, which represents the first effort of applying one integrated deep learning framework for SE that is capable of addressing the uncertainties involved in both grid topology and state measurement simultaneously. Unlike the

existing data-driven approaches that can only handle a finite topology space, the proposed method by applying conditional GAN with spectral normalization is capable of tackling a variety of out-of-sample topologies. Besides, a detector and hinge loss function is used in the TAGAN model to improve SE accuracy. Extensive experiments have been carried out to examine the influence of data contaminations with respect to different ratios and types, which are rarely considered by most existing works. Thus, the proposed method is proved to be robust to the corrupted measurements or missing measurements, making the SE viable even in an unobservable network.

1.5 Thesis Layout

The remainder of the thesis is organized as follows:

In Chapter 2, provided that the power grid is observable via the SCADA system with enough redundancy, a reliability-based probabilistic optimal placement of supplementary PMUs approach is proposed to minimize voltage magnitude estimation uncertainty in SE based on various operating scenarios. In this Chapter, PMU measurement reliability is modeled when estimating the system states. In the modeling of PMU measurement reliability, PMU measurement system components' random outages are considered. These random outages may cause part of the network to be unobservable, which may endanger the power system's safety. Therefore, in the estimation process, it is important to take these random outages as well as PMU measurement reliability into account. In addition, unlike the traditional OPP problem, this method takes into account operating uncertainties randomly, including power generation and load patterns. In order to better address these uncertainties, PLF is applied to represent different operating scenarios and then obtain a PMU placement solution for the planning. Finally, the IEEE 9-bus and IEEE 14-bus systems are used to verify the effectiveness of the proposed OPP model.

Chapter 3 proposes a model-free and fully data-driven deep learning approach for power system static SE based on conditional GAN. Unlike the power grid is observable via the SCADA system in chapter 2, in this chapter, all measurements are assumed to be provided by PMUs. Compared with the conventional SE approach, e.g., the WLS based methods, any appropriate knowledge of the system model is unnecessary for the proposed method. Without knowing the specific model, GAN can learn the inherent physics of underlying state variables purely relying on historical samples. Once the model has been trained, it can estimate the corresponding system state accurately based on the system raw measurements, which are sometimes characterized by incompletions and corruptions in addition to noises. Case studies on the IEEE 118-bus system and a 2746-bus Polish system validate the effectiveness of the proposed approach, and the mean absolute error is less than 1.2e-3 p.u. and 5.3e-3 rad for voltage magnitude and phase angle, respectively, which indicates a high potential for practical applications.

Chapter 4 proposes a novel data-driven and model-free TAGAN-SE approach considering varying topologies for distribution system SE. In this chapter, all measurements are assumed to be provided by PMUs. In the TAGAN-SE, the detector is designed to generate a “mask” as GAN’s input, and the hinge loss function is applied to enhance the training process. Besides, the spectrum normalization is employed in the discriminator enabling the discriminator function to be Lipschitz continuous and solve the problems (including vanishing gradients and mode collapse) in the classical GAN, which improves the performance with varying topologies, especially for the out-of-sample topology. Different measurement contamination types and varying topologies, including the seen topologies during the training and the out-of-sample topologies, were examined in the model testing. The effectiveness of the proposed TAGAN-SE is verified with IEEE 33-node distribution system and IEEE 118-bus distribution systems.

Eventually, Chapter 5 summarizes this thesis with some valuable conclusions drawn.

Chapter 2 Reliability-based Probabilistic Optimal PMU Placement Considering State Estimation Uncertainty

2.1 Introduction

Phasor measurement unit (PMU) is a device used to synchronize wide-area measurements and record the measurement time with high accuracy in the power system, i.e., less than one microsecond [1]. With the assumption that PMUs have enough channels, the PMU deployed on the system bus is able to measure this bus's voltage phasor and the neighboring branches' current phasor. With the increase in the availability of substation's PMUs, the performance of various basic functions such as monitoring, protection, and control of the associated system has been enhanced [8]. As a result, the PMU-based wide-area measurement system (WAMS) has become an important measure to guarantee safety and stability in the power system. As an essential part of any WAMS, PMU has received significant attention for various research topics, including optimal PMU placement.

As the main application of PMUs, SE plays a vital role in contemporary energy management systems (EMSs). The accurate SE is essential for power system control, optimization, and security analysis [21]. Conventional SE is based on supervisory control and data acquisition (SCADA) system with the measurements of the remote terminal units (RTUs). Since SCADA measurements are less accurate, asynchronous, and have a low sampling rate, they are difficult to capture fast changing system dynamics [10]. Thus, in many studies, PMUs are placed in the power system to improve the SE process benefiting from the advantages of PMUs. The easiest way to be fully monitored and maximum measurement redundancy is that all buses are deployed on PMUs. Nevertheless, because of the high expenses of PMU, this is almost impossible in practice. Besides, it is expensive to connect many PMUs to the communication network of the control center [20]. Thus, the available PMUs are limited in power grids. Considering the current situation that the existing SCADA measurements are extensively deployed in the system, and the number of PMUs is limited, now it has been broadly accepted that the SCADA measurements and PMUs can be jointly employed to achieve satisfactory SE in the system. Therefore, the optimal PMU placement (OPP) problem can be formed into a constrained optimization problem, with supplementary PMUs installed in the power grid, which is observable via the SCADA system.

In this chapter, provided that the power grid is observable via the SCADA system with enough redundancy, a reliability-based probabilistic OPP approach is proposed to obtain minimal voltage magnitude estimation uncertainty based on various operating scenarios, with supplementary PMUs installed in the power grid, which belongs to above mentioned the second class of OPP problems. PMU measurement reliability is modeled when estimating the system states. In the modeling of PMU measurement reliability, PMU measurement system components' random outages are considered. These random outages may cause part of the network to be unobservable. Hence the system operator cannot monitor any violations or events in these unobservable buses, which may endanger the security of the power system. Besides, unlike the conventional OPP problem, this method takes into account operating uncertainties in a random manner, including power generation and load patterns. In order to better address these uncertainties, Probabilistic Load

Flows (PLFs) are applied to reflect a variety of practical operating scenarios, and then an OPP solution for PMU installation can be investigated.

2.2 State Estimation Uncertainty

2.2.1 Branch Current State Estimation (BCSE)

Two main categories of WLS algorithms have been conceived for SE: 1) node- voltage (NV) and 2) branch current (BC) estimators. The main difference among the available approaches is the choice of the state variables to be used within the algorithm. When the same settings of measurements are used, the WLS algorithms provide the same accuracy performance despite the choice of the state variables. Besides, BC estimators allow achieving fewer average execution times than NV estimators. The difference of average execution times becomes more significant in the system with PMUs because the gain matrix of WLS is constant [60]. Therefore, in this Chapter, the branch-current state estimator proposed in [60] (BCSE) is adopted.

The general measurement model adopted for SE is

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \tag{2.1}$$

where $\mathbf{z} = [z_1 \ldots z_M]^T$ denotes the vector of the M measurements obtained from the system, $\mathbf{x} = [x_1 \ldots x_N]^T$ denotes the vector of the N state variables, $\mathbf{h} = [h_1 \ldots h_M]^T$ represents the vector of the measurement functions, and $\mathbf{e}$ denotes the vector of the measurement noise, which is generally set as random variables with zero mean and covariance matrix $\Sigma_{\mathbf{z}}$.

In BCSE, the state vector $\mathbf{x}$ needs to comprise a reference bus voltage and the rectangular currents in the N_{br} branches. If synchrophasor measurements are used, the state vector $\mathbf{x}$ can be written as

$$\mathbf{x} = [V_s, \theta_s, i_1^r \ldots i_{N_{br}}^r, i_1^x \ldots i_{N_{br}}^x]^T \tag{2.2}$$

with $N = 2 + 2N_{\mathrm{br}}$ elements, where V_s and θ_s are the voltage amplitude and phase angle of the slack bus chosen as a reference, while i^r and i^x are the currents' real part and imaginary part, respectively. It is important to highlight that the slack bus can be arbitrarily chosen because it is only necessary to complete the state in the BC formulation in order to estimate the voltage profile accurately [47]. However, different choices of the slack bus do not affect the estimates and their uncertainties. For the case of conventional measurements, the slack voltage phase angle is now included and can be estimated. Furthermore, all the phase angles can be made a reference to the absolute reference given by the Coordinated Universal Time (UTC).

In the BCSE method, the voltage and current are estimated iteratively through alternate forward scanning and WLS steps. Therefore, the SE problem can be formulated as a WLS optimization problem [60]:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{e}^T \mathbf{W} \mathbf{e} \tag{2.3}$$

where $\hat{\mathbf{x}}$ is the estimated state vector, and $\mathbf{W}$ is the weight matrix.

During the process of WLS, the estimation of the state vector is updated by calculating the normal equation:

$$\Delta \mathbf{x}_n = \mathbf{x}_{n+1} - \mathbf{x}_n = \mathbf{G}_n^{-1} \mathbf{H}_n^T \mathbf{W} [\mathbf{z} - \mathbf{h}(\mathbf{x}_n)] \tag{2.4}$$

where $\mathbf{x}_n$ is the state vector in iteration n, $\mathbf{H}_n$ denotes the Jacobian of the measurement function concerning the state variable, $\mathbf{G}_n = \mathbf{H}_n{}^T \mathbf{W} \mathbf{H}_n$ denotes the Gain matrix. Coherently with the known measurement properties, for an efficient WLS estimator, $\mathbf{W}$ is given as the inverse of the covariance matrix $\Sigma_{\mathbf{z}}$ of the measurement errors.

A forward scan computation tracks the WLS process during each estimation algorithm's iteration. Starting from the final estimate of the branch current and slack bus voltage, the forward scan step is able to calculate the network voltage for each bus by directly calculating the voltage drop along the line. Once the state vector's updates $\Delta \mathbf{x}_n$ are less than the selected tolerance, the algorithm will stop.

2.2.2 Voltage Magnitude Estimation Uncertainty

By inverting the gain matrix applied at the final iteration during the estimation procedure, the estimated state's covariance matrix is able to be acquired. Therefore, the dimension of $\mathbf{G}^{-1}$ is $N \times N$, and its diagonal is the estimated state's variance while the rest is covariance. Especially in consideration that the state vector is expressed as (2.2), element $\mathbf{G}^{-1}$ (1, 1) represents the variance $\sigma^2_{\hat{V}_s}$ with respect to the reference bus voltage amplitude estimate $\hat{V}_s$, while the element $\mathbf{G}^{-1}$ (2, 2) gives the variance $\sigma^2_{\hat{\theta}_s}$ of the estimated phase angle. In the following sections, the analysis will concentrate on the slack bus voltage, and the result is able to be extended to the whole system's nodes because the reference bus can be selected randomly.

To analyze the variances of the estimated slack bus voltage amplitude as well as phase angle, the Gain matrix is split into four blocks:

$$\mathbf{G} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \tag{2.5}$$

where $\mathbf{A}$ denotes a 2×2 matrix, $\mathbf{B}$ denotes a $2 \times N_{br}$ matrix, $\mathbf{C} = \mathbf{B}^T$ (for the symmetry of the Gain matrix) has $N_{br} \times 2$ size, and $\mathbf{D}$ is a $N_{br} \times N_{br}$ squared matrix.

The inverse of this block matrix is written as follow:

$$\mathbf{G}^{-1} = \begin{bmatrix} (\mathbf{A}-\mathbf{BD}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{BD}^{-1}\mathbf{C})^{-1} & (\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix} \tag{2.6}$$

Focusing on the 2×2 covariance matrix of the slack bus voltage phasor $\Sigma_{\mathbf{s}}$, it is possible to use the Woodbury matrix identity to obtain

$$\begin{aligned} \Sigma_{\mathbf{s}} &= \mathbf{G}^{-1}\Big|_{m=1,2;n=1,2} = (\mathbf{A}-\mathbf{BD}^{-1}\mathbf{C})^{-1} \\ &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \end{aligned} \tag{2.7}$$

The second block in the main diagonal of (2.6) is the covariance matrix $\Sigma_{\boldsymbol{I}}$ of the rectangular current estimations, thus (2.7) can be expressed as follows:

$$\Sigma_{\mathbf{s}} = \begin{bmatrix} \sigma^2_{\hat{V}_s} & \sigma_{\hat{V}_s,\hat{\theta}_s} \\ \sigma_{\hat{V}_s,\hat{\theta}_s} & \sigma^2_{\hat{\theta}_s} \end{bmatrix} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\Sigma_I\mathbf{B}^T\mathbf{A}^{-1} \tag{2.8}$$

where $\sigma^2_{\widehat{V}_s,\widehat{\theta}_s}$ is the covariance between magnitude and phase angle of the slack bus voltage estimation. Equation (2.8) is the generalization to synchronized measurements of the expression for conventional ones.

To understand which terms are involved in the uncertainty expression of the voltage estimations, it is necessary to analyze the contributions, pertaining to different measurement types, forming the Gain matrix. Three types of measurements can be distinguished: 1) voltage magnitude measurements, 2) voltage phase angles, and 3) the other measurements (powers or currents). The gain matrix is able to be resolved via differentiating augmentations from voltage synchrophasor measurements or other measurements. In particular, in the BCSE, all power measurements are transformed into the same current measurements. As the power measurement's influence on the voltage state vanishes, this results in a slight approximation of the gain matrix. However, this approximation exists in the gain matrix only, while it is not in the BCSE results because considering the final estimate of the voltage curve, the equivalent measurement value will be perfected in each iteration. The Gain matrix can be written as

$$\begin{aligned}\mathbf{G} = \mathbf{H}^T\mathbf{W}\mathbf{H} &= [\mathbf{H}_V^T \quad \mathbf{H}_\theta^T \quad \mathbf{H}_I^T]\begin{bmatrix}\mathbf{W}_V & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_\theta & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{W}_I\end{bmatrix}\begin{bmatrix}\mathbf{H}_V\\ \mathbf{H}_\theta\\ \mathbf{H}_I\end{bmatrix} \\ &= \mathbf{H}_V^T\mathbf{W}_V\mathbf{H}_V + \mathbf{H}_\theta^T\mathbf{W}_\theta\mathbf{H}_\theta + \mathbf{H}_I^T\mathbf{W}_I\mathbf{H}_I = \mathbf{G}_V + \mathbf{G}_\theta + \mathbf{G}_I\end{aligned} \tag{2.9}$$

where V, θ, and I indicate voltage magnitudes, voltage phase angles, and the other (power and/or currents) measurements. Correspondingly, $\mathbf{G}_V$, $\mathbf{G}_\theta$, and $\mathbf{G}_I$ are the contributions to the Gain matrix, $\mathbf{W}_V$, $\mathbf{W}_\theta$, and $\mathbf{W}_I$ are the weighting submatrices, and $\mathbf{H}_V$, $\mathbf{H}_\theta$, and $\mathbf{H}_I$ are the associated Jacobians.

By maintaining the same terminology, it is possible to split the contributions of (2.5). Since the current measurements are able to be directly represented by using the corresponding current state variable and thus there is no derivative term about the slack bus voltage in the Jacobian, $\mathbf{A}_I$, $\mathbf{B}_I$, and $\mathbf{C}_I$ are null matrices, and it illustrates that

$$\mathbf{G} = \begin{bmatrix}\mathbf{A}_V + \mathbf{A}_\theta & \mathbf{B}_V + \mathbf{B}_\theta\\ \mathbf{B}_V^T + \mathbf{B}_\theta^T & \mathbf{D}_V + \mathbf{D}_\theta + \mathbf{D}_I\end{bmatrix} \tag{2.10}$$

Taking into account the derivative terms appearing in the Jacobians $\mathbf{H}_V$ and $\mathbf{H}_\theta$, the following expression for A can be found:

$$A = \sum_{i\in\Lambda_i}\begin{bmatrix} w_{Vi}\cos^2\theta_{is} + w_{\theta i}\dfrac{\sin^2\theta_{is}}{V_i^2} & V_s(w_{Vi} - \dfrac{w_{\theta i}}{V_i})\dfrac{\sin 2\theta_{is}}{2}\\ V_s(w_{Vi} - \dfrac{w_{\theta i}}{V_i})\dfrac{\sin 2\theta_{is}}{2} & V_s^2(w_{Vi}\sin^2\theta_{is} + w_{\theta i}\dfrac{\cos^2\theta_{is}}{V_i^2})\end{bmatrix} \tag{2.11}$$

where i represents the index of the bus with the PMU measurement installed; θ_{is} is the deviations of the phase angle between bus i and the slack bus; w_{Vi} and

$w_{\theta i}$ are, respectively, the weights associated with the voltage amplitude and phase angle measurement in node *i*.

The voltage phase angles θ_{is} are pretty small, and they can be set around to zero in the first approximation. With this assumption, the mutual influences of voltage amplitude and phase angle are decoupled and, from (2.11), the following expression holds for the voltage amplitude uncertainty:

$$\sigma_{\hat{V}_s}^2 = \sigma_a^2 + \sigma_b^2 \simeq \frac{1}{\sum_i w_{Vi}} + \frac{1}{(\sum_i w_{Vi})^2}[\mathbf{b}_1^T \Sigma_I \mathbf{b}_1] \tag{2.12}$$

where $\mathbf{b}_1$ is the transpose of the first row of $\mathbf{B}$.

From (2.12), it is possible to observe that the overall uncertainty of $\hat{V}_s$ is affected by two terms from different sources for the conventional voltage measurements. The first item, σ_a^2, is determined by the amount of PMU voltage amplitude measurements as well as their precision. With the assumption that voltage measurements have the same standard deviation σ_{VPMU}, then we have:

$$\sigma_a^2 \simeq \frac{1}{M_{PMU} w_{VPMU}} = \frac{\sigma_{PMU}^2}{M_{PMU}} \tag{2.13}$$

where M_{PMU} is the entire amount of voltage phasor measurements accessible in the system and w_{VPMU} is the weight of all the voltage measurements.

Concerning the second item σ_b^2 in (2.12), the elements of $\mathbf{b}_1$ with the assumptions mentioned above are

$$\mathbf{b}_1(j) \simeq \begin{cases} \sum_i \lambda_{ji} R_{ji} w_{Vi} & \text{if } j \le N_{br} \\ \sum_i \lambda_{ji} X_{ji} w_{Vi} & \text{if } j > N_{br} \end{cases} \tag{2.14}$$

where *i* represents the index of the bus with the voltage phasor measurement installed; λ_{ji} represents a logic value, and it is 1 when branch *j* belongs to the route between bus i and the slack bus included in the Jacobian and 0 otherwise;

$R_{ji} = (-r_j \cos\theta_i - x_j \sin\theta_i)$ and $X_{ji} = (x_j \cos\theta_i - r_j \sin\theta_i)$ denotes the derivatives of the voltage amplitude measurement at bus i concerning the currents' real part and imaginary part at branch j, respectively. Also, r_j is the resistance, and x_j is the reactance at branch j. Thus, it will result in a large number of elements closely linked with the voltage drops' uncertainty between the selected slack bus and each measurement bus. As a consequence, changing the slack bus impacts on $\mathbf{b}_1$ and then on the uncertainty term σ_b^2, reflecting the different ratios of uncertainty achievable on different nodes. It is crucial to recall that changing the slack bus will not affect the estimates and the uncertainties of the voltage profile; thus, such an approach allows exploiting the presented expressions to analyze better where and how uncertainty arises at each node.

2.3 The Proposed OPP Method

2.3.1 Reliability of Measurement Devices

The definition of the reliability of a measurement is the probability that the measurement is available under these random outages [46]. In the transmission network, PMUs are installed at the substations/buses, and observability of the substation/bus may be lost under some circumstances, including malfunction of the phasor data concentrator (PDC), malfunction of the local communication system as well as unavailable of getting voltage or current measurements. These random outages may cause part of the network to be unobservable. Hence the system operator cannot monitor any violations or events in these unobservable buses, which may endanger the security of the power system. Moreover, the quality of security analysis and system operation is directly affected by the speed, accuracy, and reliability of measurements [47].

Therefore, in the estimation process, it is highly desired to take these random outages into account for a robust solution of the OPP problem, where the reliability-based studies might lead to solutions closer to actual voltage magnitude estimation uncertainty.

As mentioned in section 2.1, in this chapter, the power grid is assumed observable via the SCADA system with enough redundancy. Thus, in this OPP problem, only PMU measurement reliability is considered. Since PMUs are installed in the substation and serve as a part of the substation monitoring system, the measurements obtained from PMUs for the power system depend on the reliable operation of the substation monitoring system. Therefore, the reliability of the PMU measurement is actually the reliability of the substation monitoring system. The substation monitoring system aims to obtain the electrical measurements to send them to the national or regional control center. The reliability of the monitoring system at the substation refers to the probability when measurements can be collected in the substation and sent to the control center. In the substation, the critical elements are the primary measurement devices, which can supply the input signal for PMUs. As illustrated in Figure 2-1, it presents that the substation monitoring system consists of three parts: measurement parts, PDC, and the communication system. In particular, measurement parts include potential transformers (PTs), current transformers (CTs), and the PMU. PTs and CTs are the primary measurement devices, which measure the voltage phase value and current phase value, respectively. Both of them supply the input signal for PMUs. PDC is another critical element, which collects the measurement data according to the time-stamp. Besides, the communication system aims to deliver the measurements data between the PMU and the PDC. The reliable operation of the substation monitoring system needs all parts to operate normally. Thus, measurement parts, PDC, and the communication system should be considered in the reliability of the substation monitoring system.

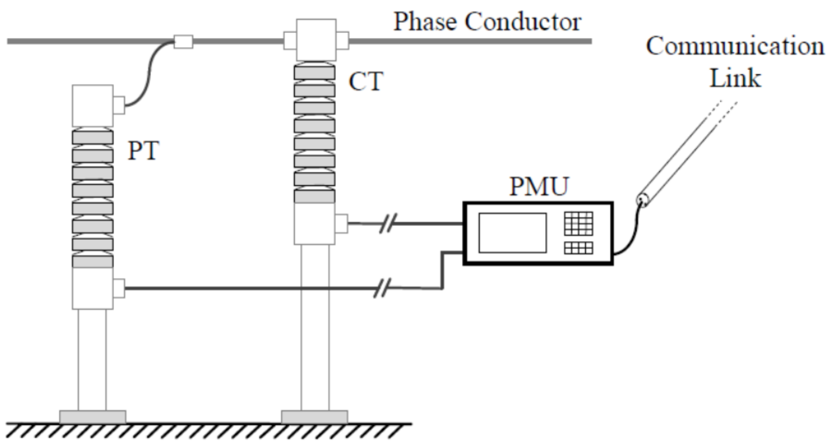


Fig. 2-1. Measurement system installed in the substation.

Thus, the reliability of the monitoring system at the substation is computed as:

$$R_{sub} = (1 - Q_{mea}) \cdot (1 - Q_{pdc}) \cdot (1 - Q_{com}) \tag{2.15}$$

where R_{sub} is the reliability of the monitoring system at the substation, Q_{mea} is the malfunction probability of the measurement parts, Q_{pdc} is the malfunction probability of the PDC, Q_{com} is the malfunction probability of the internal communication system.

There is a malfunction of measurement parts when it is unavailable to take the voltage or the current measurement from PTs and CTs in the substation. It is expressed by

$$Q_{mea} = 1 - (1 - Q_{pmu}) \cdot (1 - Q_{vm}) \cdot (1 - Q_{cm}) \tag{2.16}$$

where Q_{pmu} is the malfunction probability of the PMU; Q_{vm}, Q_{cm} is the malfunction probability of the voltage measurement and current measurement, respectively.

The malfunction probability of the voltage measurement and current measurement can be formed by

$$Q_{vm} = 1-(1-Q_{pt})\cdot(1-Q_{plink}) \tag{2.17}$$

$$Q_{cm} = 1-(1-Q_{ct})\cdot(1-Q_{clink}) \tag{2.18}$$

where Q_{pt}, Q_{ct} is the malfunction probability of PTs and CTs; Q_{plink}, Q_{clink} is the malfunction probability of the link between the PTs or CTs and the PMU.

2.3.2 Probabilistic Load Flow (PLF)

As under varying operating scenarios, the OPP solution may be different. However, conventional OPP methods only concentrate on one scenario to develop the OPP plan that may not be fit for other operation scenarios, which might cause biased solutions. Thus, it is necessary to consider the operating uncertainties to obtain an unbiased solution suitable for different operating conditions. In this chapter, in order to better address these uncertainties, the PLF is applied to reflect a variety of operating scenarios and then obtain an OPP solution for PMU installation. PLF is often used to handle the uncertainty caused by the outage rate of generators and the variation of load demands. In the PLF approach, the output of the conventional generator is commonly assumed to follow a Bernoulli distribution based on the generator's forced outage rate [14]. The load demand is assumed to follow a Gaussian distribution. With PLF computing, the probability distribution of power flows can be acquired. In addition, the system voltage magnitude's probability distribution is acquired, which can be applied to compute voltage magnitude estimation uncertainty in the whole system.

Different algorithms like the point estimate method (PEM), cumulant method with Gram–Charlier expansion, and first-order second-moment method are able to be employed to solve PLF analytically. These analytical methods suffer from complicated mathematical computation and low accuracy due to different approximations. In contrast to analytical methods, Monte Carlo (MC) simulation based PLF involves repetitive simulations sampling the uncertainty caused by the outage rate of generators and the variation of load demands, which leads to higher accuracy. Therefore, given the high accuracy of MC based PLF, MC is applied in this chapter to calculate the voltage magnitude estimation uncertainty [14].

2.3.3 The Proposed OPP Algorithm

Based on discussions and analyses in all previous sections of this chapter, provided that the power grid is observable via the SCADA system and supplementary PMUs are installed in the grid, the OPP problem aims to obtain minimal voltage magnitude estimation uncertainty with limited PMU numbers based on various operating scenarios considering PMU measurement reliability. Also, PLF is applied to reflect a variety of operating scenarios of generator output states and load states and then obtain an OPP solution for PMU installation. Therefore, the OPP problem is formulated as:

$$\min \sigma_V = \frac{\sum_{k=1}^{MC} \sigma_V^{(k)}}{MC} = \frac{\sum_{k=1}^{MC} \sqrt{\frac{\sum_{j=1}^{nbu} \sigma_j^2(n, num)}{nbu}}}{MC} \tag{2.19}$$

The objective of (2.19) is to obtain minimum overall system voltage magnitude estimation uncertainty σ_V, which is the average root mean square of all PMU buses' voltage magnitude estimation uncertainty among total MC trials in a placement plan. σ_V represents the average voltage magnitude estimation uncertainty in the whole system. σ_j is voltage magnitude estimation uncertainty at the bus j, which represents voltage magnitude estimation variance. As introduced in (2.12), σ_j is from two parts, including PMU measurements and voltage drops. *nbu* is the bus number in the system; *n* and *num* are PMU placement combinations serial number and PMU number respectively; $\sigma_V^{(k)}$ is the overall system voltage magnitude estimation uncertainty in *k* MC trials; *MC* is the total amount of MC trials.

