



THE HONG KONG  
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

---

## Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

**By reading and using the thesis, the reader understands and agrees to the following terms:**

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

### IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact [lbsys@polyu.edu.hk](mailto:lbsys@polyu.edu.hk) providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

**GRAPH ORIENTED TOPOLOGY  
IDENTIFICATION AND STATE  
MONITORING IN DISTRIBUTION  
SYSTEM**

**HUAYI WU**

**PhD**

**The Hong Kong Polytechnic University**

**2022**

The Hong Kong Polytechnic University

Department of Electrical Engineering

**GRAPH ORIENTED TOPOLOGY  
IDENTIFICATION AND STATE MONITORING IN  
DISTRIBUTION SYSTEM**

**HUAYI WU**

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

July 2022

# 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 acknowledgement has been made in the text.

\_\_\_\_\_ (Signed)

    Huayi WU     (Name of student)

# *Abstract*

In the past decades, the power system is integrated with increasing renewable energy resources (RES) to combat climate change and mitigate the energy crisis. The uncertainty and intermittency of RESs cause significant impacts on the distribution system operation. This change in the distribution system composition spurs the requirement of timely monitoring of the distribution system states. Different from the transmission system with redundant monitoring, the distribution system lacks sufficient measurements. Another distinguished difference is the frequent changing topology. Thus, there remains a gap in the distribution system topology identification and state estimation. In addition, the gap also exists in probabilistic power flow for the system planning. Therefore, this thesis focuses on addressing developing progressive state monitoring approaches based on graph-oriented artificial intelligence technologies.

To timely identify the distribution grid topology, a power distribution grid topological generative adversarial network (Gridtopo-GAN) model is proposed for the topology identification considering the challenging situations of limited measurements and meshed structure. Specifically, an innovative topology preserved node embedding architecture is introduced to represent and compress the numerous topologies such that the topology identification of large-scale systems with varying topologies can be handled. The bad measurement data and missing data inspire the application of the GAN with the generative capability to render the robustness to the proposed topology identification model. Numerical simulations represent the effectiveness and time saving of the proposed model.

To timely track the states of distribution systems with high penetration of RES, an unrolled spatiotemporal graph convolutional network (USGCN) is developed for distribution system state estimation and forecasting that is exposed to complex correlations

among the renewable power outputs. Specifically, the proposed unrolled spatiotemporal graph model can capture the spatiotemporal correlations simultaneously to obtain ameliorated forecasting accuracy. Then, the node-embedding is proposed to represent the hidden spatiotemporal correlations so that automatically learning the correlations and distribution system parameters can be achieved. Furthermore, the multiple stacking spatiotemporal convolutional layers can achieve the ahead-of-time system states. The simulation results verify the accuracy and efficiency of the proposed model.

To represent the uncertain distribution system states quantificationally, a graph-aware deep learning network (GADLN) is proposed. To fully capture the mapping from the fluctuated power injections and the uncertain system states, the convolutional operation is introduced to aggregate the correlations among renewable power outputs to facilitate the PPF. In this way, improved effectiveness and speed-up calculation can be achieved in the proposed model. Moreover, the numerical results show the superior of the GADLN over the state-of-art with accurate and effective manners.

To calculate the probabilistic power flow (PPF) with complex correlations, a graph attention enabled convolutional network (GAECN) is proposed to approximate PPF. Specifically, the graph attention enabled convolutional layer is proposed to aggregate the correlations of the power injections during the training process. Within this layer, the full self-adaptive graph convolutional operation is proposed to capture and learn any implicit correlation automatically so that significantly enhanced accuracy can be achieved. The improved accuracy and efficiency achieved by the proposed model are indicated by the simulation results.

# *Publications Arisen from the Thesis*

## **Refereed Journal Publications**

- 1 **Huayi Wu**, Zhao Xu, Jian Zhao, and Songjian Chai, “Gridtopo-GAN for Distribution System Topology Identification”, *IEEE Transactions on Industrial Informatics*, 2022. (Accepted.)
- 2 **Huayi Wu**, Minghao Wang, Zhao Xu, and Youwei Jia, “Graph Attention Enabled Convolutional Network for Distribution System Probabilistic Power Flow”, *IEEE Transactions on Industry Applications*, 2022. (Accepted.)