Conventional optimization approaches are inapplicable for solving this OPP problem since various operating scenarios of generator output states and load states are involved. Therefore, in order to solve this OPP problem, an ad-hoc approach is proposed to obtain minimal voltage magnitude estimation uncertainty and PMU placement plans, as shown in Fig. 2-2. The proposed reliability-based probabilistic OPP method includes the following steps: first, input the test case data. Then, input the number of PMUs and enumerate all possible combinations of placement plans. Afterward, for each placement plan, 10000 trials of Monte Carlo (MC) simulation, which is considered sufficient to guarantee the convergence of the PLF, are performed to calculate the overall system voltage magnitude estimation uncertainty considering various operation scenarios. In the *k* MC trial, the system generates output states, load states, and PMU operational states are sampled, followed by calculating the power flow. Next, the estimated voltage magnitude estimation uncertainty at each bus σ_j can be calculated, which is used to calculate the overall system voltage magnitude estimation uncertainty $\sigma_V^{(k)}$. When the MC simulation is completed, the overall system voltage magnitude estimation uncertainty in 10000 scenarios can be obtained, which is used to calculate the average overall system voltage magnitude estimation uncertainty σ_V can be calculated in this PMU placement plan. In the same way, the average overall system voltage magnitude estimation uncertainty σ_V can be calculated in every PMU placement plan with the given PMU number. Next, determine the minimum overall system voltage magnitude estimation uncertainty σ_{V_min} among these PMU placement plans under the given PMU number. If σ_{V_min} meets the overall system voltage magnitude estimation uncertainty criterion σ_0, then collect the minimal overall system voltage magnitude estimation uncertainty σ_{V_min} and PMU placement plans. Otherwise, add one PMU number and repeat the steps as mentioned above until *num* equals the bus number of the input system.

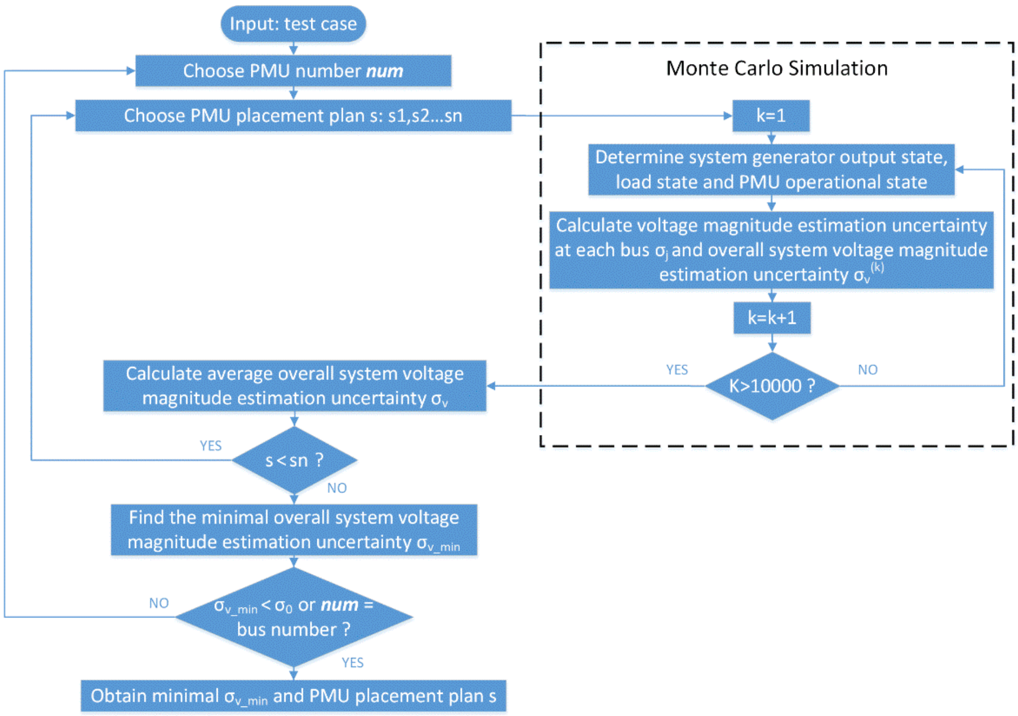


Fig. 2-2. Flowchart of the proposed reliability-based probabilistic OPP method.

2.4 Numerical Results

The proposed OPP method is tested on IEEE 9-bus as well as IEEE 14-bus systems. Three comparable experiments are conducted to investigate the optimal OPP plan with minimum voltage magnitude estimation uncertainty for each system. Firstly, to study the influence of the overall system voltage magnitude estimation uncertainty criterion on the OPP plan, in terms of various overall voltage magnitude estimation uncertainty criteria, Test 1 is conducted to find the minimal PMU number and the optimal OPP solution under this minimal PMU number whose overall system voltage magnitude estimation uncertainty is minimal. Secondly, to study the influence of PMU number on overall system voltage magnitude estimation uncertainty, Test 2 is conducted to find the optimal OPP solution with minimal overall system voltage magnitude estimation uncertainty under various numbers of PMUs. Thirdly, to study the influence of PMU measurement precision on OPP plan, in terms of various PMU measurement precision, Test 3 is conducted to find the optimal OPP solution with minimal overall system voltage magnitude estimation uncertainty under various numbers of PMUs. In these experiments, 10,000 trials MC simulation is performed with each placement solution. Also, the MATPOWER toolbox is employed to compute load flow.

2.4.1 Parameter Setting

In the PLF process, the output of the generator is set to follow a Bernoulli distribution, and each generator's forced outage rate is set as 0.03 [14]. Besides, load capacity is set to follow Gaussian distribution, which sets the mean value the same as the measured load value, and sets standard deviation as 0.5 times of its mean value [14].

For the settings of measurements, it is supposed that the measurements from both PMUs and RTUs follow Gaussian distribution, where the standard deviation is assumed as one-third of the precision of measuring equipment [14]. As for PMUs, the precision is set as: 0.1% for voltage amplitude and 0.01 rad for phase angle, respectively. As for RTU measurements, they have an inferior precision of 5% for current amplitude and power flow [14]. In particular, it is assumed that generation and/or power injections of entire buses are measured by RTUs.

The malfunction probability of each component in the substation monitoring system is also assumed to conform to Gaussian distribution, of which the mean values are given in Table 2-1, and the standard deviation is 0.1% [46].

Table 2-1 Malfunction probability of each component in the substation monitoring system.

Malfunction	Probability
Q_{pmu}	0.00450232
Q_{ct}	0.00041553
Q_{pt}	0.00145762
Q_{link}	0.001
Q_{pdc}	0.001

2.4.2 Optimal PMU Placement with Different Overall System Voltage Magnitude Estimation Uncertainty Criterion

As the economic criterion in conventional OPP, the number of PMUs to be deployed is limited by the financial budget. Likewise, the overall system voltage magnitude estimation uncertainty criterion is applied for the decision-maker to the minimal PMU number and the optimal OPP solution. In the OPP solution, minimum voltage magnitude estimation uncertainty cannot exceed this criterion.

Through the proposed OPP method, in terms of various overall system voltage magnitude estimation uncertainty criteria, the minimal PMU number and the optimal OPP solution are obtained under this minimal PMU number whose overall system voltage magnitude estimation uncertainty is minimal. The results are presented in Table 2-2. With the corresponding optimal solution, each buses' voltage magnitude estimation uncertainty under different overall system voltage magnitude estimation uncertainty criteria is shown in Fig.2-3.

Table 2-2 Optimal PMU placement plans with different overall system voltage magnitude estimation uncertainty criterion.

Overall system voltage magnitude estimation uncertainty criterion	Minimum voltage magnitude estimation uncertainty	PMU number	Placement plan (bus No.)
0.5%	0.4451%	1	9
0.4%	0.3495%	2	5,7
0.3%	0.2346%	3	5,7,9
0.2%	0.1761%	4	3,5,7,9

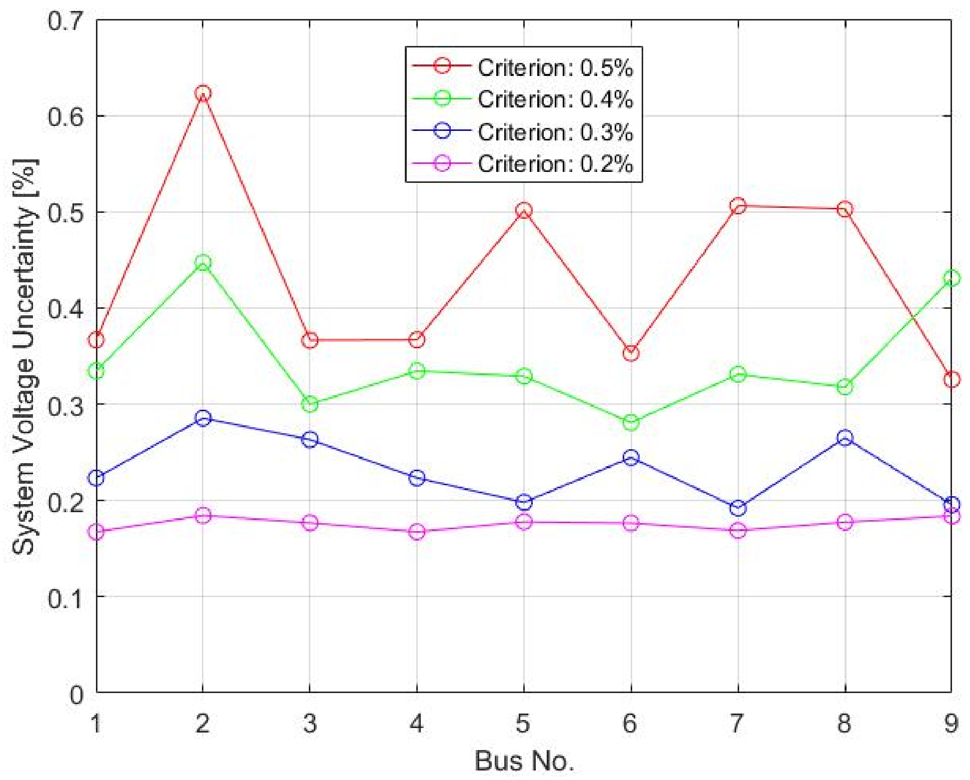


Fig. 2-3. Each bus's voltage magnitude estimation uncertainty of optimal PMU placement solution under different overall system voltage magnitude estimation uncertainty criteria.

From Table 2-2 and Fig.2-3, the stricter the overall system voltage magnitude estimation uncertainty criterion, the more PMUs are required to be placed in the system. In Fig.2-3, when only one PMU is provided, the optimal solution is placing the PMU at bus 9; thus, the voltage magnitude estimation uncertainty is minimal among all buses in the system. However, when two PMUs are provided, the optimal solution is placing the PMUs at buses 5 and 7; the voltage magnitude estimation uncertainty at bus 9 is greater than one PMU placed in the system.

2.4.3 Voltage Magnitude Estimation Uncertainty with Different PMU Number

By utilizing the proposed method, the PMU placement solutions with various PMU numbers are acquired.

From Fig. 2-4 and Table 2-3, it is suggested that the overall system voltage magnitude estimation uncertainty decreases with more PMUs installed in the system. Generally, the marginal benefit of installing additional voltage measuring devices decreases as the number of such measuring devices increases. When a few PMUs are installed in the system (like 1-4 PMU number), the overall system voltage magnitude estimation uncertainty decreases notably. When the PMU number is greater than four, the overall system voltage magnitude estimation uncertainty drops slower. Finally, when all buses in the whole network are installed PMU, the overall system voltage magnitude estimation uncertainty is close to the theoretical limit of uncertainty. However, it cannot reach the theoretical limit because each buses' voltage magnitude estimation uncertainty is not only from their own PMU measurement but from another buses' PMU, and the voltage magnitude estimation uncertainty is spread via power flow. Table 2-3 gives the detailed PMU placement plans with different PMU numbers.

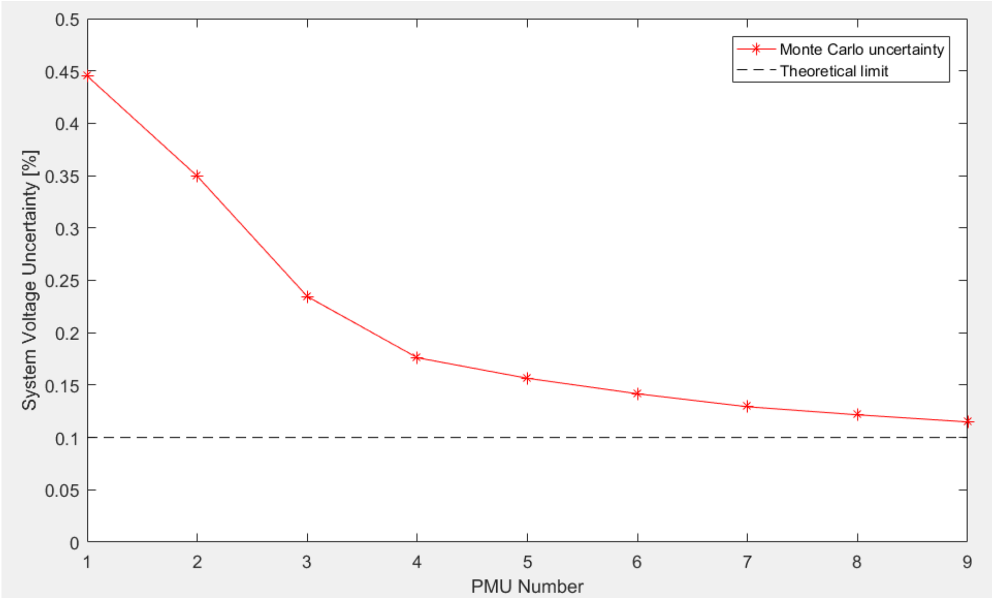


Fig. 2-4. Overall system voltage magnitude estimation uncertainty with different PMU numbers.

Table 2-3 Optimal PMU placement plans with different PMU numbers.

PMU number	Placement plans (bus No.)	Minimum voltage magnitude estimation uncertainty
1	9	0.4451%
2	5,7	0.3495%
3	5,7,9	0.2346%
4	3,5,7,9	0.1761%
5	3,4,5,7,9	0.1565%
6	2,3,4,5,7,9	0.1417%
7	2,3,4,5,6,7,9	0.1293%
8	1,2,3,4,5,6,7,9	0.1216%
9	1,2,3,4,5,6,7,8,9	0.1148%

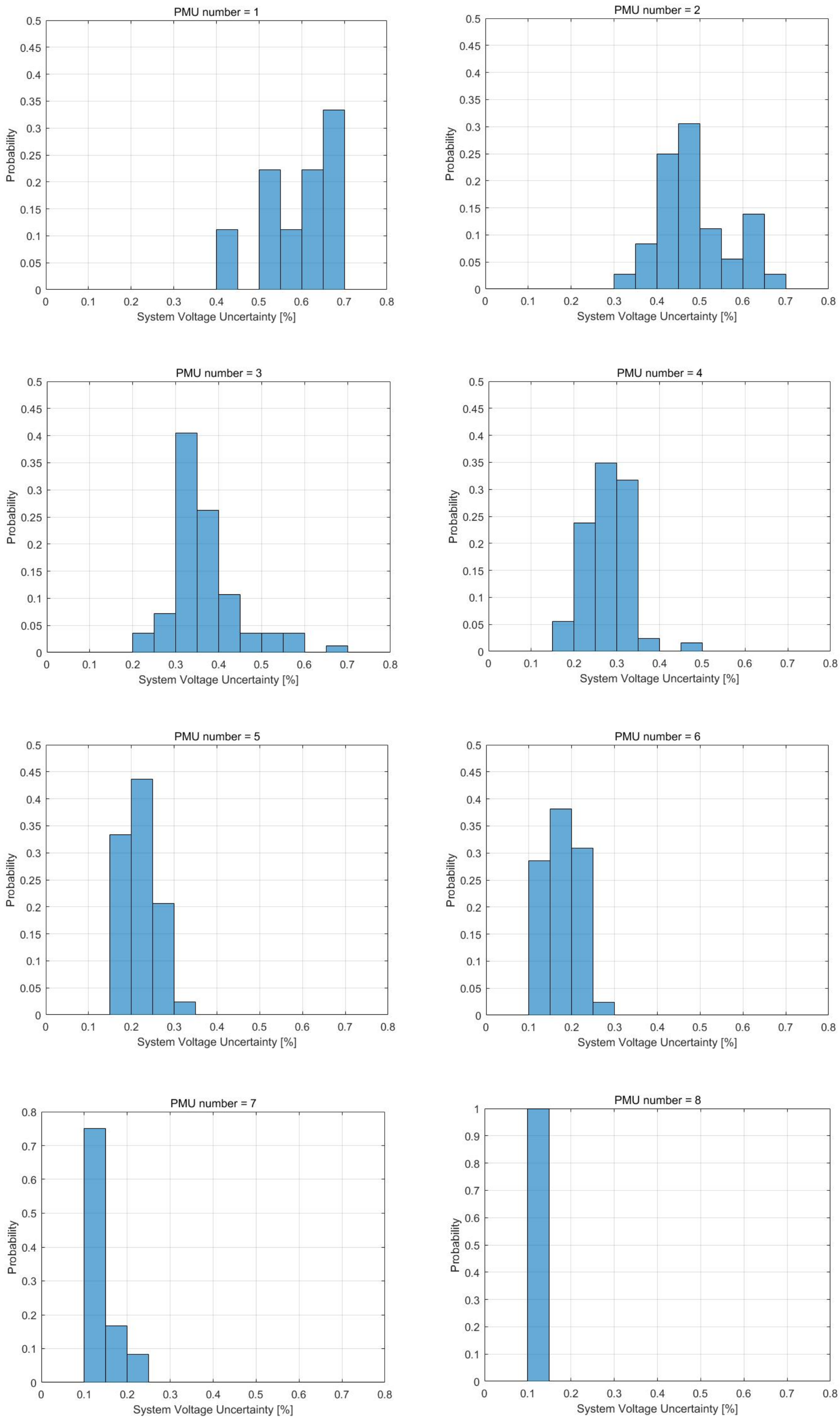


Fig. 2-5. Probability distribution of overall system voltage magnitude estimation uncertainty.

The eight graphs in Fig.2-5 show the probability distribution of overall system voltage magnitude estimation uncertainty among all possible PMU placement combinations with different PMU numbers. Not all PMU placement combinations with more PMUs cause a less overall system voltage magnitude estimation uncertainty than PMU placement combinations with fewer PMUs. For example, when one PMU is installed in the system, nearly half of PMU placement combinations whose overall system voltage magnitude estimation uncertainty is less than 0.6%, while nearly 20% of PMU placement combinations have 2 PMUs whose overall system voltage magnitude estimation uncertainty is less than 0.6%. However, on the whole, the more PMUs placed in the system, the overall system voltage magnitude estimation uncertainty is more likely to be less. The range of overall system voltage magnitude estimation uncertainty with different PMU numbers is presented in Fig. 2-6.

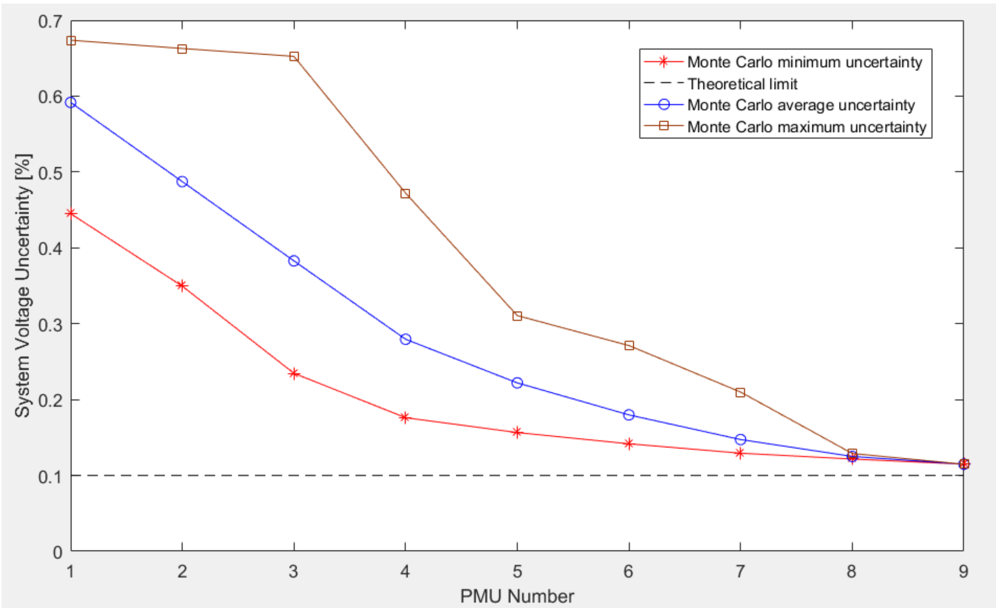


Fig. 2-6. The range of overall system voltage magnitude estimation uncertainty with different PMU numbers.

2.4.4 Voltage Magnitude Estimation Uncertainty with Different PMU Precision

By utilizing the proposed method, the PMU placement solutions with various PMU measurement precision are acquired.

When the PMU precision changes from 0.1% to 0.15%, the overall system voltage magnitude estimation uncertainty becomes greater. Under 0.15% precision, the minimal overall system voltage magnitude estimation uncertainty cannot also reach its theoretical limit (0.15%). Fig. 2-7 shows the result.

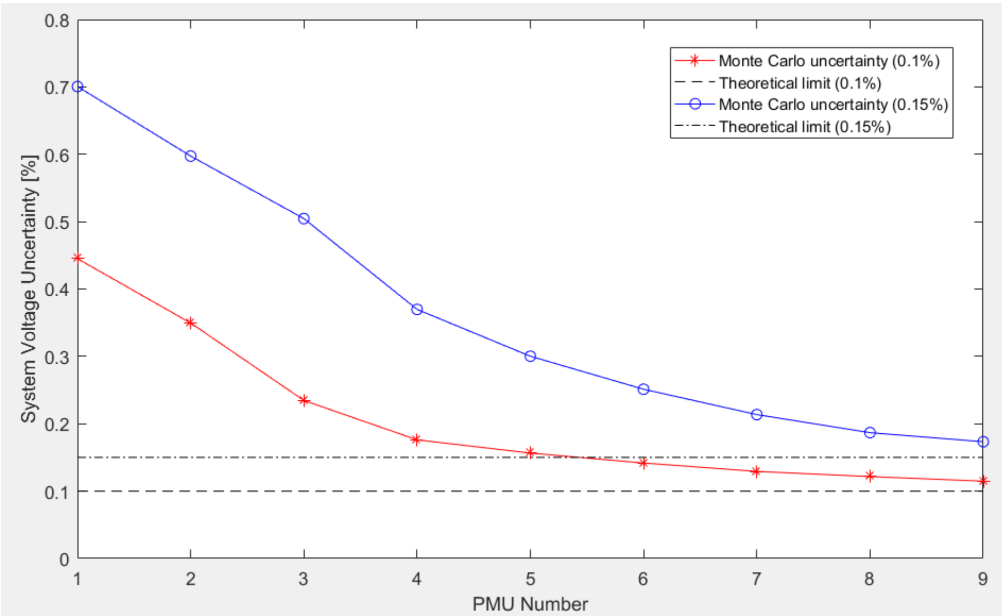


Fig. 2-7. The overall system voltage magnitude estimation uncertainty with different PMU precision.

As introduced in (2.12), the overall system voltage magnitude estimation uncertainty consists of two parts: PMU measurement devices' uncertainty and voltage drop uncertainty. If the PMU precision is different, then the proportion of the PMU measurement devices' uncertainty in overall system voltage magnitude estimation uncertainty will be changed. Thus, with different PMU precision, the optimal solution with minimal overall system voltage magnitude estimation uncertainty will be different. The results that validate this inference are in Table 2-4 and Fig. 2-8 and Fig. 2-9.

Table 2-4 Optimal PMU placement plans with different PMU numbers and different PMU precision.

	PMU precision: 0.1%		PMU precision: 0.15%	
PMU number	Placement plan (bus No.)	Minimum voltage magnitude estimation uncertainty	Placement plan (bus No.)	Minimum voltage magnitude estimation uncertainty
1	9	0.4451%	9	0.5746%
2	5,7	0.3495%	5,7	0.4662%
3	5,7,9	0.2346%	5,7,9	0.3382%
4	3,5,7,9	0.1761%	3,5,7,9	0.2620%
5	**3,4,5,7,9**	0.1565%	**2,3,4,7,9**	0.2335%
6	**2,3,4,5,7,9**	0.1417%	**3,4,5,6,7,9**	0.2125%
7	2,3,4,5,6,7,9	0.1293%	2,3,4,5,6,7,9	0.1939%
8	1,2,3,4,5,6,7,9	0.1216%	1,2,3,4,5,6,7,9	0.1824%
9	1,2,3,4,5,6,7,8,9	0.1148%	1,2,3,4,5,6,7,8,9	0.1734%

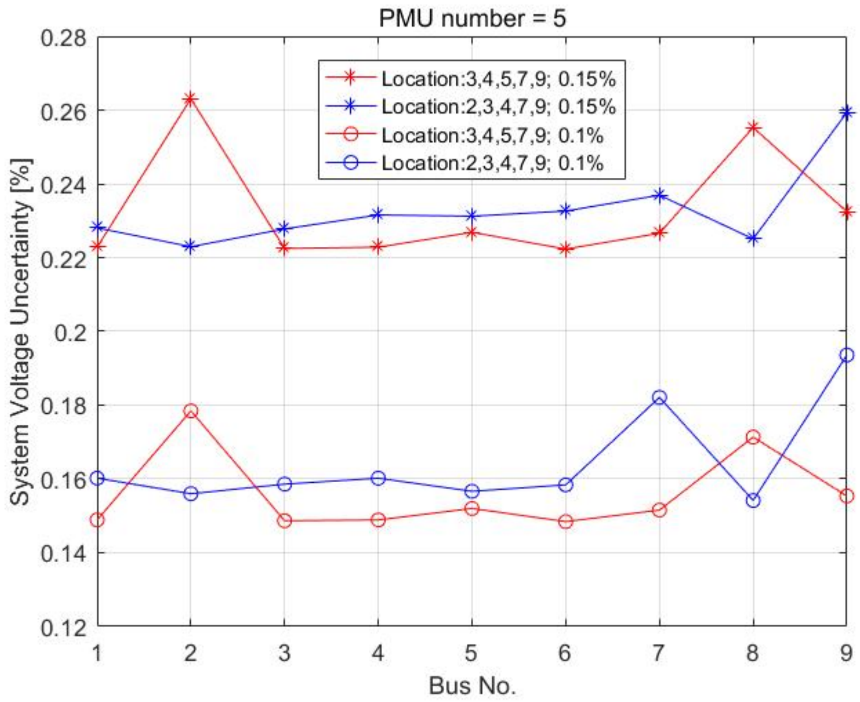


Fig. 2-8. The overall system voltage magnitude estimation uncertainty of different placement under different PMU precision.

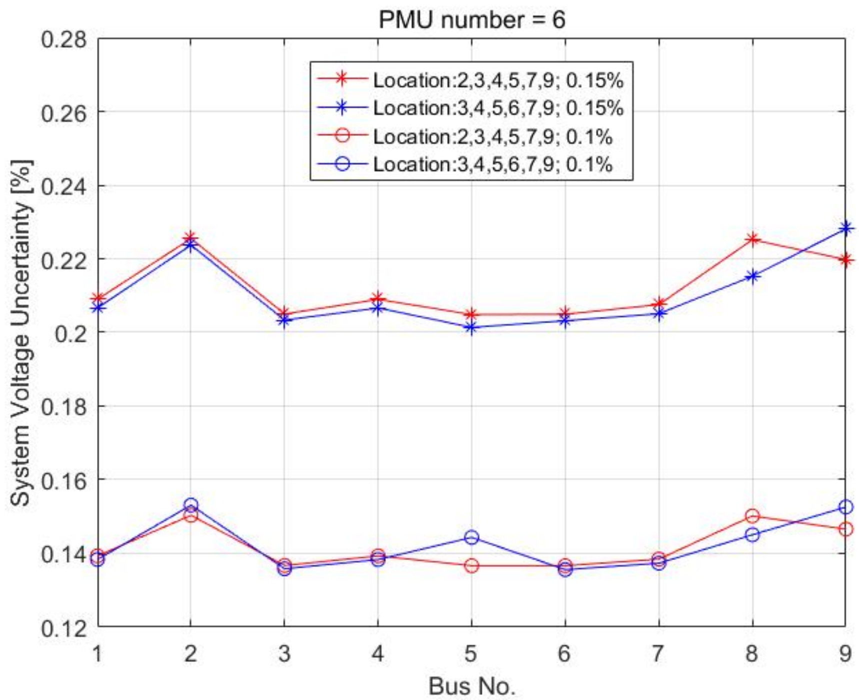


Fig. 2-9. The overall system voltage magnitude estimation uncertainty of different placement under different PMU precision.

2.4.5 Results on IEEE 14-Bus System

The proposed approach is also tested in the IEEE 14-bus system. The PMU placement solutions of different overall system voltage magnitude estimation uncertainty criteria and different PMU numbers are obtained. The results are in Table 2-5 and Fig. 2-10.

In Table 2-5, it shows that more PMUs are required to be installed in the system with the stricter overall system voltage magnitude estimation uncertainty criterion. Compared to the results on the IEEE 9-bus system in Table 2-2, more PMUs might be needed on IEEE 14-Bus System under the same overall system voltage magnitude estimation uncertainty criterion. For example, under the criterion of 0.3% and 0.2%, the minimum number of PMUs is respectively 5 and 10 while the number is 3 and 4 on IEEE 9-Bus System. From Fig. 2-10, it suggests the same trend in the results on the IEEE 9-bus system, in which the overall system voltage magnitude estimation uncertainty decreases with more PMUs installed in the system.

Table 2-5 Optimal PMU placement plans with different overall system voltage magnitude estimation uncertainty criteria and different PMU numbers.

Overall system voltage magnitude estimation uncertainty criterion	Minimum voltage magnitude estimation uncertainty	PMU number	Placement plan (bus No.)
0.5%	0.4710%	1	8
0.4%	0.3856%	2	3,10
0.3%	0.2918%	5	3,4,8,9,13
0.2%	0.1990%	10	1,2,3,4,5,9, 10,12,13,14

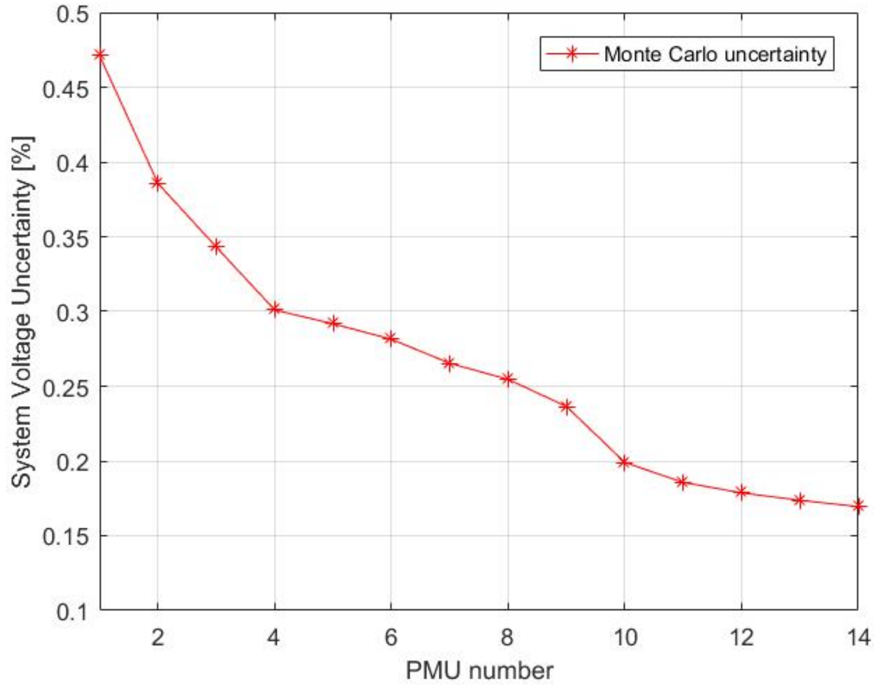


Fig. 2-10. Overall system voltage magnitude estimation uncertainty with different PMU numbers.

2.5 Summary

In this chapter, a reliability-based probabilistic OPP approach is proposed to obtain minimal voltage magnitude estimation uncertainty based on various operating scenarios, with supplementary PMUs installed in the power grid, which is observable via the SCADA system. PMU measurement system reliability is modeled when estimating the system states. In the modeling of PMU measurement reliability, PMU measurement system components' random outages are considered. These random outages may lead to the partial unobservability of the network and will endanger the power system's safety. Moreover, MC-based PLF is applied to describe a variety of operating scenarios. In this way, the load patterns and power generations are considered stochastically as the operating uncertainties so that the obtained PMU placement solution is unbiased. Finally, the proposed OPP method is tested on IEEE 9-bus as well as IEEE 14-bus systems.

Chapter 3 A Novel Approach for Transmission System State Estimation Based on PMUs

3.1 Introduction

As introduced in chapter 2, SE plays a vital role in contemporary energy management systems (EMSs). While the system observability is the prerequisite to traditional SE methods, the network is, however, not always fully observable due to, e.g., malfunction of measurement devices, miss of measurement data, or interference by malicious data attacks [17]. Hence, this chapter focuses on proposing a novel approach for transmission system state estimation, which is effective even in an unobservable(as by the classical SE methods) network considering the corrupted or bad measurements and missing measurements under contingencies.

As discussed in chapter 2, phasor measurement units (PMUs) shows great advantages over conventional supervisory control and data acquisition (SCADA) measurements on the following aspects: (i) Synchronization. Each PMU measurement is time-stamped and synchronized from the global positioning satellite system (GPS) [11]. (ii) Higher measurement accuracy. This is because network buses' voltage phasor can be measured direly, and a reference bus with a fixed voltage phase angle is not needed to choose anymore. (iii) Higher sampling rates (up to 60 samples/s), which can capture fast system dynamics while bringing about the huger amount of data as compared with the SCADA system (typically around 1 sample/ 5 s) [64]. Owing to these merits, the deployment of PMUs makes it possible for real-time monitoring of the smart power grid [65].

To overcome the unobservability issue and handle corrupted measurements as well as missing measurements while providing an accurate SE, in this chapter, a novel SE approach using a conditional generative adversarial network (GAN) is proposed. GAN is one of the most promising generative networks under a deep learning framework and has attracted great interest in recent years, especially in computer vision research due to its excellent capability in generating realistic images [85] given a collection of indistinct or incomplete images. It has several merits: (i) speed of processing. Once the model is trained well, it can give the output immediately; (ii) do not need any appropriate knowledge of the system model. This method is model-free and data-driven; (iii) fault tolerant. The output is not likely to have a large error even with fault input; (iv) fast and robust. It retains good learning ability in the context of bad or missing data [86]. This inspires this thesis to apply GAN in the SE process, where the raw system measurements can be regarded as corrupted images, and the desired system states correspond to the real images that can be directly generated through a fine-tuned GAN. Compared to the classical GAN, the conditional GAN (CGAN) is applied to appropriately adapt to the SE problem. The proposed method uses the Wasserstein distance rather than the Jensen–Shannon divergence proposed in [87], which can significantly improve training performance and obtain more accurate SE results. In this chapter, by applying conditional generative adversarial networks, the actual correlations of system states can be well captured, and the system states can be accurately estimated without prior knowledge of the system model. The deep learning based SE method with PMUs can effectively restore all system states considering the corrupted raw measurements or missing measurements under contingencies. The influence of data contaminations is fully investigated with respect to different data contamination ratios and types.

Considering the above discussed, the purely PMU-based SE method has shown various benefits compared with purely conventional SCADA based or PMU-SCADA-based SE methods, including linear measurement function, time-stamped measurements, high sampling rates, and low latency. In the future, the periodical magnitude-based measurement in the SCADA system will be gradually replaced by real-time PMU-based measurements. To develop an accurate SE method for the future smart grid, unlike the power grid is observable via the SCADA system in chapter 2, in this chapter, all measurements are assumed to be provided by PMUs.

This chapter presents a comprehensive study of transmission system state estimation, and the proposed method can handle different types and ratios of contamination measurements and effectively restore all system states even in an unobservable network. The case studies are based on two large case systems, including the IEEE 118- bus system and the 2746-bus Polish system.

3.2 Problem Formulation

This section firstly details power system SE models based on PMUs. Then, the proposed SE framework based on deep learning is given.

3.2.1 Power System State Estimation Models based on PMUs

In the system with M PMUs installed, it is assumed the whole system can be observable with the maximum redundancy by deploying all these PMUs for all system buses. This corresponds to saying that the voltage phasors

$y_j = V_j \angle \theta_j, j = 1,\dots,N$, of all buses in the system, can be measured to form a raw measurement vector **y**, which is denoted as $[y_1,\dots y_j \dots, y_N]^T$. In practice, there is a wide range of factors that might lead to corrupted measurements, such as impulsive communication noise, the failures of instruments, cyber-attacks, etc. Missing measurement is another common situation faced by system operators. As discussed in chapter 2, the unavailability of getting voltage or current measurements from potential transformers (PTs) or current transformers (CTs), failure of phasor data concentrator (PDC), and failure of a local communication system [33] may all lead to the measurement loss of different severity. Especially, the system is more likely to be unobservable without enough redundancy when the aforementioned situations occur.

As discussed in section 2.2.1, the presented BCSE model is based on PMU and SCADA measurements. The most commonly used SCADA measurements are the line power flows and bus power injections, in which the measurement function between the state variables and the measurements is non-linear [16]. While for PMUs, the measured quantities are voltage phasors, leading to a linear relationship with the state variables. The system states thus can be estimated via a linear measurement model. In the BCSE model, the state variables consist of the reference bus's phasors and currents' real part and imaginary part. While for the linear SE model based on PMUs, in an *N*-bus power system, the vector of state variables $\mathbf{x}$ is denoted by $[V_1 \dots, V_j \dots, V_N, \theta_1 \dots, \theta_j \dots, \theta_N]^T$, $j = 1,\dots,N$. V_j and θ_j are the voltage magnitude and phase angle of *j*-th bus, respectively. As the general SE formulation presented in section 2.2.1, the measurement function of the linear SE model based on PMUs is expressed in a similar form:

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \tag{3.1}$$

where $\mathbf{y}$ represents the measurement vector obtained by PMUs; $\mathbf{h}\ (\cdot)$ is the linear vector-valued measurement function established based on the state vector $\mathbf{x}$; $\mathbf{e}$ is the measurement error vector that is usually assumed to be white noise composed by zero mean with a covariance matrix $\mathbf{R}$.

As discussed in section 2.2.1, the SE formulation is presented in (2.3). As discussed before, the classical SE involves complicated and separated steps to deal with network topology, observability, SE calculation, and bad data.

Other models are based on data-driven methods. Such as auto-associative neural networks (NNs) or autoencoders and multilayer perceptron (MLP) based NNs are applied for SE. Once trained offline using historical data and/or simulated samples, NNs can be implemented for real-time SE. The accuracy of SE was not outstanding, especially for large-scale systems restricted by the development of AI technology during that period.

3.2.2 The Proposed SE Framework based on Deep Learning

Due to the poor temporal resolution nature of the pseudo-measurements in classical SE methods and the limitation by the development of AI technology during that period in data-driven methods based on NNs, the accuracy of SE cannot be guaranteed with corrupted measurements as well as missing measurements, especially for large-scale systems. Therefore, the proposed SE model is developed and includes three parts, i.e., the input, data-driven SE module based on deep learning technique, and the output. For the input, in this chapter, it is assumed that the system is observed merely by PMUs, i.e., all the obtained measurements are voltage phasors with a typical sampling rate at 0.02 second/sample. The proposed data-driven SE module is based on a deep learning framework, which includes conditional inputs and performs more outstanding than previous data-driven methods on handling contamination data (including corrupted or bad data and missing data) and will be discussed in detail in Section 3.3. The output is the estimated system states (voltage magnitudes and angles), whose updating frequency is in line with the PMUs and faster than the classical SCADA based SE. The prosed SE model provides an integral framework, which stays away from the complicated and separated steps in the classical SE methods.

3.3 The Proposed Method

In this section, to solve SE problems in 3.2, the proposed fully data-driven CGAN-SE with Wasserstein GAN is presented. The basic theory of GAN [87] will be reviewed first. Then, the section explains how the framework of Wasserstein GAN fits into the SE problem. Later, the model establishment process for SE by integrating the synthetic PMU measurement will be discussed.

3.3.1 Wasserstein GAN

As defined before, $\tilde{\mathbf{x}} = [\tilde{x}_1 \dots \tilde{x}_N]^T$ represents the true system state, $\{\tilde{x}_j^{(i)}\}_{i=1}^m$ denotes the *i*-th sample of the true system state, and *m* is the number of samples. Let $p_{data}(x)$ denote the distribution of the true system state. Suppose a group of noise inputs z follow a known distribution $z \sim p_z(z)$, e.g., uniform distribution or joint Gaussian. The goal of the method is to transform the sample z from the distribution $p_z(z)$ so that it can follow true system state distribution $p_{data}(x)$. To this end, two deep neural networks are trained simultaneously. One is the generator network G expressed as $G(z;\theta^{(G)})$, which is parametrized by $\theta^{(G)}$; the other is the discriminator network D written as $D(x;\theta^{(D)})$, whose function is parametrized by $\theta^{(D)}$. The generator and discriminator are combined to form the GAN network.

Generator: When training the generator, a large number of up-sampling operations are implemented to the inputs z, and the generator outputs are the estimated system states. The training procedure can be expressed as the following mapping:

$$G(z;\theta^{(G)}): z \to p_G(z) \tag{3.2}$$

where $p_G(z)$ is the generated distribution, which provides samples to the estimated system state. $p_G(z)$ also follows the true system state distribution $p_{data}(x)$.

Discriminator: The discriminator should be trained with the generator at the same time. Both samples from the generated distribution $p_G(z)$ and the true system state distribution $p_{data}(x)$ are served as discriminator inputs. After plenty of down-sampling operations, the output is a value p_{real} which is continuous and reflects what extent these inputs belong to the true system state distribution $p_{data}(x)$. Likewise, the training process of the discriminator can be expressed as a mapping:

$$D(x;\theta^{(D)}): x \to p_{real} \tag{3.3}$$

where x is the input vector that can be sampled either from $p_G(z)$ or $p_{data}(x)$. The discriminator is expected to learn to distinguish between $p_G(z)$ and $p_{data}(x)$, and to maximize the difference between these two distributions.

In the training stage, D is trained to maximize its capacity of discernment between true system state distribution and estimated state distribution from the generator. G and D are trained simultaneously to minimize the difference between these two distributions. The weights of G and D are updated to minimize generator loss function C_G and discriminator loss function C_D, respectively. Specifically, a batch of samples collected from distribution $p_z(z)$ is fed into G. Meanwhile, a batch of true samples drawn from distribution $p_{data}(x)$ are fed into D. A small C_G indicates that the generated samples are more realistic from the discriminator's view. That is, with respect to the application of GAN for SE in this thesis, the generated system states are more similar to the true system states. On the other hand, a small C_D reflects the D does well in distinguishing the discrepancy between the generated system states and the true system states. Also, it indicates that there is a large difference between generated state distribution $p_G(z)$ and true state distribution $p_{data}(x)$. Generator loss function C_G and discriminator loss function C_D can be expressed as [89]:

$$C_G = \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{3.4}$$

$$C_D = -\mathbb{E}_{x \sim p_{data}(x)}[\log(D(x))] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{3.5}$$

For a given D, as a large output value of discriminator p_{real} shows the generated samples are more realistic, the generator should seek to minimize

$\log(1-D(G(z)))$ by altering G to generate more realistic samples, which gives the loss function of a generator in (3.4). For a given G, as seen in (3.5), the discriminator attempts to minimize $-\log(D(x))$, thus giving a large discriminator output value p_{real}. In the meanwhile, the network minimizes $-\log(1-D(G(z)))$, which is virtually a reverse of C_G. Note that, Eq. (3.4) equals to minimize $-\log(D(G(z)))$. Hence, the GAN can be formulated by combining these two loss functions as a two-player minimax game with the value function $V(G,D)$:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{3.6}$$

where $V(G,D)$ is the negative of C_D.

At the beginning of training, the performance of generator G is poor, and the system state samples generated by G are very different from samples of $p_{data}(x)$. Consequently, the discriminator outputs a small value of p_{real} and rejects these 'fake' samples with high confidence. Under these circumstances, C_G is small, C_D is large and $V(G,D)$ is also large. As the training goes on, the generator learns to produce more realistic samples and the discriminator learns to distinguish these samples from two different distributions. Finally, G defeats D, i.e., the samples generated by G are almost as real as true samples, also, D fails to distinguish samples from $p_G(z)$ and $p_{data}(x)$.

According to the Kantorovich-Rubinstein duality [90], The Wasserstein distance (Earth-Mover distance) is the dual of the minimax objective in (3.6). x and y are two random variables and $\prod(\mathbb{P}_r, \mathbb{P}_g)$ is the set of all joint distributions $\gamma(x,y)$, whose marginals are $\mathbb{P}_r$ and $\mathbb{P}_g$, respectively. Then the Wasserstein distance between x and y is defined as:

$$W(\mathbb{P}_r,\mathbb{P}_g) = \inf_{\gamma\in\prod(\mathbb{P}_r,\mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|] \tag{3.7}$$

The Wasserstein distance can be viewed with the "cost" of the optimal plan that moves all the "mass" $\prod(\mathbb{P}_r,\mathbb{P}_g)$ from location x to location y in order to transform the distribution $\mathbb{P}_r$ into the distribution $\mathbb{P}_g$. $\gamma(x,y)$ can be described as the quantity of the moved "mass" at one time.

The objective of GAN is to make the generated sample distribution $p_z(D(G(z)))$ close to the true system states distribution $p_{data}(D(x))$. Thus, the Wasserstein distance between the true system state and the generated sample can be expressed as:

$$W(D(x),D(G(z)) = \sup_D \mathbb{E}_{x\sim p_{data}(x)}[D(x)] - \mathbb{E}_{z\sim p_z(z)}[D(G(z))] \tag{3.8}$$

When the Wasserstein distance converges, the optimal plan of moving "mass" is found, and the optimal generator G^* is also found. As reported in the literature [89], the JS divergence applied in the original GAN cannot reflect the extent to which two distributions $p_G(z)$ and $p_{data}(x)$ are close when they are very different from each other, which makes GAN sensitive to the parameters. As a result, the generated system states almost follow the pattern with the highest occurring probability regardless of the inputs. However, applying the Wasserstein distance as the loss function of GAN [91] successfully addresses these limits and gives the accurate distance between two distributions. Therefore, the generator of WGAN can mimic the true system operating scenarios to generate diversified system states rather than the same ones produced by the original GAN.

3.3.2 Conditional GAN SE

The classical GAN model uses merely the noise vector as input and has no extra limitations for the generated output. The classical GAN can be extended to a conditional counterpart where both the generator and discriminator are conditioned on some extra information [91]. In conditional GAN (CGAN), the generated samples should satisfy this condition y.

This architecture is more fit to SE problems, where the raw system measurement can be regarded as a condition y, and the generated system states should be guaranteed to be as close as possible to the true system states while satisfying the corresponding raw measurement. The CGAN is implemented by feeding y into both the generator and discriminator as additional inputs. Eventually, Eq. (3.8) can be rewritten as:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[D(x|y)] - \mathbb{E}_{z\sim p_z(z)}[D(G(z|y))] \quad (3.9)$$

Fig. 3-1 illustrates the CGAN architecture for SE, and the algorithm used in the proposed method is described in Algorithm 3-1.

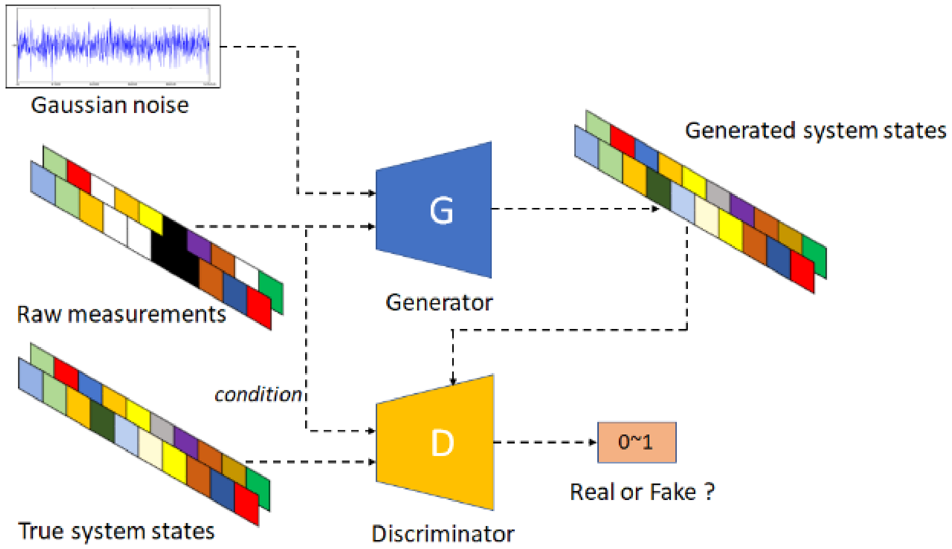


Fig. 3-1. The architecture of CGAN, including the input and output of the generator and discriminator, respectively.

In algorithm 3-1, $G(z)$ and $D(x)$ are neural networks with parameter $\theta^{(G)}$ and $\theta^{(D)}$, respectively. Both networks consist of multilayer perceptron (MLP), convolution, normalization, max-pooling and Rectified Linear Units (ReLU). The parameters are tuned within several training batches. The training algorithms for discriminator and generator are slightly different, where the former is based on gradient ascend and the latter is gradient descend. Besides, the Root Mean Square Propagation (*RMSProp*) algorithm is applied in both generator and discriminator to allow the learning rate to be self-adjustable. *RMSProp* is a method in which the learning rate is adapted for each of the parameters. The idea is to divide the learning rate for a weight by calculating the average of recent gradients magnitudes [92]. It should be noted that weight clipping is applied in discriminator training to meet specific conditions and avoid gradient explosion [89]. The model setting for CGAN will be presented in Section 3.4.1.

Algorithm 3-1 CGAN with Wasserstein Distance for SE

Require: α , the learning rate; c , the clipping parameter; m , the batch size; k_{dis} , the number of iterations of the discriminator per generator iteration.

Require: $\theta_0^{(D)}$, initial discriminator's parameters; $\theta_0^{(G)}$, initial generator's parameters.

while $\theta_0^{(G)}$ has not converged **do**

for $t = 0, \ldots, k_{dis}$ **do**

• Sample batch of m noise sample $\{(z^{(i)}, y^{(i)})\}_{i=1}^{m}$ from noise prior distribution $p_z(z)$.

• Sample batch of m examples $\{(x^{(i)}, y^{(i)})\}_{i=1}^{m}$ from the true system state data $p_{data}(x)$

• Update the discriminator by ascending its gradient:

$$g_{\theta^{(D)}} \leftarrow \nabla_{\theta^{(D)}} \frac{1}{m} \sum_{i=1}^{m} [D(x^{(i)} \big| y^{(i)}) - D(G(z^{(i)} \big| y^{(i)}))]$$

$$\theta^{(D)} \leftarrow \theta^{(D)} + \alpha \cdot RMSProp(\theta^{(D)}, g_{\theta^{(D)}})$$

$$\theta^{(D)} \leftarrow clip(\theta^{(D)}, -c, c)$$

end for

• Sample batch of m noise samples $\{(z^{(i)}, y^{(i)})\}_{i=1}^{m}$ from noise prior distribution $p_z(z)$.

• Update the generator by descending its gradient:

$$g_{\theta^{(G)}} \leftarrow -\nabla_{\theta^{(G)}} \frac{1}{m} \sum_{i=1}^{m} D(G(z^{(i)} \big| y^{(i)}))$$

$$\theta^{(G)} \leftarrow \theta^{(G)} - \alpha \cdot RMSProp(\theta^{(G)}, g_{\theta^{(G)}})$$

end while

3.4 Numerical Results

To validate the effectiveness of the proposed SE method, the experiment is carried out on two power systems, namely the IEEE-118 bus system and the 2746-bus Polish network, respectively. The system data are simulated from MATLAB and MATPOWER toolbox by implementing the Monte Carlo power flow calculations [93]. To obtain distinct system states, it is assumed the system load satisfies the Gaussian distribution with zero mean and standard deviation of 0.1. Load samples are then drawn and fed into power flow computations to derive true system states. For the IEEE-118 bus system and a

2746-bus Polish network, there are 15,000 training examples, respectively. 80% of these samples are used for training, and the remaining 20% are used for testing. The batch size is 32, and the number of epochs is 300. Therefore, the training of the networks took 112,500 iterations (15000 * 80% / 32 * 300) for each scenario. All the programs for the conditional GAN-based SE (CGAN-SE) model are implemented using 'TensorFlow' [94] in Python on PyCharm IDE with NVIDIA GeForce RTX 2080 Ti GPU and 11GB RAM is 11GB.

3.4.1 Model Architecture and Training Details

Table 3-1 The proposed CGAN based SE model structure for the IEEE 118-bus system.

	Generator *G*	Discriminator *D*
Input	100	2 * 118
Layer 1	MLP, 1024	Conv, 64
Layer 2	MLP, 512	Conv, 256
Layer 3	Conv_transpose, 512	Conv, 512
Layer 4	Conv_transpose, 256	MLP, 1024
Layer 5	Conv_transpose, 64	

The generator *G* consists of 2 fully connected multilayer perceptron (MLP) and 3 de-convolutional layers. The first 2 MLPs are used for up-sampling, and the de-convolutional layers kernel size is 1×5 and strides size 2 are used to up-sample the input noise z. While the discriminator *D* has a reversed architecture, whose 3 convolutional layers are all with a kernel size of 1×5 and stride size of 2. Table 3-1 lists the detailed settings of the proposed GAN model on the IEEE 118-bus system.

The input measurement data are normalized to [-1, 1] in order to match with the output range of the tanh activation function in the last layer of the generator G. The models are trained by the *RMSProp* optimizer. Random initializations of neuron weights follow a normal distribution with zero mean and standard deviation of 0.02. Except for the input layer, the batch normalization is employed before each layer to stabilize the inputs to nonlinear activation functions. To be specific, it normalizes each layer's inputs by using zero mean and unit variance. Leaky-ReLU activation function is used in the discriminator, and the ReLU activation function is used in the generator, excluding the output layer. In this chapter, to achieve reliable performance, the discriminator D is trained for four times, and the generator G is trained once [85]. Thus, k_{dis} is set as 4 in Algorithm 3-1.

3.4.2 Data Generation

As mentioned in the preceding sections, the true system states $\tilde{\mathbf{x}}$ are generated via the Monte Carlo probabilistic power flow calculations with different load scenarios. Then, the raw measurements are created by adding a Gaussian noise **e** with zero mean and standard deviation of 0.001 [112] (PMU's precision) to $\tilde{\mathbf{x}}$. The next section will discuss the creation of abnormal measurements considering 3 contamination scenarios, in each of which different contamination ratios r% ranging from 0% to 100% will be considered.

Corrupted measurement data refers to the measurements that significantly differ from the normal measurement data due to various reasons such as instrument failures, and impulsive communication noise, etc. Corrupted voltage magnitude and phase angle measurements are generated by randomly choosing r% raw measurements and adding an error with 0.5 mean and 0.05

standard deviation [64]. The rest (1 - r%) are still raw measurements with only typical PMU measurement noise.