## **Submitted Journal papers for review**

- 3 **Huayi Wu**, Zhao Xu, and Minghao Wang, “Unrolled Spatiotemporal Graph Convolutional Network for Distribution System State Estimation and Forecasting”, *IEEE Transactions on Sustainable Energy*. (Revised.)
- 4 **Huayi Wu**, Zhao Xu, Youwei Jia, and Xu Xu, “Adaptive Distributed Graph Model for Multiple Line Outage Identification in Large-Scale Power System”, *IEEE Systems Journal*, 2022. (Revised.)

## **Refereed Conference Publications**

- 5 **Huayi Wu**, Minghao Wang, Zhao Xu, and Youwei Jia, “Probabilistic Power Flow of Distribution System Based on a Graph-Aware Deep Learning Network”, in *2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, Chengdu, China, July 2021. (Best Paper Award)
- 6 **Huayi Wu**, Zhao Xu, and Minghao Wang, “Convolutional Deep Learning-Based Distribution System Topology Identification with Renewables”, in *2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE)*, Chengdu,

China, Dec. 2021.

- 7 **Huayi Wu**, Zhao Xu, “Fast DC Optimal Power Flow Based on Deep Convolutional Neural Network”, in *2022 IEEE 5th International Electrical and Energy Conference (CIEEC)* Nanjing, China, May, 2022.
- 8 **Huayi Wu**, Youwei Jia, and Zhao Xu, “Adaptive Graph Convolutional Network-Based Distribution System State Estimation”, in *2022 IEEE PES General Meeting*, Denver, CO, USA., July, 2022.
- 9 **Huayi Wu**, Minghao Wang, Zhao Xu, and Youwei Jia, “Active Constraint Identification Assisted DC Optimal Power Flow”, in *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, Shanghai, China, July 2022.

# *Acknowledgements*

I am very grateful to all those people who helped, supported and accompanied me during my three-year Ph.D. study in Hong Kong PolyU.

First and foremost, I would like to express my deepest gratitude to my chief supervisor, Professor Zhao Xu, for his careful guidance, encouragement, and support throughout my Ph.D. study. He is knowledgeable and has unique insights into scientific research. Without his patient guidance and strict requirements, this doctoral dissertation would not have been completed. It is my great honor to have him as my research advisor and closely learn from him. In addition, his encouragement made me not give up easily when faced with difficulties and learned to meet challenges with an optimistic and positive heart. I will always remember his endless kindness, strong support and continuous patience.

I would also like to express my sincere appreciation to my co-supervisor Dr. Bu Siqi for his patient guidance and precious advice. Besides, I would also like to thank reviewers most warmly for their valuable comments on the thesis and papers.

I would also like to express many thanks to my lab-mates, especially Dr. Youwei Jia, Dr. Songjian Chai, Dr. Minghao Wang, Dr. Jian Zhao, and Dr. Xu Xu who gave me valuable advice and selfless help.

I would also like to express special thanks to my beloved parents for their unconditional support and encouragement. I also owe many thanks to my dear husband and son, who have always been there for me. It is their loving support that keep me focused and motivated over these three years.

Last but not least, I would like to acknowledge the support from Ph.D. studentship awarded by The Hong Kong Polytechnic University.

# *Table of Contents*

Abstract.....	I
Publications Arisen from the Thesis.....	III
Acknowledgements.....	V
Table of Contents.....	VII
List of Figures.....	X
List of Tables .....	XIII
Nomenclature.....	XV
Chapter 1 Introduction.....	1
1.1 Background and Research.....	1
1.2 Motivation and Purpose .....	5
1.3 Objectives and Primary Contributions .....	7
1.4 Outlines of the Thesis.....	10
Chapter 2 Gridtopo-GAN for Distribution System Topology Identification .....	11
2.1 Introduction.....	11
2.2 Topology Preserved Node Embedding.....	15
2.2.1 Topological Markov Blanket Model .....	15
2.2.2 Topology Preserved Node Embedding Architecture.....	17
2.3 Grid Topological Generative Adversarial Network Model .....	20
2.3.1 Gridtopo-GAN Framework .....	20
2.3.2 Gridtopo-GAN’s Learning procedure.....	22
2.3.3 Training Procedure of Gridtopo-GAN .....	24
2.4 Case study.....	25
2.4.1 The Performance of the TPNE architecture.....	28
2.4.2 The Performance of the Gridtopo-GAN Model.....	30
2.4.3 Topology Identifications with Loops .....	34
2.4.4 Results with Different Levels of Noises.....	35