Missing measurement data are generated by randomly choosing r% raw measurement and setting their voltage magnitude and voltage phase angle with zeros.

Mixed measurement data is contaminated with a mixture of bad data and missing data. To fabricate this situation, the contaminated data are generated equally for each type. i.e., the mixed contamination data accounting for r % of the dataset contains (r% / 2) corrupted data and (r% / 2) missing data.

3.4.3 Performance Evaluation

The performance of SE is evaluated by the mean average error (MAE) for the total SE error [79]:

$$\text{MAE} = \frac{1}{N * m} \sum_{j=1}^{N} \sum_{i=1}^{m} \left| \hat{x}_j^{(i)} - \tilde{x}_j^{(i)} \right| \tag{3.10}$$

where N is the number of buses, m is the number of samples, $\hat{x}_j^{(i)}$ and $\tilde{x}_j^{(i)}$ refer to estimated system states and true system states corresponding to the i-th sample and the j-th bus, respectively.

A. Overall SE accuracy under different contamination ratios and types

Different contamination ratios r% ranging from 10% to 90% with 10% increment are examined in this section to give the full-scale analysis of the model accuracy and robustness. The case with raw measurements (r% = 0%) is also tested. Additionally, to ensure the experiment is unbiased, the verified model is simulated 10 times for each result. The MAE of the proposed method for SE of both voltage magnitude and phase angle under each contamination scenario are illustrated in Fig. 3-2 (IEEE-118 bus system) and Fig. 3-3 (2746-

bus Polish network). Their average SE results of ten runs are represented by the blue/orange bars, respectively.

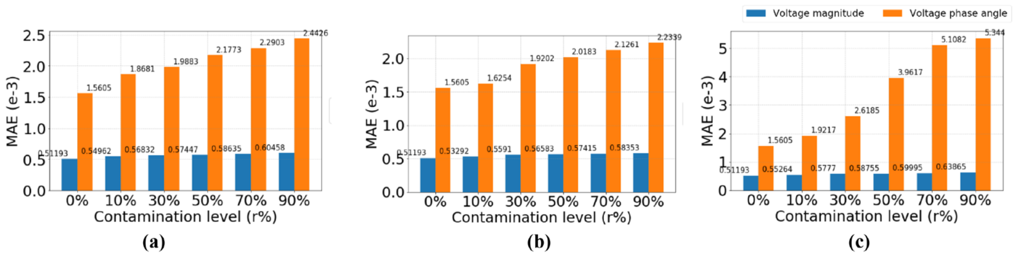


Fig. 3-2. MAE (e-3) of CGAN-SE on the IEEE 118-bus system over ten runs with respect to various contamination ratios under three measurement contamination scenarios.

(a) Corrupted Measurement, (b) Missing Measurement, (c) Mixed contamination Measurement

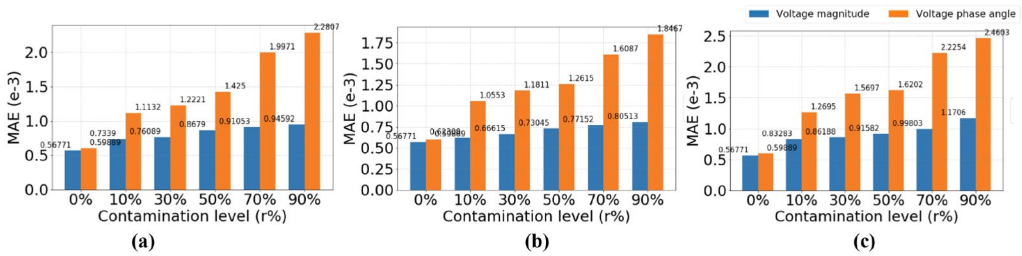


Fig. 3-3. MAE (e-3) of CGAN-SE on the 2746-bus Polish system over ten runs with respect to various contamination ratios under three measurement contamination scenarios.

(a) Corrupted Measurement, (b) Missing Measurement, (c) Mixed contamination Measurement

From the results shown in Fig. 3-2, the MAE of voltage angle is higher than the voltage magnitude under all three contamination scenarios. Especially in Fig. 3-2 (c), with mixed contamination measurement, the MAE of phase angle is significantly greater than the voltage magnitude. In the scenario with only raw measurement involved (r%=0%), the MAE of voltage magnitude and phase angle are 5.1193e-4 and 1.5605e-3, respectively. With the increase of contamination ratio from 0% to 90%, the MAE of both voltage magnitude and phase angle grows accordingly. Additionally, the MAE of phase angle has an obvious rising trend while that of voltage magnitude ascends slowly. The MAE under mixed contamination situations is greater than that of the other two cases. The MAE with the missing measurement is slightly smaller as compared to that with the corrupted measurement. The tendency of model performance for the 2746-bus Polish system under all verified contamination ratios is similar to that of the 118-bus system, as observed in Fig. 3-3. Yet, the MAE of voltage magnitude is larger than that in the 118-bus system, whereas the error of phase angle is smaller.

B. Estimated Distribution Assessment

To investigate the similarity of the distribution between the generated system states and true system states, two load buses are randomly chosen to compare their probability density distribution profiles of voltage magnitude. Fig. 3-4 and Fig. 3-5 depict the probability density histograms of generated and true system states, respectively at bus 30 (118-bus system) under 30% contamination ratio. Also, Fig. 3-6 and Fig. 3-7 depict the probability density histograms of generated and true system states, respectively at bus 245 (2746-bus system) under 30% contamination ratio. The probability in these figures is represented by the individual rectangle areas multiplied by the width of the interval, and the Y-axis value and the cumulative rectangle areas are equal to one. The distribution of voltage magnitude is closer to its true distribution than that of phase angle under three contamination scenarios in both tested systems. On the other hand, CGAN-SE is more effective in handling either missing or corrupted measurement than mixed contamination measurement, as the discrepancy of distribution profiles for the former two scenarios is less evident than that for the mixed contamination measurement case.

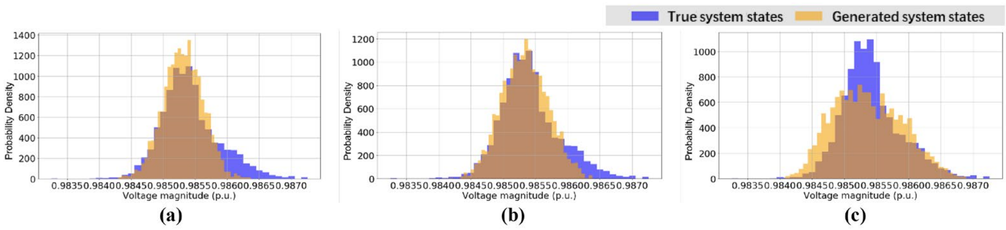


Fig. 3-4. Probability density histograms of the generated system states and true system states with different contamination measurements (30% contamination ratio) at bus 30 of the IEEE 118-bus system (voltage magnitude). (a), (b), (c) Probability density distribution of voltage magnitude with 30% ratio of corrupted measurement, missing measurement, and mixed contamination measurement, respectively.

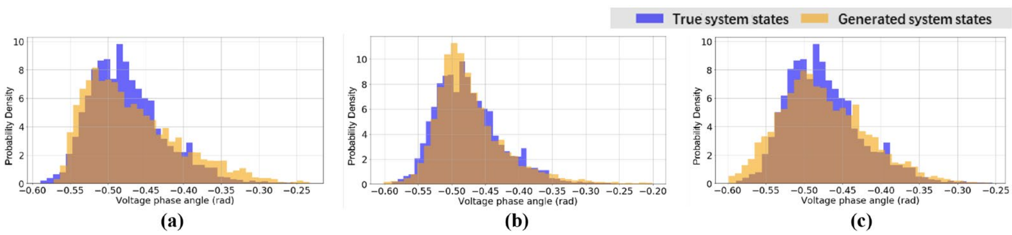
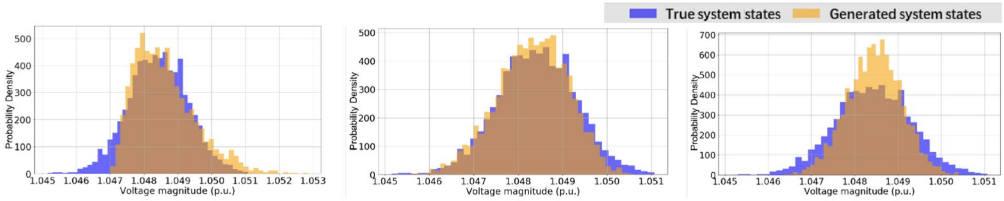


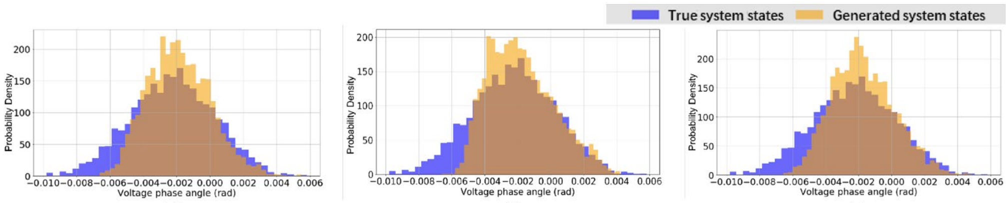
Fig. 3-5. Probability density histograms of the generated system states and true system states with different contamination measurements (30% contamination ratio) at bus 30 of the IEEE 118-bus system (voltage phase angle).

(a), (b), (c) Probability density distribution of phase angle with 30% ratio of corrupted measurement, missing measurement, and mixed contamination measurement, respectively.



(a) (b) (c)

Fig. 3-6. Probability density histograms of the generated system states and true system states with different contamination measurements (30% contamination ratio) at bus 245 of 2746-bus Polish system (voltage magnitude). (a), (b), (c) Probability density distribution of voltage magnitude with 30% ratio of corrupted measurement, missing measurement, and mixed contamination measurement, respectively.



(a) (b) (c)

Fig. 3-7. Probability density histograms of the generated system states and true system states with different contamination measurements (30% contamination ratio) at bus 245 of 2746-bus Polish system (voltage phase angle).

(a), (b), (c) Probability density distribution of phase angle with 30% ratio of corrupted measurement, missing measurement, and mixed contamination measurement, respectively.

To quantify the similarity between the generated system states and true system states, the Wasserstein distance between two distributions is calculated, and the results are shown in Table 3-2. The Wasserstein distance has been introduced in Section 3.3.1 as a natural way to compare two probability distributions, and the smaller value means the two distributions are similar. As seen from Table 3-2, the smallest Wasserstein distance is observed in the case of voltage magnitude on the IEEE 118-bus system with corrupted measurement, and the largest distance occurs in the estimated voltage phase angle on the IEEE 118-bus system with corrupted measurement. Besides, the distance of voltage magnitude is larger than that of voltage phase angle for the same system.

Table 3-2 The Wasserstein distance with different contamination measurements (30% contamination ratio) on the IEEE 118-bus system and the 2746-bus Polish system.

	Contamination Type (30% ratio)		
	Corrupted	Missing	Mixed
IEEE 118-bus system (voltage magnitude)	1.4903	1.5808	2.0910
IEEE 118-bus system (voltage phase angle)	6.1195	6.8061	8.7550
2746-bus Polish system (voltage magnitude)	2.0592	1.9056	2.4831
2746-bus Polish system (voltage phase angle)	3.9170	4.1434	5.4669

C. Spatial Correlation Assessment

To further validate the quality of generated system states, the correlation of buses is studied for voltage magnitude and phase angle, respectively. To exhibit the evolving process of correlation during training, the Pearson correlation coefficient matrix is computed at several training iterations, i.e., 200, 2000, and 20000, under 30% mixed contamination scenario. For the voltage magnitude of the 118-bus system, as shown in Fig. 3-8(a), both generated voltage magnitude and true voltage magnitude show weak spatial correlations as the correlation coefficients tend to be zeros. Thus, CGAN-SE can yield correlations that are almost similar to the true system state. This is also confirmed in Fig. 3-8 (c) on the 2746-bus system. By contrast, the spatial correlation of the phase angle between buses is stronger than that of voltage magnitude. At the beginning of the training, though the correlation profile of phase angles is far from the true ones, with the learning carried on, it can learn the spatial interdependency and finally gives a better result. Additionally, the dark cross line around bus 70 in both generated phase angles and true phase angles means this bus has no spatial correlation with all of the other buses, which represents the phase angle of this bus is a constant. This conforms to the system model that bus 69 is a slack bus. Therefore, CGAN-SE has the capability of learning the spatial correlation of voltage magnitude and phase angle between buses.

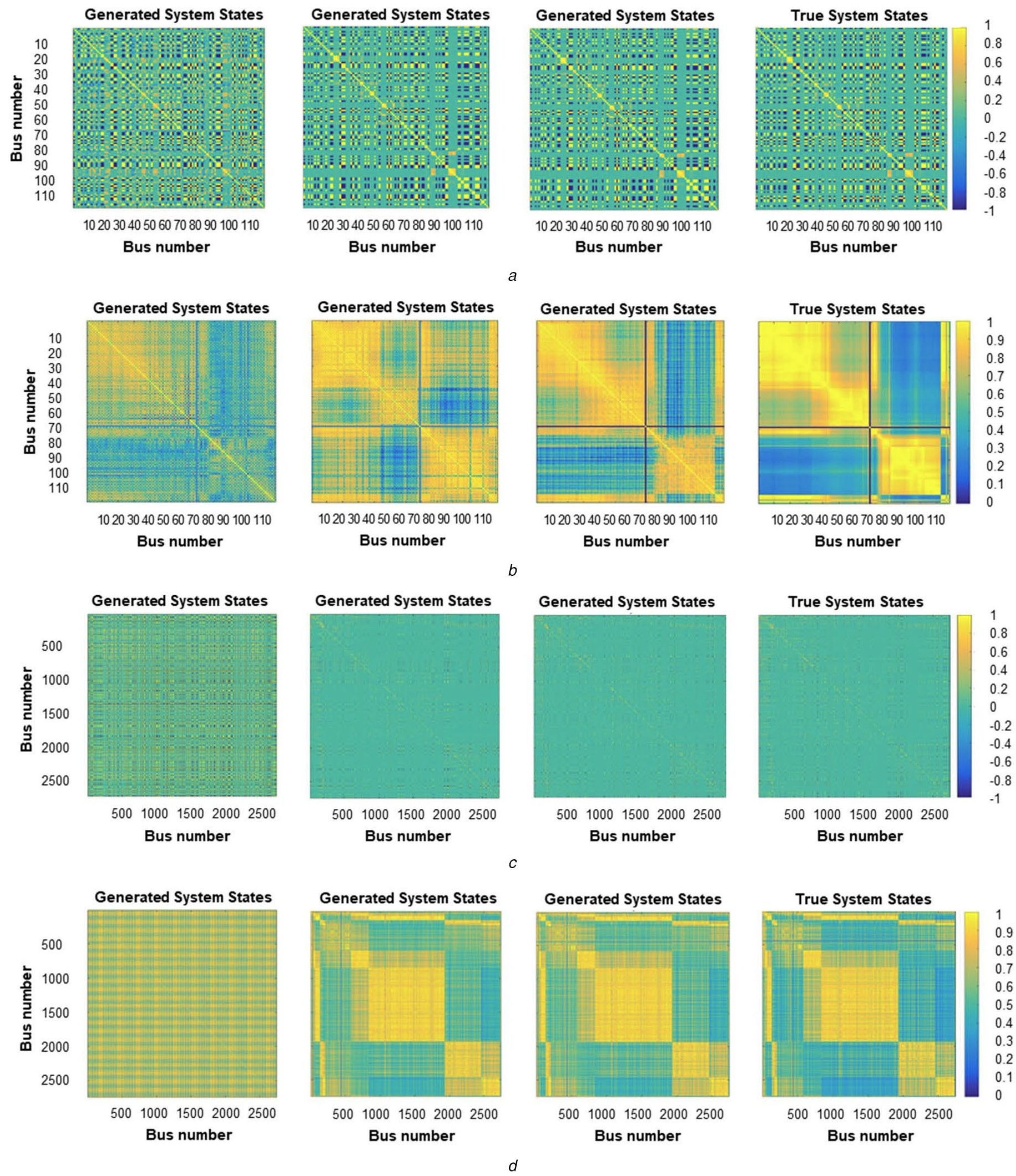


Fig. 3-8. The spatial correlation coefficients matrix colormap for different training iterations. From left to right: 200 iterations, 2000 iterations, 20000iterations, true system states. All results are tested with 30% contamination ratio of mixed contamination measurement (the right color bar is the correlation coefficient).

(a) and (b) are the voltage magnitude and phase angle spatial correlation on IEEE 118-bus system, respectively; (c) and (d) are the voltage magnitude and phase angle spatial correlation on 2746-bus Polish system, respectively.

3.4.4 Accuracy Comparison

As discussed before, only PMU measurements are considered in the network, and the measurement model is linear. Thus, the proposed method is compared with a linear WLS state estimator (LWLS-SE) [76]. To investigate the robustness of CGAN-SE, the WLAV state estimator (WLAV-SE) is also implemented by minimizing the L_1 norm between true system states and estimated states [82]. The experiments for LWLS-SE and WLAV-SE are carried out by comparing the MAE of voltage magnitude and phase angle for the IEEE 118-bus system and the 2746-bus Polish system, respectively. Especially, three different types of contamination measurement with the ratio ranging from 0% to 90% are considered.

In practice, the unobservable scenarios could occur with any ratio of missing input measurements, which can hardly be directly handled by the classical SE methods and are handled by using pseudo-measurements to replace the missing input measurements. As preliminary experiments, missing data ratios ranging from 10%~90% are considered to represent the unobservable scenarios. Since the occurrence of the missing measurement data is stochastic, r%(10%~90%) raw measurements are randomly chosen and set as missing data and assumed as zeros in the experiments. For example, if 10% of raw measurements are missing data in an adjacent area of the IEEE 118-bus system, some buses would not be monitored, which corresponds to saying that the system is unobservable. Besides, with the increased number of missing measurements, the system tends to be more unobservable, particularly by classical SE methods. The missing data problem is also tangled with the possible situation of partial system outages, where the missing input data to SE models could be attributed to the actual outages of system elements. Under such situations, the proposed SE method should provide estimated states properly, which is worthy of further investigations in the future.

The same measurement data and the experiments are applied and apply LWLS-SE and WLAV-SE to estimate voltage magnitude and phase angle. Besides, a comparison study between a deep convolutional neural network-based SE (DCNN-SE) method and CGAN-SE is conducted to show the advantages of CGAN-SE over other neural networks with deep learning [95].DCNN is a generator network, which is trained by the same dataset and verified via the same contamination ratio from 0% to 90%.

The comparative results of the proposed CGAN-SE against three benchmarks, LWLS-SE, WLAV-SE, and DCNN-SE under both test systems are listed in Table 3-3 and Table 3-4, respectively. In particular, the bold number in Tables 3-3, 3-4 are the minimum MAE of voltage magnitude and phase angle, respectively in each scenario. For the measurements containing noises only (r% = 0), the MAE of both voltage magnitude and phase angle for LWS-SE and WLAV-SE method has the PMU measuring precision close to 0.001. In the cases with corrupted measurements for both systems, DCNN-SE achieves significant improvements over the traditional LWLS-SE and WLAV-SE but is inferior to the proposed CGAN-SE method.

In the context of missing and mixed contamination measurements, the proposed CGAN-SE consistently outperforms LWLS-SE, WLAV-SE, and DCNN-SE. With increased contamination ratios, the MAE of LWLS-SE and WLAV-SE increase dramatically. The reason is that more unobservable scenarios occur with a larger number of missing measurements or mixed measurements, especially in the case of measurement loss, where it shows the largest MAE. On the other hand, CGAN-SE performs well because it is a data-driven and model-free approach. By applying the conditional generative adversarial networks, the actual correlations of system states can be well captured, and the system states can be accurately estimated without prior knowledge of the system model. Therefore, even the system is unobservable, CGAN-SE can still estimate system states with small errors.

Especially for robust WLAV-SE implemented on IEEE 118-bus system with 10% corrupted measurement, the smallest MAE for voltage magnitude and voltage phase angle is 9.032e-03 and 1.115e-02, respectively. While its MAE becomes larger as the missing ratio or corrupted ratio increases. In these cases, the system operator cannot monitor any violations or events at the buses with bad SE results, which may lead to catastrophic outcomes. The large MAE owes to a large number of bad data, this is distinct from the experiments in most studies that only very little bad data (usually less than 5%) is considered. In the latter cases, the system can always be observable, and WLAV-SE can perform well with minor errors.

Besides, the MAE of DCNN-SE grows rapidly, especially for mixed contamination measurement. The phenomenon indicates that the method is unable to estimate system states accurately. In contrast, the proposed method is not significantly influenced and still can maintain the error within the acceptable range. The major reason for these results is that the GAN in the proposed method consists of a generator and a discriminator, while DCNN only has a generator, and the discriminator can enhance the performance of the generator by providing the feedback (Wasserstein distance) between the true system states and the generated system states during the training process.

Table 3-3 Performance comparison with benchmarks on the IEEE 118-bus system.

	Contamination Type	Contamination Level r%	IEEE 118-bus System							
			Voltage Magnitude (p.u.)				Phase Angle (rad)			
			LWLS-SE	WLAV-SE	DCNN-SE	CGAN-SE	LWLS-SE	WLAV-SE	DCNN-SE	CGAN-SE
MAE	Noise only	0%	1.030e-03	1.152e-03	6.031e-04	**5.119e-04**	**9.910e-04**	1.226e-03	1.984e-03	1.561e-03
	Corrupted	10%	5.086e-02	9.032e-03	8.004e-04	**5.496e-04**	6.203e-02	1.115e-02	4.972e-03	**1.868e-03**
		30%	1.308e-01	4.801e-02	8.092e-04	**5.683e-04**	1.427e-01	5.207e-02	5.669e-03	**1.988e-03**
		50%	2.205e-01	1.104e-01	8.102e-04	**5.745e-04**	2.282e-01	1.331e-01	6.242e-03	**2.177e-03**
		70%	3.103e-01	2.768e-01	8.185e-04	**5.864e-04**	3.265e-01	2.959e-01	6.945e-03	**2.290e-03**
		90%	4.301e-01	4.437e-01	8.289e-04	**6.046e-04**	4.432e-01	5.520e-01	8.079e-03	**2.443e-03**
	Missing	10%	9.796e-02	1.608e-02	8.031e-04	**5.329e-04**	1.301e-01	1.896e-02	4.993e-03	**1.625e-03**
		30%	2.761e-01	9.353e-02	8.234e-04	**5.591e-04**	2.922e-01	9.700e-02	6.912e-03	**1.920e-03**
		50%	4.631e-01	2.097e-01	8.281e-04	**5.658e-04**	4.750e-01	2.277e-01	8.187e-03	**2.018e-03**
		70%	5.989e-01	5.572e-01	8.324e-04	**5.742e-04**	6.139e-01	5.781e-01	9.952e-03	**2.126e-03**
		90%	8.863e-01	8.958e-01	8.331e-04	**5.835e-04**	8.930e-01	8.894e-01	1.486e-02	**2.234e-03**
	Mixed	10%	8.036e-02	1.338e-02	8.245e-04	**5.526e-04**	9.029e-02	1.537e-02	1.163e-02	**1.922e-03**
		30%	2.072e-01	7.824e-02	8.432e-04	**5.777e-04**	2.220e-01	7.981e-02	1.823e-02	**2.619e-03**
		50%	3.299e-01	1.940e-01	8.795e-04	**5.876e-04**	3.387e-01	2.253e-01	2.916e-02	**3.962e-03**
		70%	4.615e-01	4.309e-01	8.978e-04	**6.000e-04**	4.723e-01	5.435e-01	4.450e-02	**5.108e-03**
		90%	6.663e-01	6.813e-01	9.111e-04	**6.387e-04**	6.798e-01	6.993e-01	6.770e-02	**5.344e-03**

Table 3-4 Performance comparison with benchmarks on 2746-bus Polish system.

	Contamination Type	Contamination Level r%	2746-bus Polish System							
			Voltage Magnitude (p.u.)				Phase Angle (rad)			
			LWLS-SE	WLAV-SE	DCNN-SE	CGAN-SE	LWLS-SE	WLAV-SE	DCNN-SE	CGAN-SE
MAE	Noise only	0%	1.020e-03	1.290e-03	8.343e-04	**6.677e-04**	1.070e-03	1.238e-03	4.900e-03	**5.989e-04**
	Corrupted	10%	5.487e-02	9.812e-03	1.305e-03	**7.339e-04**	6.385e-02	1.176e-02	1.228e-02	**1.113e-03**
		30%	1.404e-01	5.191e-02	1.320e-03	**7.609e-04**	1.503e-01	5.479e-02	1.402e-02	**1.222e-03**
		50%	2.366e-01	1.263e-01	1.339e-03	**8.679e-04**	2.420e-01	1.418e-01	1.542e-02	**1.425e-03**
		70%	3.328e-01	2.909e-01	1.355e-03	**9.105e-04**	3.473e-01	3.248e-01	1.715e-02	**1.997e-03**
		90%	4.612e-01	4.827e-01	1.409e-03	**9.459e-04**	4.723e-01	4.944e-01	1.996e-02	**2.281e-03**
	Missing	10%	1.013e-01	1.515e-02	1.334e-03	**6.231e-04**	1.336e-01	1.634e-02	1.533e-02	**1.055e-03**
		30%	2.819e-01	9.428e-02	1.368e-03	**6.662e-04**	2.972e-01	9.560e-02	1.707e-02	**1.181e-03**
		50%	4.712e-01	2.244e-01	1.381e-03	**7.305e-04**	4.836e-01	2.319e-01	2.022e-02	**1.262e-03**
		70%	6.106e-01	5.652e-01	1.447e-03	**7.715e-04**	6.254e-01	5.773e-01	2.458e-02	**1.609e-03**
		90%	8.921e-01	8.845e-01	1.475e-03	**8.051e-04**	8.908e-01	8.924e-01	3.670e-02	**1.847e-03**
	Mixed	10%	7.967e-02	1.282e-02	1.468e-03	**8.328e-04**	8.214e-02	1.335e-02	3.131e-02	**1.270e-03**
		30%	2.294e-01	7.869e-02	1.480e-03	**8.619e-04**	2.423e-01	7.919e-02	3.927e-02	**1.570e-03**
		50%	3.493e-01	1.831e-01	1.506e-03	**9.158e-04**	3.575e-01	1.925e-01	6.762e-02	**1.620e-03**
		70%	4.786e-01	4.495e-01	1.512e-03	**9.980e-04**	4.979e-01	4.561e-01	1.456e-01	**2.225e-03**
		90%	6.815e-01	7.003e-01	1.522e-03	**1.171e-03**	7.001e-01	7.147e-01	2.153e-01	**2.460e-03**

In the 2746-bus Polish system, the MAE in the DCNN-SE method increases sharply, especially for phase angle and mixed contamination cases. The MAE of these cases is larger than 1.0e-2, which does not satisfy the requirement of SE. In contrast, CGAN-SE can maintain a high degree of accuracy for both voltage magnitude and phase angle.

In summary, the MAE of voltage magnitude for the IEEE 118-bus system and 2746-bus Polish system in all scenarios ranges in [5.1e-4, 6.4e-4] p.u. and [6.7e-4, 1.2e-3] p.u., respectively. Also, the MAE of phase angle for the IEEE 118-bus system and 2746-bus Polish system in all scenarios ranges in [1.6e-3, 5.3e-3] rad and [6e-4, 2.5e-3] rad, respectively.

3.5 Summary

In this chapter, a model-free and data-driven deep learning based method is proposed for the SE of a power system. This method is based on CGAN, where the Wasserstein distance is applied as the loss function to improve training performance. With the corrupted or missing measurement at different contamination ratios, the proposed method can perform better than the traditional and state-of-the-art methods, i.e., LWLS-SE, WLAV-SE, and DCNN-SE. The proposed method CGAN-SE not only can estimate the system states with high accuracy but can also capture the statistical properties of the system measurements either from the probability distribution of system states or spatial correlation of buses. Moreover, in the case studies, the proposed SE method may produce satisfactory SE results for abnormal situations with 10%-90% missing data. Notably, missing input measurement data can frequently occur in practice, where the classical SE approaches have to handle such situations by using pseudo-measurements to replace the missing input measurements. It is important to point out that the study is preliminary for the proposed SE method to handle the missing data situations. Besides, the missing data problem is also tangled with the possible situation of partial system outages, where the missing input data to SE models could be attributed to the actual outages of system elements. Under such situations, it is very challenging to distinguish between the fundamental reasons for missing data due to either actual data missing during the data transportation/communication or the system outages. This issue is worthy of further investigation in the future but out of the scope of this Thesis. The effectiveness of the proposed CGAN-SE is validated through testing on the IEEE 118-bus system, as well as a large system with 2746 buses. The mean absolute error is less than 1.2e-3 p.u. and 5.3e-3 rad for voltage magnitude and phase angle, respectively, which is significantly better than traditional methods.

Chapter 4 A Novel Approach for Distribution System State Estimation Based on PUMs Considering Network Topology Changes

4.1 Introduction

State estimation (SE) is traditionally only applied in transmission systems. In recent years, SE based on PMUs or micro-PMUs is also discussed to enhance real time monitoring and control of distribution systems with increased penetration of renewables and other emerging technologies that introduce tremendous uncertainties and risks into system operation. Therefore, distribution system state estimation (DSSE) becomes a fundamental part of the distribution management systems (DMSs) needed for the monitoring and control of the future smart grid [96]. DSSE supplies the real-time system operating states to a variety of DMS applications as their inputs [97].

The topology of the network can be defined directly by the status of switching devices [88]. The classical SE methods involve complicated and separated steps to deal with network topology, observability, SE calculation, and bad data. In transmission systems, the system topology is generally deemed to be very constant. Therefore, many methods premise that the system topology is unchanged and completely known based on monitoring devices. However, this assumption is invalid for distribution systems, where the branch switch statuses may be unknown or doubtful due to frequent reconfiguration actions and limited topology measurements [18]. As a consequence, classical SE methods are inapplicable without an ascertained topology [32]. In order to estimate system states accurately in distribution systems, the appropriate method of DSSE considering the topology changes is essential.

To handle varying network topology issues in an integral framework and improve the SE performance for the out-of-sample topologies, in this chapter, based on chapter 3, a novel data-driven SE approach considering topology changes and data contaminations for distribution systems based on the topology-aware generative adversarial network (TAGAN) model with only PMU measurements applied, where a multilayer perceptron (MLP) based detector (for measurement data quality assessment and labeling), the spectral normalization discriminator and the hinge loss function are considered to improve SE performance. TAGAN model is based on the classical GAN, which has several merits, including 1) easier to compute than the classical GAN method due to its simple mathematical formulation of the loss function. 2) faster to be trained, since during most of the training epochs, the gradient of the loss function is zero, and the network weights do not need to be updated. 3) more robust against data contaminations due to its nonlinear loss functions [115]. This inspires us to apply TAGAN in the DSSE process, which can accurately estimate system states without prior knowledge of the system topology. The proposed TAGAN based SE method can effectively reconstruct the real system states considering the corrupted raw measurements or even missing measurements under varying topologies. Especially for the out-of-sample topologies, the proposed method can still estimate the system states accurately because the spectrum normalization is applied in the discriminator.

The main contribution of this chapter lies in that a novel data-driven TAGAN model is proposed for DSSE, which represents the first effort of applying one integrated deep learning framework for SE that is capable of addressing the uncertainties involved in both grid topology and state measurement simultaneously. Unlike the existing data-driven approaches that can only handle a finite topology space, the proposed method by applying conditional GAN with the spectral normalization is capable of tackling a variety of out-of-sample topologies. Besides, a detector for measurement data quality assessment and labeling and the hinge loss function are applied in the TAGAN model to improve SE accuracy. Extensive experiments have been carried out to examine the influence of data contaminations with respect to different ratios and types, which are rarely considered by most existing works. Thus, the proposed method is proved to be robust to the corrupted measurements or missing measurements, making the SE viable even in a traditionally unobservable network.

The remaining sections of this chapter are organized as follows: Section 4.2 introduces the problem formulation. Section 4.3 presents the proposed TAGAN based SE model. Section 4.4 gives the results and the analysis on the IEEE 33-node system and IEEE 118-bus distribution system. Finally, Section 4.5 concludes this chapter.

4.2 The Proposed DSSE Models Considering Contaminated Measurements under Varying Topologies

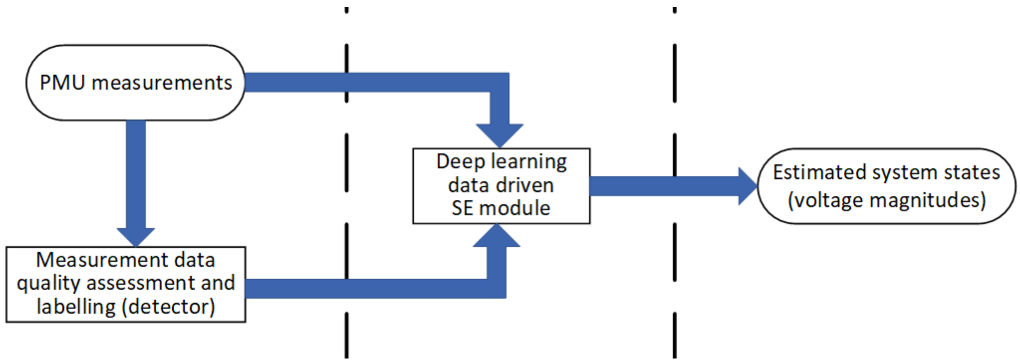


Fig. 4-1. The overall framework of the proposed DSSE model.

As shown in Fig. 4-1, the proposed DSSE model includes four parts, the input, measurement data quality assessment and labeling (detector), deep learning data-driven SE module, and the output. For the input, in this chapter, it is assumed that the distributed system is observed merely by PMUs, i.e., all the obtained measurements are voltage magnitudes while phase angles are usually not considered in distribution systems [98]. The data-driven SE module cannot fully handle corrupted and missing data when PMU measurements are directly input in the SE module. Therefore, it is meaningful to preprocess PMU measurements so that the measurement data quality can be well assessed and labeled as in the measurement data type vector, which will be input into the SE module with PMU measurements simultaneously in order to enhance the overall performance of SE. The following sections will give the details. The proposed data-driven SE module is based on a deep learning framework capable of handling contamination data (including corrupted data and missing data), and it will be detailed discussed in Section 4.3. The output is the estimated system states (voltage magnitudes), whose updating frequency is in line with the PMUs and faster than SCADA measurements. The proposed

DSSE model is an integral framework to directly provide SE results, which stays away from the complicated and separated steps in classical SE methods.

For an N-bus distribution system, the vector of state variables $\mathbf{x}$ is denoted by $[V_1 \ldots, V_j \ldots, V_N]^T$, $j = 1, \ldots, N$. V_j is the voltage magnitude of j-th bus. As the general SE formulation presented in section 2.2.1 and the linear SE model with PMUs presented in 3.2.1, the same measurement model is also applicable to the distribution system.

In the proposed DSSE model, the topology change refers to the change of the branch switch statuses while the bus number is unchanged. In the DSSE model, the number of branches is NB, the topology measurements can be represented by the vector of switch statuses sw and expressed as $[sw_1, \ldots sw_j \ldots, sw_{NB}]^T$, $j = 1, \ldots, NB$. For sw_j, "0" represents that the branch j is open, while "1" represents that the branch j is closed.

In the DSSE model, the raw measurement vector $\mathbf{y}$ is composed of the measured voltage magnitude y_j, $j = 1, \ldots, N$ of all buses, and this vector can be expressed as $[y_1, \ldots y_j \ldots, y_N]^T$. As introduced in section 3.2.2, the raw measurement vector with corrupted system measurements, i.e., bad data, is denoted as $\mathbf{y_C} = [y_1 \ldots y_{n_C} \ldots y_N]^T$, where y_{nc} denotes the corrupted measurements. Likewise, as introduced in section 3.2.3, the raw measurement vector with missing measurements is denoted as $\mathbf{y_L} = [y_1 \ldots 0 \ldots y_N]^T$, where the missing measurements y_j, resulting from measurement loss, are substituted by zero.

Besides, the measurement data type vector, which is generated by the measurement data quality assessment and labeling module, i.e., the detector based on the raw measurement input, is denoted by $\mathbf{t}=[t_1 \ldots t_j \ldots t_N]^T$, $j=1,\ldots,N$, which is specially introduced to improve SE accuracy. The details of the detector and the measurement data type vector will be introduced in Section 4.3.1. For t_j, "1" represents normal measurement data, and "0" represents corrupted or missing measurement data.

The objective of this DSSE problem is to train a generative model based on TAGAN by using raw measurements and true system states. Let $\tilde{\mathbf{x}}$ denote the true system states, i.e., voltage magnitudes in this chapter, and the training samples are obtained as the pairs of measurements and true system states $((\mathbf{y},\mathbf{t}),\tilde{\mathbf{x}})$. In this model, the inputs are voltage measurements and measurement data types $(\mathbf{y},\mathbf{t})$, each of the inputs $(\mathbf{y},\mathbf{t})$ is stamped with a given label (true system states $\tilde{\mathbf{x}}$) as a condition. Once plenty of $((\mathbf{y},\mathbf{t}),\tilde{\mathbf{x}})$ pairs are obtained, the goal is to train a TAGAN network to generate the estimated system states $\hat{\mathbf{x}}$ that are expected to be close to the true system states $\tilde{\mathbf{x}}$ as much as possible. Then, this model can be formulated into:

$$\begin{cases} \hat{\mathbf{x}} = \arg\min\limits_{\mathbf{x}} \left\| f\left(\mathbf{x} \middle| (\mathbf{y},\mathbf{t})\right) - d_0 \right\| \\ d_0 = f\left(\tilde{\mathbf{x}} \middle| (\mathbf{y},\mathbf{t})\right) \end{cases} \tag{4.1}$$

where *f(·)* represents the TAGAN model, and it is a value function of the two-player minimax game; d_0 is the value of *f(·)* with the true system states input. The single output of *f(·)* represents the training process of the TAGAN model. The difference between $f\left(\mathbf{x} \middle| (\mathbf{y},\mathbf{t})\right)$ and d_0 reflects the similarity between $\mathbf{x}$ and $\tilde{\mathbf{x}}$. The details of the TAGAN and the two-player minimax game value function will be presented in Section 4.3.1.

4.3 The Proposed Topology-Aware Generative Adversarial Network

The topology-aware GAN (TAGAN) model for SE is presented in this section, where the spectral normalization is introduced to handle topology variations. This section describes the structure of TAGAN models first. In the meanwhile, how the framework fits into the SE problem with varying topologies is discussed. Then, the learning procedure of TAGAN for SE by integrating the synthetic PMU measurement is presented.

4.3.1 Structure of TAGAN Model

In this section, each part of the TAGAN model, including the detector, the hinge loss-based conditional GAN and spectral normalization based discriminator will be introduced. Then, how the TAGAN model can fit well into the SE problem with varying topologies will be discussed in detail.

A. The Detector

In computer vision research, image inpainting work is aimed to mend damaged images due to pollution or deficiency. In general, as shown in Fig. 4-2, a mask that the damage region of the image is in white, and the rest is in black is applied as a part of the input. The mask of damage region is the pretreatment step of the image inpainting, which plays a key role in the ultimate effect and improves the inpainting performance [114]. Inspired by this, a detector to detect and accordingly “mask” (label) the normal raw measurement or missing/corrupted raw measurement is designed to output a measurement data type vector like the “mask” applied in image inpainting.

Fig. 4-2. A mask in image inpainting.

Let y denote the raw measurement from PMUs, which may be corrupted or involve missing data. The detector T is designed by a multilayer perceptron (MLP) based neural network because of its satisfactory performance on detecting/labeling the input measurements and simple architecture. The detector is expressed as $T(y;\theta_t^{(T)})$, which is parameterized by $\theta_t^{(T)}$. The detector's input is the raw measurement from PMUs, i.e., y, and after training, the detector's output is the measurement data type vector t consisting of 0s and 1s. The training procedure can be expressed as the following mapping:

$$T(y;\theta_t^{(T)}): y \rightarrow t \tag{4.2}$$

The measurement data type vector t corresponds to the classical system buses voltage magnitude type, of which 1 denotes normal data, and 0 denotes missing/corrupted data. Fig. 4-3 illustrates the structure of the detector. For raw measurements y, the color block denotes the voltage magnitude, in particular, white blocks represent missing measurements, and black blocks represent corrupted measurements. The detector serves for detecting and labeling out different types of data, which corresponds to performing a binary classification task by the conventional MLP neural network. The detailed architecture of the detector is given in the following sections.

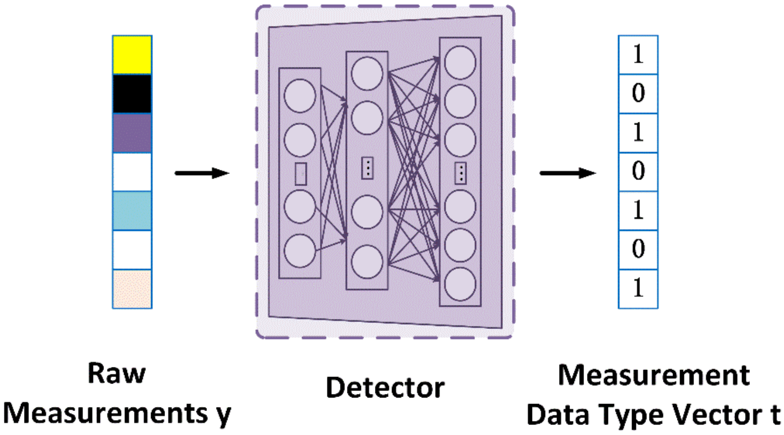


Fig. 4-3. The structure of the detector.

B. The Hinge Loss-Based Conditional GAN

As defined in section 4.2, $\tilde{\mathbf{x}} = [\tilde{x}_1 \ldots \tilde{x}_N]^T$ represents the true system state, i.e., voltage magnitude, and $\{\tilde{x}_j^{(i)}\}_{i=1}^{m_t}$ denotes the *i*-th sample of the true system state, and m_t is the number of samples. Let $p_{data}(x)$ denote the distribution of the true system state. Suppose a group of noise inputs z follow a known distribution $z \sim p_z(z)$, e.g., uniform distribution or joint Gaussian. The goal is to transform the sample z from the distribution $p_z(z)$ so that it can follow true system state distribution $p_{data}(x)$. To this end, two deep neural networks are trained simultaneously. One is the generator network *G* denoted as $G(z;\theta_t^{(G)})$, which is parameterized by $\theta_t^{(G)}$; the other is the discriminator network *D* denoted as $D(x;\theta_t^{(D)})$, which is parameterized by $\theta_t^{(D)}$. The details of classical GAN, including the generator and the discriminator, have been introduced in section 3.3.1.

The classical GAN can be extended to a conditional counterpart, where both the generator and discriminator are conditioned on some extra information, i.e., condition. The proposed TAGAN model adopts the conditional GAN frameworks, where the condition consists of the raw measurement y and the measurement data type vector t obtained by the pretreated detector *T*. Then the generator and discriminator of the proposed TAGAN can be rewritten as:

$$G(z|(y,t);\theta_t^{(G)}): z|(y,t) \rightarrow p_G(z) \tag{4.3}$$

$$D(x|(y,t);\theta_t^{(D)}): x|(y,t) \rightarrow p_{real} \tag{4.4}$$

where $z|(y,t)$ and $x|(y,t)$ denote the inputs to the generator and discriminator with the conditions (y,t) added, respectively.

Formally, *G* and *D* are playing a two-player minimax game with the value function $V(G,D)$:

$$\begin{aligned}\min_G \max_D V(G,D) = {} & \mathbb{E}_{x\sim p_{data}(x)}[\log(D(x|(y,t)))] \\ & +\mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z|(y,t))))]\end{aligned} \tag{4.5}$$

For a given *D*, since the output value of the discriminator p_{real} is close to 1 indicating that the generated samples are more realistic, the generator is expected to minimize $\log(1-D(G(z|(y,t))))$ by changing the parameter of *G* to generate more authentic samples. For a given *G*, the discriminator tries to maximize $\log(D(x|(y,t)))$, thus making the output value of the discriminator p_{real} close to 1. At the same time, the discriminator maximizes $\log(1-D(G(z|(y,t))))$.

The hinge loss is based on a soft-margin support vector machine (SVM) balancing two competing objectives to maximize the margin while penalizing points on the wrong side of the margin [115], which refers to the region bounded by two hyperplanes in a high or infinite dimensional input space of

SVM. In the TAGAN model, the hinge loss is applied with serval merits compared to other loss functions like the cross entropy and the Wasserstein distance based ones: 1) easier to compute than other loss functions and the same with its gradient due to its simple mathematical formulation. 2) faster to be trained, since during much of the training epochs the gradient of the loss function is zero, and the network weights do not need to update. 3) more robust than other loss functions against data contaminations due to its nonlinear loss function [115].

Therefore, when integrated with spectral normalization of weights, introduced in the next part, the hinge loss can significantly enhance GAN's generative performance [116] and improve the SE accuracy. The numerical experiments are conducted in section 4.4.3, where the training performance of the proposed TAGAN model with the hinge loss outperforms the training performance very much with the cross entropy and Wasserstein distance in all data sets. Besides, the TAGAN model with the hinge loss converges faster and is more robust against data contaminations than the other two loss functions. The hinge loss of the discriminator C_{D_hinge} and the generator C_{G_hinge} can be expressed as:

$$\begin{aligned} C_{D\_hinge} &= \mathbb{E}_{z\sim p_{data}(x)}[\min(0,1-D(x))] \\ &\quad + \mathbb{E}_{z\sim p_z(z)}[\min(0,1+D(G(z))] \end{aligned} \tag{4.6}$$

$$C_{G\_hinge} = -\mathbb{E}_{z\sim p_z(z)}[D(G(z))] \tag{4.7}$$

In the training stage, *D* is trained to maximize the ability to discern the true state distribution of the system and the generator estimated state distribution. *G* and *D* are trained at the same time to minimize the difference between these two distributions. The weights of *G* and *D* are updated to minimize the generator loss function C_{G_hinge} and the discriminator loss function C_{D_hinge}, respectively. To be specific, the training starts with inputting a batch of samples from the distribution $p_z(z)$ and the conditions (y,t) into *G*. At the same time, inputting a batch of real samples taken from distribution $p_{data}(x)$ and the conditions (y,t) into *D*. Small C_{G_hinge} indicates that from the perspective of the discriminator, the generated sample is more authentic. In other words, the generated system state is more similar to the real system state. On the other hand, small C_{D_hinge} indicates that the generator *D* is excellent in distinguishing the difference between the generated system state and the real system state. Moreover, this shows that the generated state distribution $p_G(z)$ is very different from the real state distribution $p_{data}(x)$.

C. Spectral Normalization Discriminator Addressing Topology Variations

In this Chapter, grid topology variation refers to the change of the branch switch statuses while the buses in operation remain unchanged in power grids, which could often occur due to faults or maintenance works in practice. The classical GAN based SE method in Chapter 3 has difficulty in addressing power grid topology variations. The main reason is that, firstly, in the classical GAN, the overall loss function is cross entropy. For the generator, the generator's loss function can be equivalent to the JS divergence [87]. The JS divergence applied as the loss function during training the generator in the classical GAN cannot reflect the similarity between the two distributions $p_G(z)$ and $p_{data}(x)$ when they do not intersect in the distribution space wherein JS divergence is a constant $-2\log 2$ rather than a variable varying with the similarity between the distributions, the generator may collapse to produce limited varieties of samples regardless of the inputs, i.e., mode collapse. Especially for DSSE with varying topologies, once mode collapse occurs, the generator is more likely to generate the system states corresponding to the topologies with smaller training loss among all topologies in the training instead of the system states corresponding to its true topology [87]. An experiment is conducted at the end of this subsection to prove that mode collapse is more likely to occur using the classical GAN only with JS divergence in generating system states with varying topologies. Secondly, the classical GAN is hard to reach the Nash equilibrium in handling SE considering varying topologies, and its generative performance is sensitive to its initial values of parameters. Thus, the classical GAN may not converge properly during the training with inappropriate initial values of parameters. Then, if the discriminator is trained to be too good in the discernment capacity, the generator training can fail due to no gradient of the loss function to update the loss, commonly referred to as vanishing gradients. Given the above two issues, the spectral normalization (SN) is applied in the TAGAN's discriminator to handle these problems and thus the TAGAN can effectively estimate system states under varying network topologies. Likewise, an experiment is conducted at the end of this subsection to show that vanishing gradients is more likely to occur using the classical GAN without spectral normalization in generating system states with varying topologies, and the application of spectral normalization can mitigate vanishing gradients.

According to [87], for a fixed generator, the optimal discriminator for the standard form of (4.5) is given by:

$$D_G^* = \frac{p_{data}(x|(y,t))}{p_{data}(x|(y,t)) + p_G(x|(y,t))} = sigmoid(\mathtt{f}^*(x|(y,t))) \tag{4.8}$$

where *sigmoid*() is the sigmoid activation function and $\mathtt{f}^*(x|(y,t))$ is the optimal discriminator network without activation function. Then, the equation can be solved:

$$\mathtt{f}^*(x|(y,t)) = \log p_{data}(x|(y,t)) - \log p_G(x|(y,t)) \tag{4.9}$$

Also, its derivative is:

$$\begin{aligned}\nabla_x \mathtt{f}^*(x|(y,t)) = &\frac{1}{p_{data}(x|(y,t))}\nabla_x p_{data}(x|(y,t)) \\ &- \frac{1}{p_G(x|(y,t))}\nabla_x p_G(x|(y,t))\end{aligned} \tag{4.10}$$

This derivative is unbounded or even incomputable, and this results in above mentioned problems (mode collapse and vanishing gradient) in GANs training. Controlling the Lipschitz constant of the discriminator can solve these problems, and its effectiveness has been verified in [89]. For the discriminator D from the set of K-Lipschitz continuous functions, specifically,

$$\underset{\|\mathtt{f}\|_{Lip} \le K}{\arg\max}\, V(G,D) \tag{4.11}$$

where $\|\mathtt{f}\|_{Lip}$ denotes Lipschitz norm of $\mathtt{f}(x|(y,t))$ and it is the smallest value P such $\|\mathtt{f}(x_1) - \mathtt{f}(x_2)\| / \|x_1 - x_2\| \le P$ for any x_1, x_2, with the norm being the l_2 norm; Lipschitz constant K is a nonnegative real number.

Once $\mathtt{f}(x|(y,t))$ is Lipschitz continuous for a real constant K, then, the problems of diminished gradient and mode collapse can be handled. Therefore, the spectral normalization is applied to enable f to be Lipschitz continuous, that is, $\|f\|_{Lip}$ can be constrained by *K* for each layer of the discriminator.

In the discriminator *D*, for each layer $g: s_i \to s_{out}$. $\sigma(C)$ is the spectral norm of the matrix *C* (L_2 matrix norm of *C*),

$$\sigma(C) := \max_{s \neq 0} \frac{\|Cs\|_2}{\|s\|_2} = \max_{\|s\|_2 \leq 1} \|Cs\|_2 \tag{4.12}$$

where $\sigma(C)$ is equivalent to the largest singular value of *C*. Therefore, by definition of Lipschitz norm, $\|g\|_{Lip}$ is equal to $\sup_s \sigma(\nabla g(s))$.

For a linear layer of neural networks $g(s) = Ws$, there is $\|g\|_{Lip} = \sup_s \sigma(\nabla g(s)) = \sup_s \sigma(W) = \sigma(W)$ and it has the inequality:

$$\|g_1 \circ g_2\|_{Lip} \leq \|g_1\|_{Lip} \cdot \|g_2\|_{Lip} \tag{4.13}$$

where '$\circ$' denotes the product of two linear layer functions, and '$\cdot$' denotes the product of their respective Lipschitz norms.

If the Lipschitz norm of the activation function $\|b_l\|_{Lip}$ equals 1 (with ReLU or leaky ReLU), then have the bound on $\|f\|_{Lip}$:

$$\begin{aligned} \|f\|_{Lip} \leq & \left\|s_L \to W^{L+1} s_L\right\|_{Lip} \cdot \|b_L\|_{Lip} \cdots \left\|s_1 \to W^2 s_1\right\|_{Lip} \cdot \|b_1\|_{Lip} \\ & \cdot \left\|s_0 \to W^1 s_0\right\|_{Lip} = \prod_{l=1}^{L+1} \left\|s_{l-1} \to W^l s_{l-1}\right\|_{Lip} = \prod_{l=1}^{L+1} \sigma(W^l) \end{aligned} \tag{4.14}$$

Then, the spectral normalization can be applied to normalize the spectral norm of the weight matrix *W*:

$$\bar{W}_{SN}(W) := W / \sigma(W) \tag{4.15}$$

Therefore, $\sigma(\bar{W}_{SN}) = \sigma(W) / \sigma(W) = 1$ and the function of the discriminator $\|f\|_{Lip}$ is 1- Lipschitz.

Rather than using singular value decomposition (SVD) to calculate the spectral norm $\sigma(W)$ at each iteration, which is very computationally heavy, the power iteration method [117] is applied to estimate $\sigma(W)$. This fast method is detailed in Section 4.3.2.

Since the JS divergence is applied as the generator's loss function in classical GAN, the degree of mode collapse is hard to be measured by the loss function and needs other metrics to measure. The intra multi-scale structural similarity (MS-SSIM) is a prevalent similarity metric to measure the degree of mode collapse in generating system states [121]. Typically, the MS-SSIM values range between 0.0 and 1.0; higher MS-SSIM values represent more similar system states samples, which indicates a higher degree of mode collapse [122]. The MS-SSIM is given by [121],

$$\text{MS-SSIM}(\mathbf{x_1}, \mathbf{x_2}) = [l_M(\mathbf{x_1}, \mathbf{x_2})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x_1}, \mathbf{x_2})]^{\beta_j} [s_j(\mathbf{x_1}, \mathbf{x_2})]^{\gamma_j} \tag{4.16}$$

where,

$$l(\mathbf{x_1}, \mathbf{x_2}) = \frac{2\mu_{x_1}\mu_{x_2} + C_1}{\mu_{x_1}^2 + \mu_{x_2}^2 + C_1} \tag{4.17}$$

$$c(\mathbf{x_1}, \mathbf{x_2}) = \frac{2\sigma_{x_1}\sigma_{x_2} + C_2}{\sigma_{x_1}^2 + \sigma_{x_2}^2 + C_2} \tag{4.18}$$

$$s(\mathbf{x_1}, \mathbf{x_2}) = \frac{\sigma_{x_1 x_2} + C_3}{\sigma_{x_1}\sigma_{x_2} + C_3} \tag{4.19}$$

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2, C_3 = C_2 / 2 \tag{4.20}$$

M is the number of scales; the exponents α_M, β_j and γ_j are applied to adjust the relative importance of different components; by default, $\alpha_j = \beta_j = \gamma_j$ for all j and $\sum_{j=1}^{M} \gamma_j = 1$; functions l, c and s are three components of MS-SSIM; $\mathbf{x_1}$ and $\mathbf{x}_2$ are two samples of system states; μ_{x_1}, μ_{x_2} are the mean of $\mathbf{x_1}, \mathbf{x_2}$, respectively; σ_{x_1}, σ_{x_2} are the variance of $\mathbf{x_1}, \mathbf{x_2}$, respectively; $\sigma_{x_1 x_2}$ is the covariance of $\mathbf{x_1}$ and $\mathbf{x}_2$; C_1, C_2 and C_3 are small constants; K_1, K_2 are two scalar constants and K_1=0.01, K_2=0.03 by default; MS-SSIM is originally used for measuring the similarity between two images in image processing [121, 122]. Higher MS-SSIM scores indicate higher similarity between images and a higher degree of mode collapse [122]. In the application of image processing, L is defined as the dynamic range of the pixel values and is set as 2^n-1 (n is a positive integer). Typically, L should be just larger than the number of pixels [121]. For the SE problem, the number of buses in a system is analogous to the number of pixels in an image. Since L should be just larger than the number of buses, e.g., for the 33-bus system, n=6 and L=63 are adopted.