2.4.5	Applying Gridtopo-GAN in Large Scale Distribution System .....	37
2.5	Summary .....	38
Chapter 3 Unrolled Spatiotemporal Graph Convolutional Network for Distribution		
	System State Estimation and Forecasting .....	39
3.1	Introduction .....	39
3.2	State Estimation with Spatiotemporal Correlations .....	43
3.3	Unrolled Spatiotemporal Graph Convolution Networks.....	45
3.3.1	Unrolled Spatiotemporal Graph.....	45
3.3.2	Self-adaptive Unrolled Spatiotemporal Graph .....	46
3.3.3	Unrolled Spatiotemporal Graph Convolution Operation.....	47
3.3.4	Gated Operation.....	48
3.4	USGCN Model.....	49
3.4.1	USGCN for State Estimation.....	49
3.4.2	USGCN for State Forecasting .....	51
3.5	Case study .....	52
3.5.1	USGCN in State Estimation .....	55
3.5.2	USGCN in State Forecasting.....	63
3.5.3	USGCN in Large-scale Distribution System.....	66
3.6	Summary.....	68
Chapter 4 Probabilistic Power Flow of Distribution System Based on a Graph-Aware		
	Deep Learning Network .....	69
4.1	Introduction .....	69
4.2	Probabilistic Power Flow Problem.....	72
4.2.1	Deterministic Power Flow .....	72
4.2.2	Probabilistic Power Flow.....	73
4.2.3	The Uncertainty of the RES Power Output .....	74
4.2.4	The Correlation Among RES.....	75
4.3	Graph-aware deep learning Network .....	76
4.3.1	Graph-Aware Scheme.....	76
4.3.2	Structure of GADNL Model.....	77
4.4	Case study .....	77

4.4.1	Considering Independent RES Power Outputs.....	80
4.4.2	Considering Correlation on RES Units.....	81
4.4.3	The performance of the proposed model .....	82
4.5	Summary .....	84
<b>Chapter 5 Graph Attention Enabled Convolutional Network for Distribution System</b>		
Probabilistic Power Flow.....		85
5.1	Introduction.....	85
5.2	Graph Attention Enabled Convolutional Network.....	86
5.2.1	Graph Convolutional Operation .....	87
5.2.2	Self-adaptive Graph Convolutional Operation .....	88
5.2.3	Convolutional Neural Network.....	89
5.2.4	Structure of GAECN.....	91
5.3	Case study .....	92
5.3.1	Considering Independent RES Power Outputs.....	94
5.3.2	Considering Correlation on RES Units.....	97
5.3.3	The Performance of the GAECN.....	101
5.4	Summary .....	104
<b>Chapter 6 Conclusions and Future Works .....</b>		<b>106</b>
6.1	Conclusions .....	106
6.2	Future Works .....	108
References.....		110

# *List of Figures*

Figure 2.1. Topological Markov Blanket model. ....	17
Figure 2.2. The structure of the TPNE. ....	19
Figure 2.3. Gridtopo-GAN's procedure.....	24
Figure 2.4. Convergence curves on onenet, GCN, and TPNE. ....	29
Figure 2.5. Results of different sizes of data sets. ....	29
Figure 2.6. F1 scores of MI, GL, and Gridtopo-GAN.....	31
Figure 2.7. F1 scores of different learning methods with different cases of missing data.....	32
Figure 2.8. The F1 scores from ten cases of the measurement locations. ....	34
Figure 2.9. (a) F1 scores of different missing data situations with mixed topology. (b) The F1 scores of different measurement locations with mixed topology. ...	34
Figure 2.10. The F1 scores from TPNE and Gridtopo-GAN under different noises. .....	36
Figure 2.11. The F1 score and accuracy from the 415-node system with different levels of missing data.....	37
Figure 3.1. Unrolled Spatiotemporal Graph Structure.....	46
Figure 3.2. (a) The structure of the USGCN for DSSE. (b) The structure of the USGCN for state forecasting.....	52
Figure 3.3. RES power output with each curve depicting one day.....	54
Figure 3.4. RES output's average correlation coefficient.....	55