To evaluate the effectiveness of spectral normalization in addressing the mode collapse, a comparison experiment is conducted using the IEEE 33-node distribution system. Classical GAN and TAGAN models with/without spectral normalization are trained in this comparison experiment. The training data are system states (nodal voltages' magnitudes) considering different system topologies with noises only. 20 topologies are included, and each topology has 1000 system states samples (the total number of samples is 20000). The batch size is 32, and the number of epochs is 625. Therefore, the training of the models takes 400,000 iterations (20000/32*625). The details of data generation and training are introduced in the following section 4.4. In the experiment, MS-SSIM scores between 100 randomly chosen pairs [122] of system states samples in each topology are computed to measure the diversity of system states with varying topologies. The higher diversity of system states results in lower mean MS-SSIM scores. On the contrary, system states samples with lower diversity have higher mean MS-SSIM scores, which means mode collapse occurs. The mean MS-SSIM score is computed for system states under 20 topologies in Fig. 4-4.

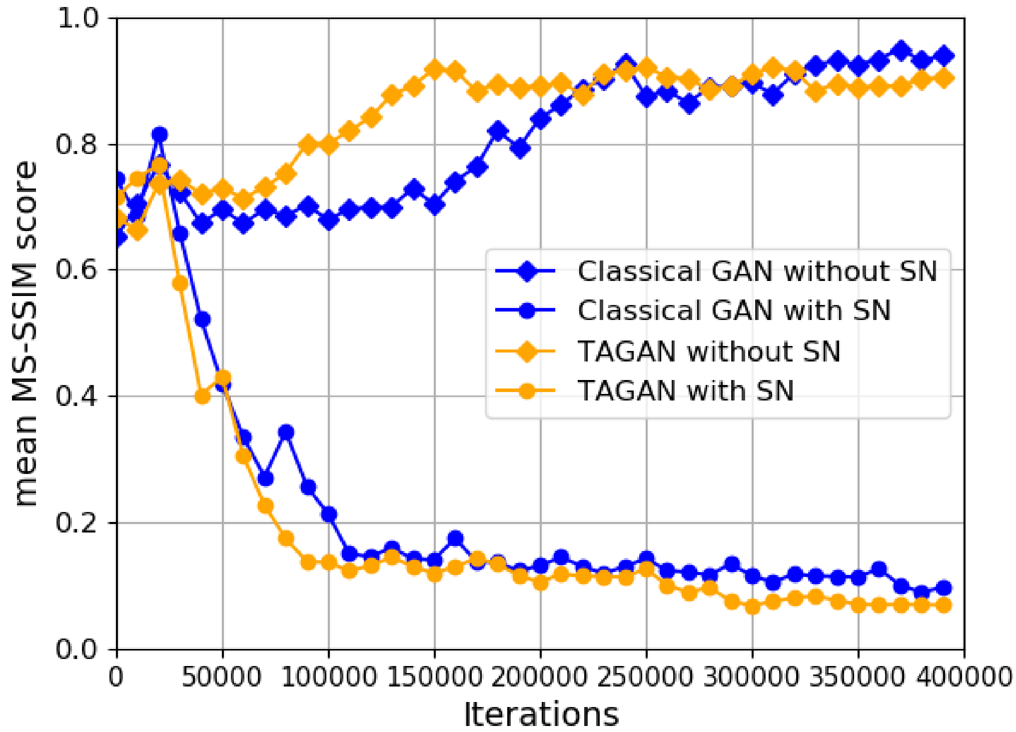


Fig. 4-4. Mean MS-SSIM score for system states under 20 topologies through training the classical GAN and TAGAN with/without spectral normalization (SN).

From the results shown in Fig. 4-4, the mean MS-SSIM score is tracked during the training process to identify whether mode collapse has occurred. The blue circle marker line is the mean MS-SSIM score of system states under 20 topologies using the classical GAN with the JS divergence as the generator's loss function and spectral normalization implemented in the discriminator, while the blue diamond marker line is derived by the classical GAN only with JS divergence. The orange lines are the mean MS-SSIM scores using TAGAN, wherein the circle marker line indicates the results using TAGAN with SN, and the diamond marker line is that without SN. During the training process, the circle marker lines decrease and gradually converge to a low MS-SSIM score, which indicates high diversity of system states samples, and mode collapse does not occur due to the applied spectral normalization. On the other hand, the diamond marker lines increase and approach 1.0, which indicates a very low diversity of system states samples, and mode collapse occurs without SN in the classical GAN or TAGAN.

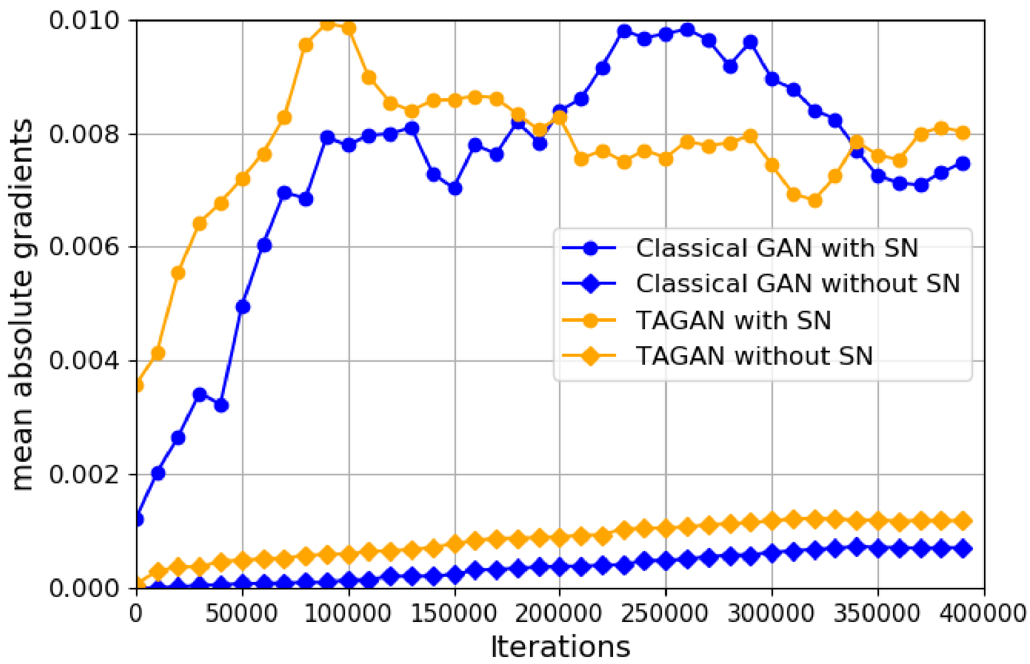


Fig. 4-5. Mean absolute gradients of the 1st layer in the generator under 20 topologies through training the classical GAN and TAGAN with/without spectral normalization (SN).

To evaluate the effectiveness of spectral normalization in addressing the vanishing gradient, a comparison experiment is conducted in the IEEE 33-node distribution system. Classical GAN and TAGAN models with/without spectral normalization are trained in this comparison experiment. The training data and model setting are the same as the experiment for the model collapse. In the experiment, the mean absolute gradients of the 1st layer in the generator are computed to examine the vanishing gradient problem [107]. The mean absolute gradients of the 1st layer in the generator are calculated by: $\frac{1}{NN_1} \cdot \sum_{i=1}^{NN_1} \left| \frac{\partial L^{(G)}}{\partial w_1^i} \right|$. NN_1 denotes the number of neurons in the 1st layer of the generator; $L^{(G)}$ is the generator loss function, which is JS divergence in the classical GAN, and the hinge loss function in the TAGAN; $\left| \frac{\partial L^{(G)}}{\partial w_1^i} \right|$ denotes the absolute gradients of the 1st layer in the generator. The smaller mean absolute gradients indicate a higher chance of vanishing gradient occurrence. From the results shown in Fig. 4-5, the mean absolute gradients of the 1st layer in the generator are tracked during the training process to identify whether the vanishing gradient has occurred. The circle marker lines are the mean absolute gradients using the classical GAN (blue) / TAGAN (orange) with spectral normalization implemented, while the diamond marker lines are derived by the classical GAN (blue) / TAGAN (orange) without spectral normalization. During the training process, the diamond marker lines are always at a very low level, which indicates that vanishing gradient occurs without spectral normalization in the classical GAN or TAGAN. On the contrary, the overall magnitudes of the gradients with spectral normalization are significantly greater than those without spectral normalization using the classical GAN or TAGAN. Thus, spectral normalization is demonstrated to be effective in addressing the vanishing gradient.

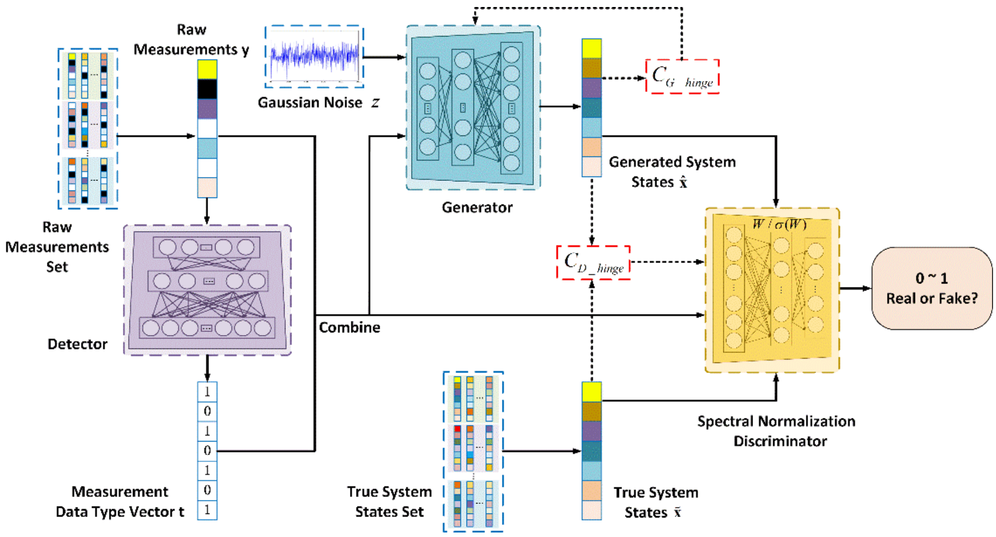


Fig. 4-6. The training process of the TAGAN model.

The training process of the TAGAN model and the learning process can be illustrated in Fig. 4-6. For raw measurements, they are sampled from the raw measurement set with a variety of topologies. The color block denotes the voltage magnitude, in particular, white blocks represent missing measurements, and black blocks represent corrupted measurements. For the measurement data type vector, it corresponds to the system buses voltage magnitude type, of which 1 denotes normal data, and 0 denotes missing/corrupted data. Combining raw measurements and measurement data type vector becomes the conditions as input for the generator and discriminator. The discriminator and generator are trained with the hinge loss C_{D_hinge} and C_{G_hinge} alternately. In each iteration, the spectral normalization discriminator (SN-D) learns from the positive samples from $p_{data}(x)$ and negative samples from $p_G(z)$. Adversarial learning between *G* and SN-D enables them to improve their ability until SN-D cannot distinguish the generated system states from *G* and the true system states.

4.3.2 Procedure of TAGAN

The algorithm used in the TAGAN model is given in Algorithm 4-1. In this algorithm, $\theta_t^{(G)}$ and $\theta_t^{(D)}$ are the parameters of the generator $G(z)$ and discriminator $D(x)$. They are both neural networks combined by several different layers, including multilayer perceptron (MLP) layers, convolution layers, normalization layers, max-pooling layers, and Rectified Linear Units (ReLU) layers. Especially for the discriminator, its training algorithm is based upon the gradient descent and updates each layer of the discriminator with spectral normalization by applying the fast power iteration method in Algorithm 4-2. Besides, for both discriminator and generator, the optimization algorithms are Adaptive Moment Estimation (*Adam*), which is able to adjust the learning rate automatically. Within several training batches, these parameters of two neural networks can be well optimized. The detailed model parameter settings for the TAGAN for case studies will be given in Section 4.4.2.

Algorithm 4-1 TAGAN model

Require: α_t, the learning rate; m_t, the batch size; k_{dis_t}, the number of iterations of the discriminator per generator iteration.
Require: $\theta_{t0}^{(D)}$, initial discriminator's parameters; $\theta_{t0}^{(G)}$, initial generator's parameters.
while $\theta_{t0}^{(G)}$ has not converged **do**

---

for $t = 0,\ldots,k_{dis\_t}$ **do**

- Sample batch of m_t noise samples $\{(z^{(i)}, y^{(i)}, t^{(i)})\}_{i=1}^{m_t}$ from noise prior distribution $p_z(z)$.
- Sample batch of m_t samples $\{(x^{(i)}, y^{(i)}, t^{(i)})\}_{i=1}^{m_t}$ from the true system state data $p_{data}(x)$
- Calculate the gradient of discriminator loss:

$$g_{\theta_t^{(D)}} \leftarrow \nabla_{\theta_t^{(D)}} \frac{1}{m} \sum_{i=1}^{m_t} \begin{matrix} [\max(0,1-D(x^{(i)} | y^{(i)}, t^{(i)})) \\ +\max(0,1+D(G(z^{(i)} | y^{(i)}, t^{(i)})))] \end{matrix}$$

- Update each layer of the discriminator with spectral normalization by applying **Algorithm 4-2**.

end for

- Sample batch of m_t noise samples $\{(z^{(i)}, y^{(i)}, t^{(i)})\}_{i=1}^{m_t}$ from noise prior distribution $p_z(z)$.
- Update the generator by descending its gradient:

$$g_{\theta_t^{(G)}} \leftarrow -\nabla_{\theta_t^{(G)}} \frac{1}{m} \sum_{i=1}^{m_t} D(G(z^{(i)} | y^{(i)}, t^{(i)}))$$

$$\theta_t^{(G)} \leftarrow \theta_t^{(G)} - \alpha_t \cdot Adam(\theta_t^{(G)}, g_{\theta_t^{(G)}})$$

end while

Algorithm 4-2 Fast power iteration method

- Initialize $\tilde{u}_l \in \mathbb{R}^{d_l}$ for $l = 1,\ldots,L$ with a random vector (sampled from isotropic distribution).
- For each update and each layer *l*:
  1. Apply the power iteration method to an unnormalized weight W^l:

     $\tilde{v}_l \leftarrow (W^l)^T \tilde{u}_l / \left\| (W^l)^T \tilde{u}_l \right\|_2$

     $\tilde{u}_l \leftarrow (W^l) \tilde{v}_l / \left\| (W^l) \tilde{v}_l \right\|_2$
  2. Calculate $\bar{W}_{SN}^l$ with the spectral norm:

     $\bar{W}_{SN}^l(W^l) = W^l / \sigma(W^l)$, where $\sigma(W^l) = \tilde{u}_l^T W^l \tilde{v}_l$
  3. Update W^l with Adam:

     $W^l \leftarrow W^l - \alpha_t \cdot Adam(\bar{W}_{SN}^l(W^l), g_{\theta_t^{(D)}})$

4.4 Numerical Results

The performance of the proposed TAGAN-SE method is validated on the IEEE 33-node distribution system and IEEE118-bus distribution system [118].

4.4.1 Data Generation

The training data are system states (nodal voltages' magnitude) considering different system topologies with or without noises or contaminations. Fig. 4-7 shows the flowchart of data generation for numerical study via Monte Carlo(MC) simulation. For each test system, the topology and system loads are randomly sampled, and then power flow computation is executed with convergence check using MATPOWER software package [93] for N_1 times. Note that different topologies are generated per a set of switch status combinations represented by a vector containing 0 or 1, which satisfies Binomial distribution with a failure probability of 0.15 [14]. In addition, the system loads are assumed to satisfy Gaussian distribution with zero mean and standard deviation of 0.1 [14].

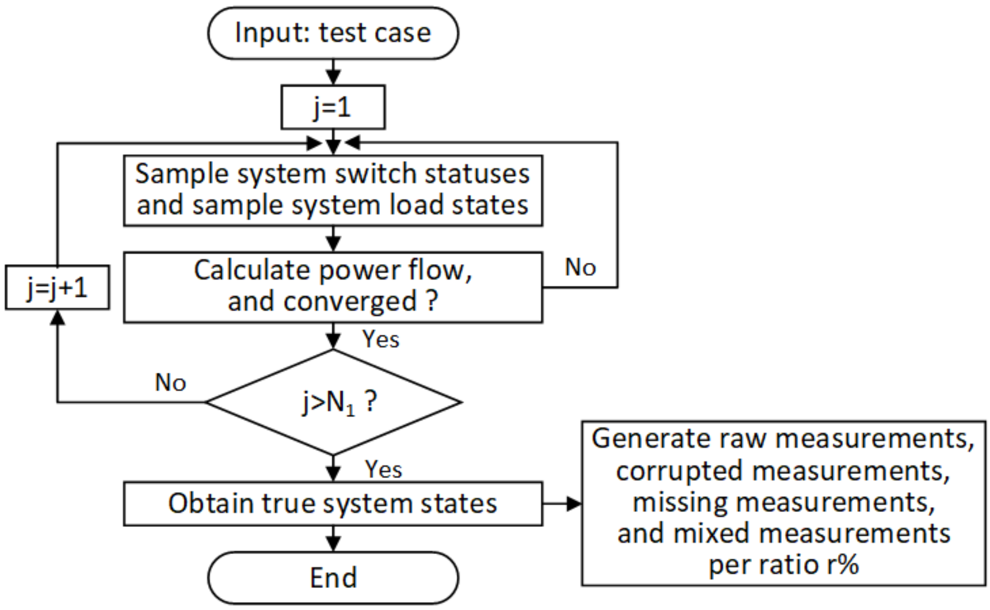


Fig. 4-7. The flowchart of data generation.

For each generated topology, after the true system states $\tilde{\mathbf{x}}$ are produced, raw measurements are generated by adding Gaussian noise with zero mean and standard deviation of 0.001 (in line with PMU's precision) [112]. In addition, abnormal measurements considering 3 types of contamination scenarios will be created, where different contamination ratio r% ranging from 0% to 50% is considered. For each measurement type, the number of training samples is 100,000.

Corrupted measurement data refers to the measurements that significantly differ from the normal raw measurement data. For each topology, the corrupted voltages are produced by randomly choosing r% raw measurements and adding an error with 0.5 mean and 0.05 standard deviation [112]. The rest (1 - r%) data are still the raw measurement with normal measurement noise.

In the case study, missing measurement data are assumed due to the time delay or interruption, data package loss, etc., during communication. Missing measurement data are produced by randomly choosing r% raw measurement and substituting by white noise with zero mean and standard deviation of 0.01.

Mixed contamination measurement data is contaminated with a mixture of bad data and missing data. To create such data, the contaminated data are equally generated per type. i.e., the mixed contamination data accounting for r % of the dataset contains (r% / 2) corrupted data and (r% / 2) missing data.

4.4.2 Training Details and Model Architecture

The case studies are carried on the two test systems. In each test system, 100 typical different topologies are considered to generate system states samples. For each topology, 1,000 samples are generated, and accordingly there are 100,000 (100*1000) samples for each test system. As mentioned before, the out-of-sample topology refers to the topology which is out of the topology library for the training data and unseen during the training, while the trained topology refers to the topology which is included in the training data. For 100 considered topologies, 20 of them belong to out-of-sample topologies, and 80 of them belong to trained topologies. In these 100,000 samples, the out-of-sample topologies' system states samples, i.e., 20,000 (20*1000) samples, are used for testing only without being trained by TAGAN, and the set of these samples is named as the test set 2. In the remaining trained topologies' system states samples, i.e., 80,000 (80*1000) samples, the training set includes 80% of them, i.e., 64,000 (80,000*80%) samples for TAGAN training, while the rest 16,000 (80,000-64,000) samples belong to the test set 1, and they are used for testing trained topologies. The batch size is 32, and the number of epochs is 200. Therefore, the training of the TAGAN takes 400,000 iterations (64000 / 32 * 200) for each contamination scenario. All programs for the TAGAN-SE model are implemented using 'TensorFlow' [120] in Python on PyCharm IDE with NVIDIA GeForce RTX 2080 Ti GPU and 11GB RAM.

In the proposed TAGAN model, the detector *T* is constructed using multilayer perceptron (MLP), which is the most widely employed neural network structure [119]. After experimenting with different hidden layers (from 1 to 5) and neurons (16, 32, 64, 128, 256), the detector *T* with single-layered MLP involving 64 neurons is found efficient to detect measurement data type accurately. The grid search method is applied to determine the optimal model setting for TAGAN via examining possible hyperparameter combinations [120]. The hyperparameters to be tuned consist of the number of layers, the number of hidden units, and the filter size. The generator *G* includes 2 fully connected MLP and 3 de-convolutional layers. The first 2 MLPs are designed for up-sampling, and the de-convolutional layers kernel size is 1×5

and strides size 2 are designed to up-sample the input noise z. Likewise, the discriminator *D* has an inversed architecture, whose 3 convolutional layers are all with a kernel size of 1×5 and stride size of 2. Table 4-1 lists the detailed settings of the TAGAN model contracted for the IEEE 33-node distribution system.

Table 4-1 The proposed TAGAN-SE model structure.

	Detector *T*	Generator *G*	Discriminator *D*
Input	33	100+66	33+66
Layer 1	MLP, 64	MLP, 1024+66	Conv, 64+66
Layer 2		MLP, 512+66	Conv, 256+66
Layer 3		Conv_transpose, 512+66	Conv, 512+66
Layer 4		Conv_transpose, 256+66	MLP, 1024+66
Layer 5		Conv_transpose, 64+66	

4.4.3 Performance Evaluation

Similarly, the performance of SE is evaluated by the conventional mean average error (MAE) for total SE error:

$$\text{MAE} = \frac{1}{N_t * m_t} \sum_{j=1}^{N_t} \sum_{i=1}^{m_t} \left| \hat{x}_j^{(i)} - \tilde{x}_j^{(i)} \right| \tag{4.21}$$

where N_t is the number of buses, m_t is the number of samples, $\hat{x}_j^{(i)}$ and $\tilde{x}_j^{(i)}$ refer to estimated system states and true system states corresponding to *i*-th sample and *j*-th bus, respectively.

A. Overall SE accuracy with different contamination ratios and types under varying topologies

Different contamination ratio $r\%$ including trained and out-of-sample topologies ranging from 10% to 50% with 20% increment is tested in this section to give a comprehensive analysis of the model accuracy and robustness. The case with raw measurements only ($r\% = 0\%$) is also examined. In addition, to guarantee the case study is unbiased, the proposed model is trained and tested 10 times for performance assessment. The MAE of TAGAN-SE on IEEE 33-node system for voltage with corrupted measurement, missing measurement, and mixed contamination measurement are illustrated in Fig. 4-8. Results from different contamination types and ratios are contrasted between the trained topology situations and the out-of-sample topology situations. Their average value of 10 runs is displayed by the green/blue bars, respectively.

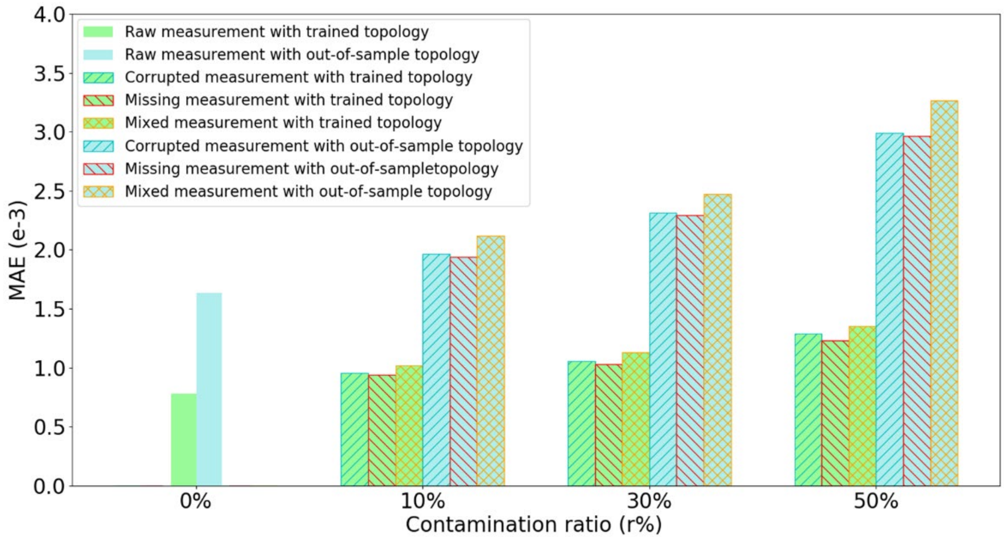


Fig. 4-8. MAE (e-3) of TAGAN-SE on the IEEE 33-node system under the trained / out-of-sample topology. The results are tested with various ratios of contaminated measurements.

From the results shown in Fig. 4-8, the MAE of voltage with out-of-sample topologies is satisfactory and reasonably larger than that with the trained topologies under all three contamination scenarios. In the situations with raw measurement only (r%=0%), the MAE of voltage with the trained topologies and out-of-sample topologies are 7.834e-4 and 1.6397e-3, respectively. They are both less than the MAE with the contaminated measurements. Besides, with the growth of contamination ratio from 0% to 50%, the MAE of both voltage with the trained topologies and the out-of-sample topologies increases correspondingly. The MAE with mixed contamination measurements is larger than that of the other two cases for both trained topologies and out-of-sample topologies scenarios. In particular, the MAE with the missing measurement is a little smaller than that with the corrupted measurement. Thus, the proposed TAGAN-SE is robust even under the scenarios with corrupted or missing data. This is because that the proposed TAGAN model is able to learn the underneath topology between buses based on the input of measurement without prior knowledge of the system model and has a strong capability of handing corrupted or missing data. Even input with some abnormal data, the model can still perform well with the underneath topology learned by analyzing the relationship between voltages of buses. The overall results of the IEEE 118-bus distribution system are presented in Table 4-3 in Section 4.4.4. It is important to point out that the missing data experiments here are preliminary to assume the system is under normal operation while the communication systems for SE cause the missing data due to time delay or data package loss, etc. In fact, missing data could also be caused by partial outages of system elements, and this should be well detected and handled differently by the proposed SE method, which is worthy of further investigations in the future.

B. Detailed Performance Assessment

The previous section only studies the overall SE performance. To study the detailed performance, the accuracy of SE on each bus is further evaluated, and the distributions between the TAGAN-SE system states and true system states are compared. Detailed assessments are studied on different topologies, and their performance is similar; thus, from both the trained topology set and the out-of-sample topology set, one topology is randomly chosen to present in IEEE 33-node test distribution system (denoted as topology 1 and topology 2) and IEEE 118-bus test distribution system (denoted as topology 3 and topology 4), where the branch numbers refer to those of the respective test systems:

- topology 1 (trained topology on the IEEE 33-node system) with switch-off branches: {13, 15, 20, 33, 37}.
- topology 2 (out-of-sample topology on IEEE 33-node system) with switch-off branches: {6, 17, 26, 33, 34}.
- topology 3 (trained topology on IEEE118-bus distribution system) with switch-off branches:

  {6,8,21,30,41,51,69,73,81,97,104,121,123,124,130}.
- topology 4 (out-of-sample topology on IEEE118-bus distribution system) with switch-off branches:

  {9,29,31,39,44,53,56,60,72,73,76,102,116,119,129}.

Fig. 4-9 and Fig. 4-10 are voltage mean and voltage standard deviation (std) of each bus under trained topology and out-of-sample topology on each test system, respectively. They are tested with 30% mixed contamination measurement. The blue dashed line denotes true system states mean, and the orange dot-dashed line denotes TAGAN-SE mean, while the light blue dashed line denotes true system states std, and the yellow dot-dashed line denotes

TAGAN-SE std. As shown in Fig. 4-9 and Fig. 4-10, for each test system, under both trained topology and out-of-sample topology, the voltage magnitudes estimated by TAGAN are very similar to the true system states and deviate slightly on the part of buses. With mixed contamination measurement input, the TAGAN-SE is robust to corrupted or missing measurement and even effective for out-of-sample topology.

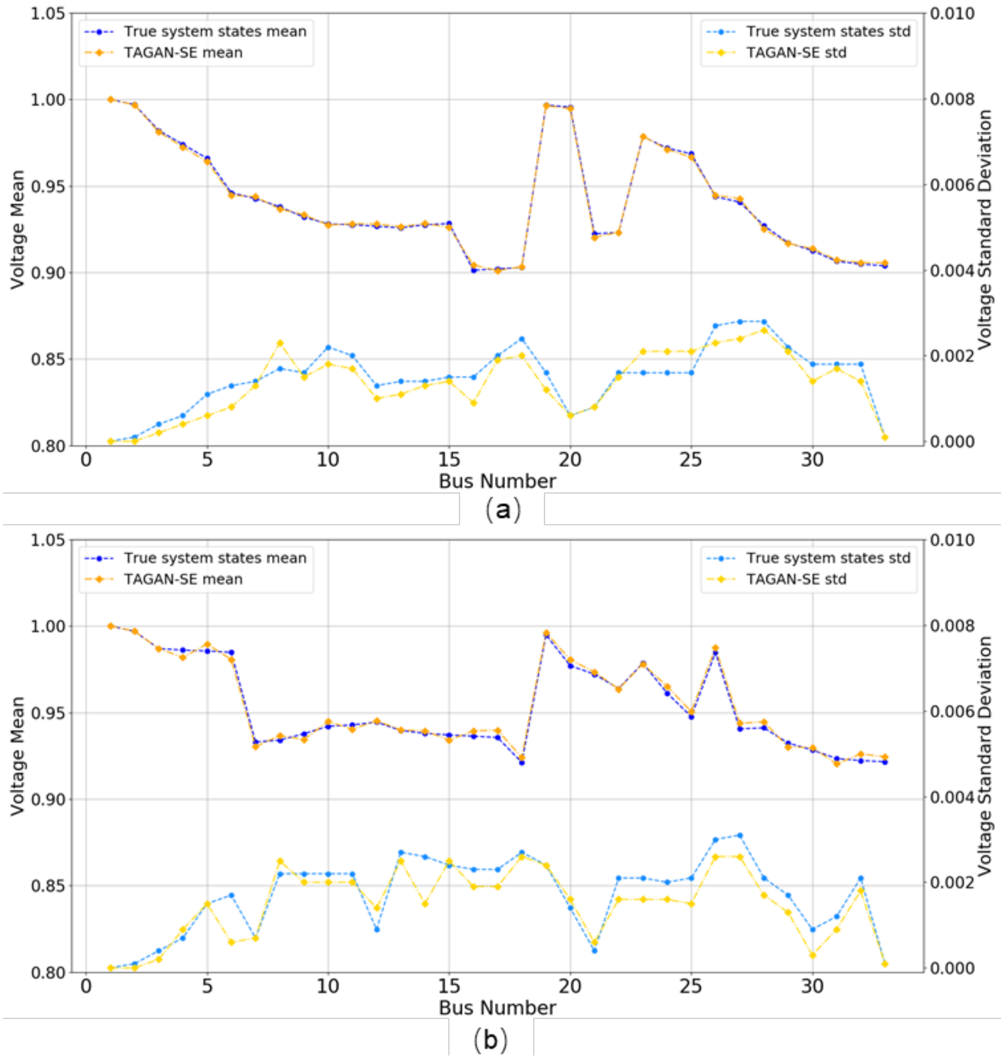


Fig. 4-9. Voltage mean and standard deviation of TAGAN-SE and true system states on the IEEE 33-node distribution system under (a) the trained topology (topology 1), and (b) the out-of-sample topology (topology 2). The result is tested with 30% mixed contamination measurement.

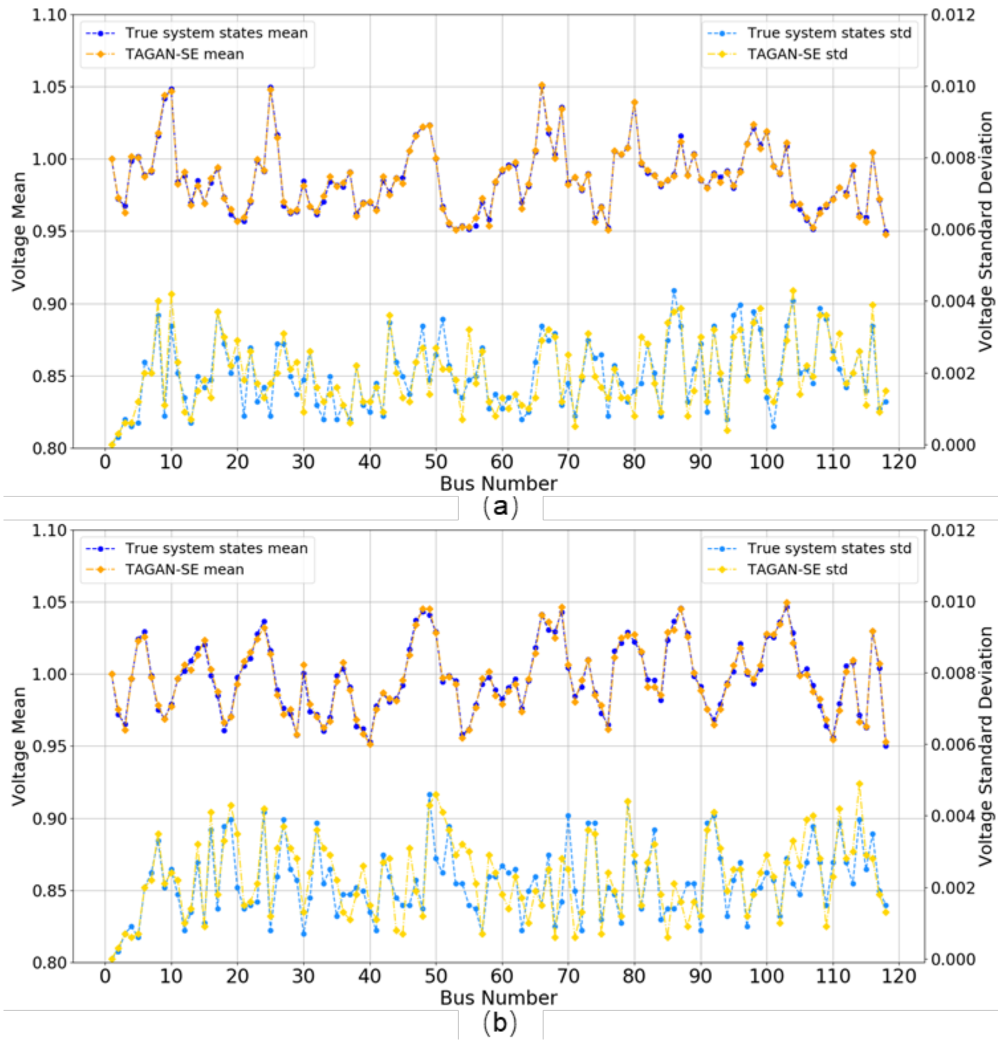


Fig. 4-10. Voltage mean and standard deviation of TAGAN-SE and true system states on the IEEE 118-bus distribution system under (a) the trained topology (topology 3), and (b) the out-of-sample topology (topology 4). The result is tested with 30% mixed contamination measurement.

To further examine the similarity between two voltage distributions, their probability density distribution profiles are compared, and bus 23 of the IEEE 33-node system is randomly chosen at topology 1 and topology 2. Fig. 4-11 describes the probability density histograms of generated and true system states, respectively at bus 23 with the trained (left subfigure) / out-of-sample topology (right subfigure). Their inputs are with three contamination types of measurement under 30% contamination ratio. In these figures, the X-axis value is the voltage magnitude (p.u.), and the Y-axis value is its probability density. The probability is denoted by the separate orthogon areas multiplied by the width of the voltage magnitude interval and its probability density. In particular, the cumulative orthogon areas are equivalent to one. Both distributions of the true system and TAGAN system states approximately follow the normal distribution. Besides, in all scenarios, the distribution of voltage with the trained topology (topology 1) is more similar to the distribution of its true system states than that with the out-of-sample topology (topology 2). This conforms to the results of previous overall assessments, indicating the proposed TAGAN-SE performs better to handle either corrupted or missing measurement than mixed of them, as the difference of distribution profiles for the former two situations is less distinct than that for the mixed situations.

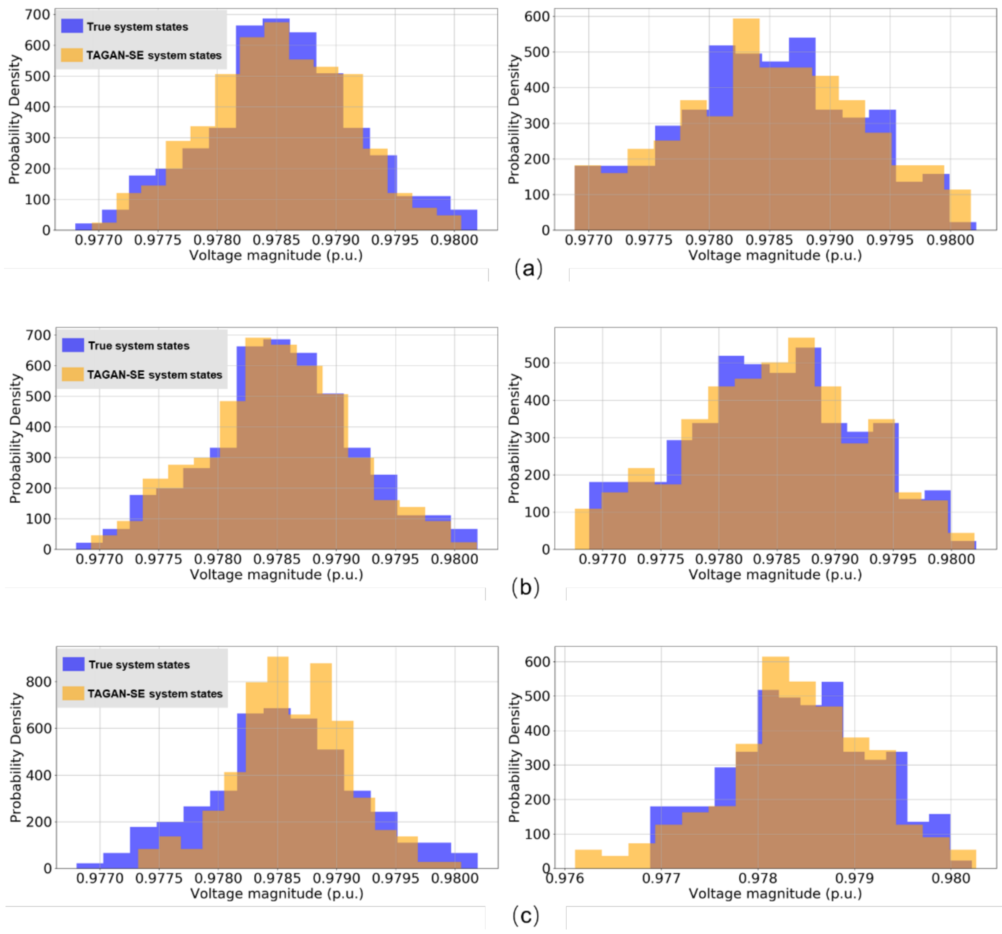


Fig. 4-11. Probability density histograms of true system states and the generated system states with the trained (left subfigure) / out-of-sample topology(right subfigure). The results are tested with (a) corrupted measurements, (b) missing measurements, and (c) mixed contamination measurements (30% contamination ratio) at bus 23 of the IEEE 33-node system.

C. Spatial Correlation Assessment

In this part, the correlation of buses for voltage is investigated further to test the quality of generated system states by TAGAN-SE. For the IEEE 33-node system, the Pearson correlation coefficients matrix colormap under the trained topology (topology 1) and the out-of-sample topology (topology 2) are depicted in Fig. 4-12 and Fig. 4-13, respectively.

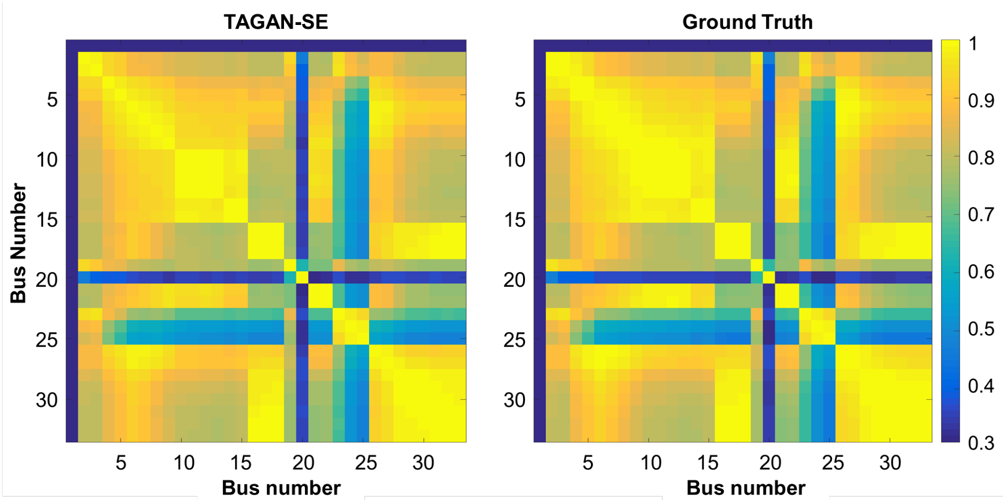


Fig. 4-12. The spatial correlation coefficients matrix colormap under the trained topology (topology 1). All results are tested with 30% ratio of mixed contamination measurement. (The right color bar is the correlation coefficient)

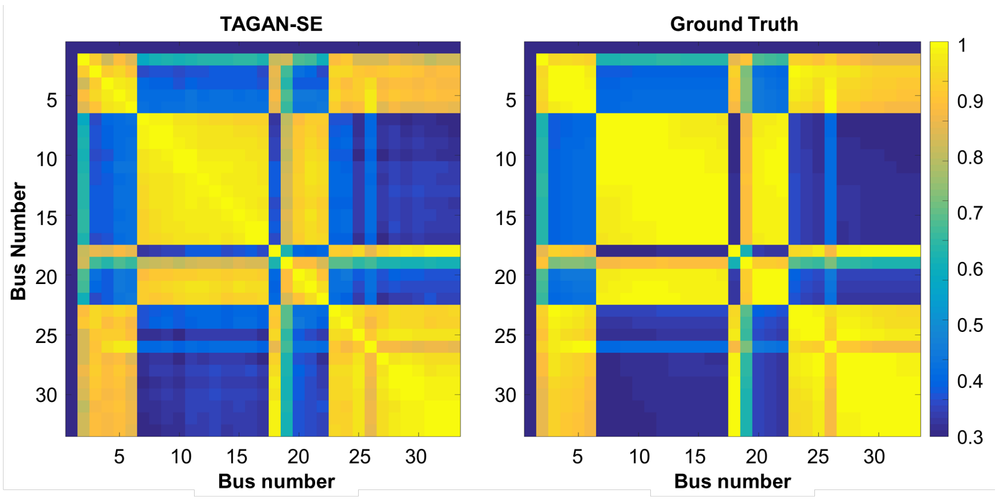


Fig. 4-13. The spatial correlation coefficients matrix colormap under the out-of-sample topology (topology 2). All results are tested with 30% ratio of mixed contamination measurement. (The right color bar is the correlation coefficient)

All results are obtained with 30% mixed contamination measurements. As seen from Fig. 4-12 and Fig. 4-13, both the correlation profiles of systems' voltage are very similar to their ground truth. Hence, TAGAN-SE can learn the spatial correlations involved in system buses well. Besides, the correlations under the trained topology (topology 1) are more similar to their ground truth than those under the out-of-sample topology (topology 2). Therefore, TAGAN-SE performs better with trained topologies, which confirms the result of previous overall assessments.

D. Loss Functions Comparison

To examine the merits of the hinge loss compared to other loss functions, the MAE of voltage is compared during the training process with the hinge loss, the cross entropy, and the Wasserstein distance in Fig. 4-14. This experiment is also performed with 30% mixed contamination measurement under out-of-sample topologies on IEEE 118-bus distribution system. As shown in Fig. 4-14, the orange line denotes the hinge loss, and it converges faster than the Wasserstein distance (blue line) and the cross entropy (red line), respectively. Besides, the hinge loss finally converges to the smaller MAE than the other two loss functions, reflecting the higher learning performance of the proposed TAGAN model through applying the hinge loss. On the other hand, the hinge loss curve is smoother than the other two loss curves, indicating that the hinge loss is more robust than the Wasserstein distance and the cross entropy.

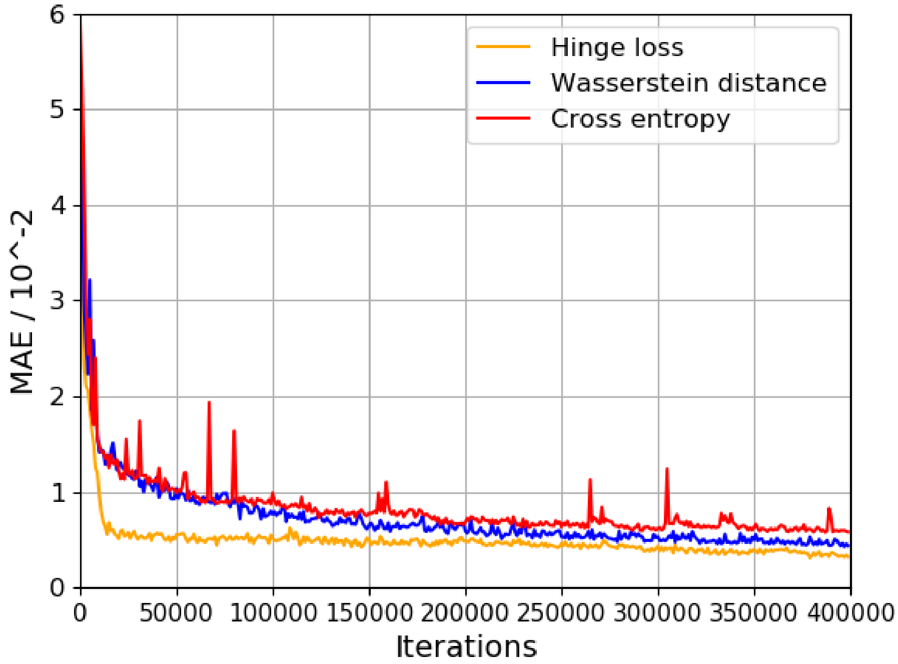


Fig. 4-14. Learning curves of the TAGAN model regarding different loss functions under the out-of-sample topology set on the IEEE 118-bus distribution system.

4.4.4 Comparison to other methods

The comparative study is carried out on the IEEE 33-node distribution system and the IEEE 118-bus distribution system by taking into account the influence of different contamination ratios. In particular, three different types of contamination measurement, including corrupted measurement, missing measurement, and mixed contamination measurement with ratios from 0% to 50% are considered.

To further validate the effectiveness of the proposed TAGAN-SE, a deep convolutional neural network-based SE (DCNN-SE) approach and another GAN-based approach, i.e., WGAN-SE, are employed. The same training and testing datasets are applied in DCNN-SE [95] and WGAN-SE [89]. The comparative results of TAGAN-SE against the benchmark DCNN-SE and WGAN-SE on IEEE 33-node and IEEE 118-bus distribution systems are listed in Table 4-2 and Table 4-3, respectively.

Table 4-2 presents the case of the IEEE 33-node distribution system where the inputs are measurements with noise only (r%=0) and measurements with different contamination types and ratios. In the same scenarios, the MAE of TAGAN-SE is smaller than the other two methods. Besides, with the increased contaminated measurements, the final MAE of TAGAN-SE increases slowly. Thus, the proposed TAGAN-SE performs better than the DCNN-SE and the WGAN-SE. The main reasons may lie in the following aspects. First of all, for DCNN-SE, there is only a single network for generating states, while TAGAN-SE consists of two networks, including a discriminator and a generator. By employing GANs in TAGAN-SE, the discriminator is able to improve the effect of the generator by giving the proper response between the generated states and real states during the training process. Compared to the WGAN-SE, the detector of TAGAN-SE improves the performance by generating a "mask" as GAN's input. Besides, the hinge loss function can be easier to compute, faster to train, and more robust than other loss functions against data contamination. Therefore, it enhances the training process and decreases the voltage MAE.

It is worth noticing that under the out-of-sample topologies for both test systems, the MAE of DCNN-SE and WGAN-SE is much larger than the same scenarios under trained topologies, while the MAE of the proposed TAGAN-SE increases less than the above two methods. Besides, the MAE of TAGAN-

SE grows slowly than DCNN-SE and WGAN-SE with the increase of contamination ratio. The reason can be attributed to that the spectrum normalization is applied in the discriminator enabling the discriminator function to be Lipschitz continuous and solve the problems (including vanishing gradients and mode collapse) in the classical GAN which improves the performance with varying topologies, especially for the out-of-sample topology. Thus, even under the cases with out-of-sample topologies, the proposed TAGAN-SE is still viable, giving a smaller MAE as compared with DCNN-SE and WGAN-SE.

In summary, the MAE of voltage for the trained topologies in all scenarios ranges in [7.8e-4, 1.3e-3] p.u and [8.7e-4, 1.9e-3] p.u, for IEEE 33-node and IEEE 118-bus distribution systems, respectively. Also, the MAE of the voltage for out-of-sample topologies in all scenarios ranges in [1.6e-3, 3.2e-3] p.u and [1.8e-3, 4.3e-3] p.u for both test systems, respectively.

Table 4-2 Performance comparison with the benchmarks on the IEEE 33-node system.

	Contamination Type	Contamination Level r%	Voltage Magnitude (p.u.)					
			Trained Topology			Out-of-sample Topology		
			DCNN-SE	WGAN-SE	TAGAN-SE	DCNN-SE	WGAN-SE	TAGAN-SE
MAE	Noise only	0%	1.418e-03	1.091e-03	7.834e-04	4.035e-03	2.320e-03	1.640e-03
	Corrupted	10%	2.433e-03	1.512e-03	9.545e-04	6.944e-03	2.991e-03	1.964e-03
		30%	2.774e-03	1.623e-03	1.054e-03	8.553e-03	3.878e-03	2.316e-03
		50%	3.590e-03	2.109e-03	1.287e-03	1.113e-02	5.624e-03	2.992e-03
	Missing	10%	2.428e-03	1.501e-03	9.421e-04	6.902e-03	2.955e-03	1.938e-03
		30%	2.751e-03	1.595e-03	1.031e-03	8.497e-03	3.836e-03	2.293e-03
		50%	3.527e-03	2.058e-03	1.232e-03	1.004e-02	5.562e-03	2.961e-03
	Mixed	10%	2.493e-03	1.575e-03	1.017e-03	7.580e-03	3.405e-03	2.119e-03
		30%	2.878e-03	1.706e-03	1.133e-03	9.246e-03	4.241e-03	2.472e-03
		50%	3.749e-03	2.265e-03	1.350e-03	1.202e-02	5.897e-03	3.263e-03

Table 4-3 Performance comparison with the benchmarks on the IEEE 118-bus system.

	Contamination Type	Contamination Level r%	Voltage Magnitude (p.u.)					
			Trained Topology			Out-of-sample Topology		
			DCNN-SE	WGAN-SE	TAGAN-SE	DCNN-SE	WGAN-SE	TAGAN-SE
MAE	Noise only	0%	1.611e-03	1.226e-03	8.725e-04	4.738e-03	2.782e-03	1.803e-03
	Corrupted	10%	2.876e-03	1.838e-03	1.169e-03	7.948e-03	3.882e-03	2.384e-03
		30%	3.351e-03	2.064e-03	1.388e-03	9.813e-03	5.053e-03	2.947e-03
		50%	4.219e-03	2.725e-03	1.751e-03	1.295e-02	7.228e-03	3.912e-03
	Missing	10%	2.861e-03	1.826e-03	1.155e-04	7.892e-03	3.846e-03	2.368e-03
		30%	3.326e-03	2.040e-03	1.372e-03	9.747e-03	5.003e-03	2.923e-03
		50%	4.280e-03	2.687e-03	1.709e-03	1.182e-02	7.174e-03	3.871e-03
	Mixed	10%	3.142e-03	2.111e-03	1.347e-03	8.780e-03	4.586e-03	2.793e-03
		30%	2.666e-03	2.462e-03	1.589e-03	1.067e-02	5.690e-03	3.219e-03
		50%	4.724e-03	3.005e-03	1.942e-03	1.399e-02	7.877e-03	4.395e-03

4.5 Summary

In this chapter, a novel data-driven and model-free TAGAN-SE approach considering varying topologies is proposed for distribution system SE. Because the spectrum normalization is applied in the discriminator, TAGAN-SE can handle topology variations. Different measurement contamination types (corrupted data, missing data, and mixed contamination data) and varying topologies, including trained topology and out-of-sample topology scenarios, were examined. Regardless of the corrupted and missing measurements under varying topologies due to, e.g., switch faults or maloperations, which can make the system even unobservable, the proposed TAGAN-SE is able to be robust to estimate the system states accurately. In addition, it can learn the statistical properties of the system states well, including the spatial correlation of buses and the probability distributions of system states. Particularly, for the out-of-sample topologies, TAGAN-SE achieves significant improvements over other data-driven based methods, e.g., DCNN-SE and WGAN-SE, while the conventional WLS-based method is inapplicable in these scenarios due to lack of the ascertained topology information. Thus, TAGAN-SE is convincingly suitable for the distribution system, of which the topology frequently changes, where many existing topologies are out of the topology library for model training. The effectiveness of the proposed TAGAN-SE is verified with the IEEE 33-node distribution system and the IEEE 118-bus distribution systems. Besides, the missing data experiments are preliminary to assume the system is under normal conditions while the communication systems for SE cause the missing data due to time delay or data package loss, etc. In fact, missing data could also be caused by partial outages in the system, and this should be well detected and handled differently by the proposed SE method, which is worthy of future research.

Chapter 5 Conclusions and Future Scope

5.1 Conclusions

The rapid growing penetration of renewable generations in smart grids exerts substantial influences on the normal operation of power systems and thus forces the system operators and planners to reconsider the mechanisms of their decision-making processes. PMU is becoming one of the most widely used advanced measuring equipment serving for real-time monitoring and control of power systems. A comprehensive study is carried out in this thesis to investigate PMUs based power system state estimation, where the optimal placement of supplementary PMUs considering the existing SCADA system as well as deep learning based SE is investigated. Specifically, the study is investigated in the following aspects, given as follows:

Firstly, a reliability-based probabilistic OPP approach is proposed in Chapter 2 to obtain minimal voltage magnitude estimation uncertainty based on various operating scenarios, with supplementary PMUs installed in the power grid, which is equipped with the SCADA system. The contribution is that the PMU measurement reliability is modeled in OPP, where state estimation uncertainty is considered. PMU measurement system components' random outages are considered in the modeling of PMU measurement reliability. Besides, the framework applies the PLF in OPP to represent different operating scenarios. In this way, the load patterns and power generators' on/off status are considered stochastically as the operating uncertainties, so that the obtained PMU placement solution is unbiased for planning purposes. With PLF carried out, the OPP scheme is more suitable for different operating scenarios because different operating uncertainties are considered.

Then, a model-free and data-driven deep learning method is proposed in Chapter 3 for SE in transmission systems. This method is based on conditional WGAN, where the Wasserstein distance is applied to improve training performance. With the corrupted or missing measurement at different ratios, the proposed method can perform better than the traditional and state-of-the-art methods, i.e., LWLS-SE, WLAV-SE, and DCNN-SE. The proposed method CGAN-SE not only can estimate the system states with high accuracy, but can also capture the statistical properties of the system measurements either from the probability distribution of system states and the spatial correlation of buses. Moreover, the proposed SE method can still be effective even in an unobservable network. The experiments validate the effectiveness of the proposed method on a large system (2746-bus Polish system), and this is the first experiment using deep learning models for power system SE processes on a large-scale system.

Finally, a topology-aware data-driven, model-free approach for distribution system SE is proposed in Chapter 4. By applying TAGAN in the distribution system, the actual correlations of system states can be well learned, and the true states of the system can be estimated precisely by only applying one model, which considers the various topologies rather than decoupling the SE process into topology identification and state estimation. Regardless of the corrupted and missing measurements under varying topologies due to, e.g., switch faults or maloperations, the proposed method TCGAN-SE is robust and can estimate the system states accurately. In particular, for the out-of-sample topologies, where the conventional WLS-based method is inapplicable without an ascertained topology, TAGAN-SE performs much better than other data-driven based methods, e.g., DCNN-SE and WGAN-SE. Therefore, this method well fits the SE of distribution systems since the topology of the distribution system changes frequently, and there is countless number of scenarios.

5.2 Future Scope

This study is preliminary for the proposed SE method to handle the missing data situations. Besides, the missing data problem is also tangled with the possible situation of partial system outages, where the missing input data to SE models could be attributed to the actual outages of system elements. Under such situations, it is very challenging to distinguish between the fundamental reasons for missing data due to either actual data missing during the data transportation/communication or the system outages. This issue is worthy of further investigation in the future.

Besides, in my current work, all test data are from simulation. Although these proposed deep learning-based SE methods perform well in the SE problem, it is hard to evaluate whether they can outperform existing methods in real power grids. Thus, in the future, the deep learning-based SE methods are expected to be tested by using actual system's data and apply it into real applications.

With the development of the performance of computer processors (CPUs & GPUs) and larger storage media for retaining huge training datasets, deep learning methods have begun to outperform traditional algorithms in many areas. There is a trend that researchers try to better handle on how deep learning technologies can be integrated into existing and new critical infrastructures while safeguarding and representing important values such as safety, sustainability, and equity. In this thesis, deep learning technologies are introduced to handle state estimation in power grids. Besides, the proposed SE methods present great advantages, especially in aspects of unobservability and topology issues. As the power grid is still undergoing transitions and a large amount of data such as historical power system operation records, future renewables generation, and load forecasts are providing us more information and insights of the underlying grids, a critical challenge in distribution grid operation and control is that the grid itself may be unobservable and fast-changing. Based on the achievements of distribution system state estimation in this thesis, topology recovery and advanced control algorithms may be further developed by deep learning technologies to improve distribution grid observability and controllability.

References

[1] IEEE Standard for Synchrophasors for Power Systems, *IEEE Standard* C37.118-2005, Jun. 2005.

[2] J. Bertsch, C. Carnal, D. Karlson, J. McDaniel, and V. Khoi, "Wide-Area Protection and Power System Utilization," *Proceedings of the IEEE,* vol. 93, no. 5, pp. 997-1003, 2005.

[3] Z. Y. Dong, Y. Xu, P. Zhang, and K. P. Wong, "Using IS to Assess an Electric Power System's Real-Time Stability," *IEEE Intelligent Systems,* vol. 28, no. 4, pp. 60-66, 2013.

[4] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in power,* vol. 6, no. 2, pp. 10-15, 1993.

[5] J. S. Thorp, A. G. Phadke, and K. J. Karimi, "Real time voltage-phasor measurements for static state estimation," *IEEE Trans. Power App. Sys.*, vol. PAS-104, no. 11, pp. 3098–3106, Nov. 1985.

[6] A. G. Phadke, J. S. Thorp, and K. J. Karimi, "State estimation with phasor measurements," *IEEE Trans. Power Syst.*, vol. 1, no. 1, pp. 233–241, Feb. 1986.

[7] A. G. Phadke and J. S. Thorp, Synchronized Phasor Measurements and Their Applications. *New York: Springer*, 2008.

[8] B. Gou, "Optimal placement of PMUs by integer linear programming," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1525-1526, Aug. 2008.

[9] R. Kavasseri and S. K. Sirnivasan, "Joint placement of phasor and power flow measurements for observability of power systems," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1929-1936, Nov. 2011.

[10] Enshaee, A., Hooshmand, R.A., and Fesharaki, F.H., 'A New Method for Optimal Placement of Phasor Measurement Units to Maintain Full Network Observability under Various Contingencies', *Electric Power Systems Research*, 2012, 89, pp. 1-10.

[11] Kekatos, V., Giannakis, G.B., and Wollenberg, B., 'Optimal Placement of Phasor Measurement Units Via Convex Relaxation', *IEEE Transactions on Power Systems*, 2012, 27, (3), pp. 1521-1530.

[12] Huang, L., Sun, Y., Xu, J., Gao, W., Zhang, J., and Wu, Z., 'Optimal Pmu Placement Considering Controlled Islanding of Power System', *IEEE Transactions on Power Systems*, 2013, 29, (2), pp. 742-755.

[13] Qi, J., Sun, K., and Kang, W., 'Optimal Pmu Placement for Power System Dynamic State Estimation by Using Empirical Observability Gramian', *IEEE Transactions on Power Systems*, 2014, 30, (4), pp. 2041-2054.
[14] Zhang, C., Jia, Y., Xu, Z., Lai, L.L., and Wong, K.P., 'Optimal Pmu Placement Considering State Estimation Uncertainty and Voltage Controllability', *IET Generation, Transmission & Distribution*, 2017, 11, (18), pp. 4465-4475.
[15] A.Monticelli, *State estimation in electric power systems: a generalized approach*, vol. 507. Springer Science & BusinessMedia, 1999.
[16] A. Abur and A. Exposito, *Power system state estimation: theory and implementation*, vol. 24. CRC, 2004.
[17] Zhu, J. and Abur, A., 'Identification of Network Parameter Errors Using Phasor Measurements', in, *2009 IEEE Power & Energy Society General Meeting*, (IEEE, 2009)
[18] Meneghetti, Rogério, Antonio Simões Costa, Vladimiro Miranda, and Larah Brüning Ascari. "Information Theoretic Generalized State Estimation in power systems." *Electric Power Systems Research* 182 (2020): 106251.
[19] R. Kavasseri and S. K. Srinivasan, "Joint placement of phasor and conventional power flow measurements for fault observability of power systems," *IET Generat., Transmiss. Distrib.*, vol. 5, no. 10, pp. 1019-1024, Oct. 2011.
[20] N. M. Manousakis and G. N. Korres, "A weighted least squares algorithm for optimal PMU placement," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3499-3500, Aug. 2013.
[21] N. C. Koutsoukis, N. M. Manousakis, P. S. Georgilakis, and G. N. Korres, "Numerical observability method for optimal phasor measurement units placement using recursive Tabu search method," *IET Generat., Transmiss. Distrib.*, vol. 7, no. 4, pp. 347-356, Apr. 2013.
[22] A. Garcia, A.Monticelli, and P. Abreu, "Fast decoupled state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS- 98, pp. 1645–1652, Sept 1979.
[23] Mili, L., Cheniae, M.G., and Rousseeuw, P.J., 'Robust State Estimation of Electric Power Systems', *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 1994, 41, (5), pp. 349-358.
[24] Aminifar, F., Fotuhi-Firuzabad, M., Shahidehpour, M., and Khodaei, A., 'Observability Enhancement by Optimal Pmu Placement Considering Random Power

System Outages', *Energy Systems*, 2011, 2, (1), pp. 45-65.

[25] Monticelli, A. and Wu, F.F., 'Network Observability: Identification of Observable Islands and Measurement Placement', *IEEE Transactions on Power Apparatus and systems*, 1985, PAS-104, (5), pp. 1035-1041.

[26] Massignan, J.A.D., London, J.B.A., Maciel, C.D., Bessani, M., and Miranda, V., 'Pmus and Scada Measurements in Power System State Estimation through Bayesian Inference', in, *2019 IEEE Milan PowerTech*, (2019)

[27] Zhao, J., Zhang, G., Das, K., Korres, G.N., Manousakis, N.M., Sinha, A.K., and He, Z., 'Power System Real-Time Monitoring by Using Pmu-Based Robust State Estimation Method', *IEEE Transactions on Smart Grid*, 2016, 7, (1), pp. 300-309.

[28] Mestav, K.R., Luengo-Rozas, J., and Tong, L., 'Bayesian State Estimation for Unobservable Distribution Systems Via Deep Learning', *IEEE Transactions on Power Systems*, 2019, 34, (6), pp. 4910-4920.

[29] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid,* vol. 10, no. 2, pp. 2312-2322, 2018.

[30] R. Meneghetti, A. S. Costa, V. Miranda, and L. B. Ascari, "Information Theoretic Generalized State Estimation in power systems," *Electric Power Systems Research,* vol. 182, p. 106251, 2020.

[31] G. N. Korres and N. M. Manousakis, "A state estimation algorithm for monitoring topology changes in distribution systems," in *Proc. PES General Meeting*, San Diego, CA, USA, 2012.

[32] D. V. Kumar, S. Srivastava, S. Shah, and S. Mathur, "Topology processing and static state estimation using artificial neural networks," *IEE Proceedings-Generation, Transmission and Distribution,* vol. 143, no. 1, pp. 99-105, 1996.

[33] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 12-19, Jan. 2009.

[34] B. Milosevic and M. Begovic, "Nondominated sorting genetic algorithm for optimal phasor measurement placement," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 69-75, Feb. 2003.

[35] F. Aminifar, A. Khodaei, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Contingency-constrained PMU placement in power networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 516-523, Feb. 2010.

[36] N. H. Abbasy and H. M. Ismail, “A unified approach for the optimal PMU location for power system state estimation,” *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 806-813, May 2009.
[37] M. Esmaili, K. Gharani, and H. A. Shayanfar, “Redundant observability PMU placement in the presence of flow measurements considering contingencies,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3765-3773, Nov. 2013.
[38] R. Emami and A. Abur, “Robust measurement design by placing synchronized phasor measurements on network branches,” *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 38-43, Feb. 2010.
[39] S. Chakrabarti and E. Kyriakides, “Optimal placement of phasor measurement units for power system observability,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1433-1440, Aug. 2008.
[40] S. M. Mazhari, H. Monsef, H. Lesani, and A. Fereidunian, “A multi-objective PMU placement method considering measurement redundancy and observability value under contingencies,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2136-2146, Aug. 2013.
[41] C. Rakpenthai, S. Premrudeepreechacharn, S. Uatrongjit, and N. R. Watson, “An optimal PMU placement method against measurement loss and branch outage,” *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 101-107, Jan. 2007.
[42] S. Azizi, A. S. Dobakhshari, S. A. Nezam Sarmadi, and A. M. Ranjbar, “Optimal PMU placement by an equivalent linear formulation for exhaustive search,” *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 174-182, Mar. 2012.
[43] L. Zhang and A. Abur, “Strategic placement of phasor measurements for parameter error identification,” *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 393-400, Feb. 2013.
[44] K.-P. Lein, C.-W. Liu, C.-S. Yu, and J.-A. Jiang, “Transmission network fault location observability with minimal PMU placement,” *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1128-1136, Jul. 2006.
[45] J. Chen and A. Abur, “Enhanced topology error processing via optimal measurement design,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 845-852, Aug. 2008.
[46] K. G. Firouzjah, A. Sheikholeslami, and T. Barforoushi, “Reliability improvement of power system observability with minimum phasor measurement units,” *IET Generat., Transmiss. Distrib.*, vol. 7, no. 2, pp. 118-129, Feb. 2013.
[47] B. Gou and R. G. Kavasseri, “Unified PMU placement for observability and bad

data detection in state estimation," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 25732580, Nov. 2014.

[48] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1608-1615, Nov. 2006.

[49] A. Chakrabortty and C. F. Martin, "Optimal measurement allocation algorithms for parametric model identification of power systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 5, pp. 1801-1802, Sep. 2014.

[50] E. Caro, R. Singh, B. C. Pal, A. J. Conejo, and R. A. Jabr, "Participation factor approach for phasor measurement unit placement in power system state estimation," *IET Generat., Transmiss. Distrib.*, vol. 6, no. 9, pp. 922-929, Sep. 2012.

[51] L. Zhang and A. Abur, "Single and double edge cutset identification in large scale power networks," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 510-516, Feb. 2012.

[52] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 911–916, May 1996.

[53] M. Baran and A. Kelley, "A branch-current-based state estimation method for distribution systems," *IEEE Trans. Power Syst.*, vol. 10, no. 1, pp. 483–491, Feb. 1995.

[54] W.-M. Lin, J.-H. Teng, and S.-J. Chen, "A highly efficient algorithm in treating current measurements for the branch-current-based distribution state estimation," *IEEE Trans. Power Del.*, vol. 16, no. 3, pp. 433–439, Jul. 2001.

[55] R. Hoffman, "Practical state estimation for electric distribution networks," in *Proc. IEEE PES PSCE*, Nov. 2006, pp. 510–517.

[56] J.-H. Teng, "Using voltage measurements to improve the results of branch-current-based state estimators for distribution systems," *IEE Proc. Generat., Transmiss. Distrib.*, vol. 149, no. 6, pp. 667–672, Nov. 2002.

[57] M. Baran, J. Jung, and T. McDermott, "Including voltage measurements in branch current state estimation for distribution systems," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2009, pp. 1–5.

[58] H. Wang and N. Schulz, "A revised branch current-based distribution system state estimation algorithm and meter placement impact," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 207–213, Feb. 2004.

[59] G. D'Antona, C. Muscas, and S. Sulis, "State estimation for the localization of harmonic sources in electric distribution systems," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1462–1470, May 2009.

[60] Muscas, C., Pau, M., Pegoraro, P.A*., et al.*: ‘Uncertainty of voltage profile in PMU-based distribution system state estimation’, *IEEE Trans. Instrum. Meas.*, 2016, 65, (5), pp. 988–998.
[61] Gómez-Expósito, A., de la Villa Jaén, A., Gómez-Quiles, C., Rousseaux, P., and Van Cutsem, T., 'A Taxonomy of Multi-Area State Estimation Methods', *Electric Power Systems Research*, 2011, 81, (4), pp. 1060-1069.
[62] Kekatos, V. and Giannakis, G.B., 'Distributed Robust Power System State Estimation', *IEEE Transactions on Power Systems*, 2012, 28, (2), pp. 1617-1626.
[63] Zhang, L., Wang, G., and Giannakis, G.B., 'Real-Time Power System State Estimation and Forecasting Via Deep Unrolled Neural Networks', *IEEE Transactions on Signal Processing*, 2019, 67, (15), pp. 4069-4077.
[64] Zanni, L., Power-System State Estimation Based on Pmus', (EPFL, 2017).
[65] Huang, Y.-F., Werner, S., Huang, J., Kashyap, N., and Gupta, V., 'State Estimation in Electric Power Grids: Meeting New Challenges Presented by the Requirements of the Future Grid', *IEEE Signal Processing Magazine*, 2012, 29, (5), pp. 33-43.
[66] Korres, G.N. and Manousakis, N.M., 'State Estimation and Bad Data Processing for Systems Including Pmu and Scada Measurements', *Electric Power Systems Research*, 2011, 81, (7), pp. 1514-1524.
[67] Hui, X., Qing-quan, J., Ning, W., Zhi-qian, B., Hai-tang, W., and Hong-xia, M., 'A Dynamic State Estimation Method with Pmu and Scada Measurement for Power Systems', in, *2007 International Power Engineering Conference (IPEC 2007)*, (2007).
[68] Fang, C., Xueshan, H., Zhiyuan, P., and Li, H., 'State Estimation Model and Algorithm Including Pmu', in, *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, (2008).
[69] Chakrabarti, S., Kyriakides, E., Ledwich, G., and Ghosh, A., 'Inclusion of Pmu Current Phasor Measurements in a Power System State Estimator', *IET Generation, Transmission & Distribution*, 2010, 4, (10), pp. 1104-1115.
[70] Costa, A.S., Albuquerque, A., and Bez, D., 'An Estimation Fusion Method for Including Phasor Measurements into Power System Real-Time Modeling', *IEEE Transactions on Power Systems*, 2013, 28, (2), pp. 1910-1920.
[71] Tavares, B., Freitas, V., Miranda, V., and Costa, A.S., 'Merging Conventional and Phasor Measurements in State Estimation: A Multi-Criteria Perspective', in, *2017 19th International Conference on Intelligent System Application to Power Systems (ISAP)*,

(2017).

[72] Mahmood, F.: 'Synchrophasor Based Steady State Model Synthesis of Active Distribution Networks', KTH Royal Institute of Technology, 2018.

[73] Jones, K.D., Pal, A., and Thorp, J.S., 'Methodology for Performing Synchrophasor Data Conditioning and Validation', *IEEE Transactions on Power Systems*, 2015, 30, (3), pp. 1121-1130.

[74] Tate, J.E. and Overbye, T.J., 'Extracting Steady State Values from Phasor Measurement Unit Data Using Fir and Median Filters', in, *2009 IEEE/PES Power Systems Conference and Exposition*, (IEEE, 2009).

[75] Zhou, M., Centeno, V.A., Thorp, J.S., and Phadke, A.G., 'An Alternative for Including Phasor Measurements in State Estimators', *IEEE Transactions on Power Systems*, 2006, 21, (4), pp. 1930-1937.

[76] Sarri, S., Zanni, L., Popovic, M., Boudec, J.L., and Paolone, M., 'Performance Assessment of Linear State Estimators Using Synchrophasor Measurements', *IEEE Transactions on Instrumentation and Measurement*, 2016, 65, (3), pp. 535-548.

[77] Zhu, J. and Abur, A., 'Effect of Phasor Measurements on the Choice of Reference Bus for State Estimation', in, *2007 IEEE Power Engineering Society General Meeting*, (IEEE, 2007).

[78] Xu, C. and Abur, A., 'A Fast and Robust Linear State Estimator for Very Large Scale Interconnected Power Grids', *IEEE Transactions on Smart Grid*, 2017, 9, (5), pp. 4975-4982.

[79] Göl, M. and Abur, A., 'Lav Based Robust State Estimation for Systems Measured by Pmus', *IEEE Transactions on Smart Grid*, 2014, 5, (4), pp. 1808-1814.

[80] Göl, M. and Abur, A., 'A Hybrid State Estimator for Systems with Limited Number of Pmus', *IEEE Transactions on Power Systems*, 2015, 30, (3), pp. 1511-1517.

[81] Kotiuga, W.W. and Vidyasagar, M., 'Bad Data Rejection Properties of Weighted Least Absolute Value Techniques Applied to Static State Estimation', *IEEE Transactions on Power Apparatus and systems*, 1982, (4), pp. 844-853.

[82] Mahaei, S.M. and Navayi, M.R., 'Power System State Estimation with Weighted Linear Least Square', *International Journal of Electrical & Computer Engineering (2088-8708)*, 2014, 4, (2).

[83] Rousseeuw, P.J., 'Least Median of Squares Regression', *Journal of the American statistical association*, 1984, 79, (388), pp. 871-880.

[84] Huber, P.J. and Ronchetti, E.M., 'Robust Statistics. 2009', *Hoboken: John Wiley*

& Sons, 2.
[85] Chen, Y., Wang, Y., Kirschen, D., and Zhang, B., 'Model-Free Renewable Scenario Generation Using Generative Adversarial Networks', *IEEE Transactions on Power Systems*, 2018, 33, (3), pp. 3265-3275.
[86] Nath, R.P., and Balaji, V.N., 'Artificial Intelligence in Power Systems', *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN*, 2014, pp. 2278-0661.
[87] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 'Generative Adversarial Nets', in, *Advances in neural information processing systems*, (2014).
[88] He, Y., Chai, S., and Xu, Z., 'A Novel Approach for State Estimation Using Generative Adversarial Network', in, *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, (IEEE, 2019).
[89] Arjovsky, M., Chintala, S., and Bottou, L., 'Wasserstein Gan', *arXiv preprint arXiv:1701.07875*, 2017.
[90] Villani, C., *Optimal Transport: Old and New*, (Springer Science & Business Media, 2008).
[91] Mirza, M. and Osindero, S., 'Conditional Generative Adversarial Nets', *arXiv preprint arXiv:1411.1784*, 2014.
[92] Tieleman, T. and Hinton, G., 'Lecture 6.5-Rmsprop: Divide the Gradient by a Running Average of Its Recent Magnitude', *COURSERA: Neural networks for machine learning*, 2012, 4, (2), pp. 26-31.
[93] Zimmerman, R.D., Murillo-Sánchez, C.E., and Thomas, R.J., 'Matpower: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education', *IEEE Transactions on Power Systems*, 2011, 26, (1), pp. 12-19.
[94] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., and Devin, M., 'Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems', *arXiv preprint arXiv:1603.04467*, 2016.
[95] Krizhevsky, A., Sutskever, I., and Hinton, G.E., 'Imagenet Classification with Deep Convolutional Neural Networks', in, *Advances in neural information processing systems*, (2012).
[96] J. Fan and S. Borlase, “The evolution of distribution,” IEEE Power Energy Mag., vol. 7, no. 2, pp. 63–68, Mar. 2009.
[97] A. P. S. Meliopoulos, E. Polymeneas, Z. Tan, R. Huang, and D. Zhao, “Advanced distribution management system,” *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2109–

2117, Dec. 2013.

[98] A. S. Zamzam and N. D. Sidiropoulos, “Physics-aware neural networks for distribution system state estimation,” IEEE Trans. Power Syst., vol. 35, no. 6, pp. 4347–4356, Nov. 2020.

[99] E. Caro, A. J. Conejo, and A. Abur, "Breaker status identification," *IEEE Transactions on Power Systems,* vol. 25, no. 2, pp. 694-702, 2009.

[100] Y. C. Chen, T. Banerjee, A. D. Dom´ınguez-Garc´ıa, and V. V. Veeravalli, “Quickest line outage detection and identification,” *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 749–758, Feb. 2016.

[101] S. Sihag and A. Tajer, "Power system state estimation under model uncertainty," *IEEE Journal of Selected Topics in Signal Processing,* vol. 12, no. 4, pp. 593-606, 2018.

[102] O. Alsac, N. Vempati, B. Stott, and A. Monticelli, "Generalized state estimation," *IEEE Transactions on power systems,* vol. 13, no. 3, pp. 1069-1075, 1998.

[103] E. M. Lourenço, A. S. Costa, and K. A. Clements, "Bayesian-based hypothesis testing for topology error identification in generalized state estimation," *IEEE Transactions on power systems,* vol. 19, no. 2, pp. 1206-1215, 2004.

[104] V. Freitas and A. Simoes Costa, “Integrated state & topology estimation based on a priori topology information,” in *Proc. IEEE PowerTech*, Eindhoven, the Netherlands, Jun. 2015, pp. 1–6.

[105] N. Da Silva, A. S. Costa, K. Clements, and E. Andreoli, "Simultaneous estimation of state variables and network topology for power system real-time modeling," *Electric Power Systems Research,* vol. 133, pp. 338-346, 2016.

[106] V. Miranda, J. Krstulovic, H. Keko, C. Moreira, and J. Pereira, "Reconstructing missing data in state estimation with autoencoders," *IEEE Transactions on power systems,* vol. 27, no. 2, pp. 604-611, 2011.

[107] Dinesh, *Great Learning Team*, https://www.mygreatlearning.com/blog/the-vanishing-gradient-problem/.

[108] R. Singh, E. Manitsas, B. C. Pal, and G. Strbac, "A recursive Bayesian approach for identification of network configuration changes in distribution system state estimation," *IEEE Transactions on Power Systems,* vol. 25, no. 3, pp. 1329-1336, 2010.

[109] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Transactions on Smart Grid,* vol. 6, no. 4, pp. 1964-1971, 2015.

[110] D. Singh, J. Pandey, and D. Chauhan, "Topology identification, bad data processing, and state estimation using fuzzy pattern matching," *IEEE Transactions on power systems,* vol. 20, no. 3, pp. 1570-1579, 2005.

[111] Y. He, S. Chai, Z. Xu, C. S. Lai, and X. Xu, "Power system state estimation using conditional generative adversarial network," *IET Generation, Transmission & Distribution,* vol. 14, no. 24, pp. 5823-5833, 2020.

[112] M. E. Baran and A. W. Kelley, "State estimation for real-time monitoring of distribution systems," *IEEE Trans. Power Syst.*, vol. 9, pp. 1601–1609, Aug. 1994.

[113] C. Carquex, C. Rosenberg, and K. Bhattacharya, "State estimation in power distribution systems based on ensemble Kalman filtering," *IEEE Trans. Power Syst.,* vol. 33, no. 6, pp. 6600–6610, Nov. 2018.

[114] W. Zhang, et al. "A precise-mask-based method for enhanced image inpainting." *Mathematical Problems in Engineering,* 2016.

[115] J. H. Lim and J. C. Ye, "Geometric GAN," *arXiv*:1705.02894, 2017.

[116] I. Kavalerov, et al. "A multi-class hinge loss for conditional gans." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1290-1299. 2021.

[117] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–26.

[118] D. Zhang, Z. Fu, and L. Zhang, "An improved TS algorithm for loss-minimum reconfiguration in large-scale distribution systems," *Electr Pow Syst Res*, vol. 77, no. 5-6, pp. 685-694, 2007.

[119] A. K. Palit and D. Popović, Computational Intelligence in Time Series Forecasting. *New York: Springer*, 2005.

[120] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Beijing, China: *O'Reilly Media, Inc.*, 2017.

[121] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1398–1402.

[122] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.