Figure 3.5. The estimated value of voltage magnitudes and phase angles of node 121 of the 1747-node distribution system. ....	58
Figure 3.6. The estimation time of USGCN and traditional methods. ....	58
Figure 3.7. The state estimation of voltage magnitudes and phase angles of the 118-node distribution system at hour 12.....	61
Figure 3. 8. (a) Correlations of RES. (b) Adjacent matrix of topology. (c) Parameters in the unrolled spatiotemporal graph. ....	62
Figure 3.9. The state estimation for a different number of CNN layers in USGCN.	63
Figure 3.10. The prediction results of voltage magnitudes and phase angles of node 55 in the 118-node distribution system.....	64
Figure 3.11. The prediction results of voltage magnitudes and phase angles of node 14-30 of the 118-node distribution system at hour 12.....	66
Figure 3.12. The prediction results of voltage magnitudes and phase angles for bus 500 of the 1746-node distribution system. ....	67
Figure 4.1. The structure of the GADLN. ....	77
Figure 4.2. The voltage magnitudes and line active power flow.....	80
Figure 4.3. The voltage magnitudes and line active power flow.....	81
Figure 4.4. The training performance of the GADLN.....	83
Figure 5.1. The structure of the GAECN.....	91
Figure 5.2. The voltage magnitudes and line active power flow without correlations. ....	96
Figure 5. 3. The histograms of MC and GEACN without correlations.....	97
Figure 5. 4. The voltage magnitudes and line active power flow with correlations.	99

Figure 5. 5. The errors of voltage magnitudes with correlations.....	99
Figure 5. 6. The histograms of MC and GEACN with correlations. ....	100
Figure 5. 7. (a) Adjacency matrix of topology. (b) Correlation coefficient matrix of RES. (c) Self-adaptive graph. ....	104

# *List of Tables*

Table 2.1. Confusion Matrix.....	27
Table 2.2. The Sizes of the Parameters.....	27
Table 2.3. Missing Data Numbers on the IEEE 33-node System.....	31
Table 2.4. Missing Data Numbers on the 118-node System .....	31
Table 2.5. Computation Time.....	34
Table 2.6. F1 Scores of TPNE with complete measurements and different noises..	37
Table 3.1. The locations of the wind turbines and solar panels .....	55
Table 3.2. The evaluation index of USGCN compared with traditional methods with and without correlations .....	57
Table 3.3. The Evaluation Index in State Estimation of USGCN in Comparison with Data-driven Methods with Correlations .....	60
Table 3.4. The evaluation index in State forecasting of USGCN compared with deep learning methods with correlations.....	65
Table 3.5. The evaluation index of USGCN of 1746-node distribution system .....	68
Table 4.1. Wind Speed and Solar Irradiance Distribution Functions' Parameters...	79
Table 4.2. Correlation Coefficient Matrix of the RES.....	79
Table 4.3. The mean error of comparison of different methods without correlation	81
Table 4.4. The mean error of comparison of different methods with correlation ....	82
Table 4.5. The calculation time of comparison of different methods with correlation .....	83

Table 5. 1 The Locations of the RES units.....	94
Table 5. 2 The mean error of comparison of different methods without correlation of IEEE 33 system.....	94
Table 5. 3. The mean error of comparison of different methods with the correlation of IEEE 33 system. ....	98
Table 5. 4. The mean error of different data-driven methods with correlation. ....	102
Table 5.5. The calculation time of comparison of different methods.....	104

# *Nomenclature*

CNN	Convolutional Neural Network
DSSE	Distribution System State Estimation
FCN	Fully Connected Network
GADLN	Graph-Aware Deep Learning Network
GAECN	Graph Attention Enabled Convolutional Network
GAN	Generative Adversarial Network
GB	Graph Embedding
GCN	Graph Convolutional Network
LAV	Least Absolute Value
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MB	Markov Blanket
PPF	Probabilistic Power Flow
PV	photovoltaic
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
TPNE	Topology Preserved Node Embedding
USGCN	Unrolled Spatiotemporal Graph Convolutional Network
WLS	Weighted Least Squares

# *Chapter 1 Introduction*

## **1.1 Background and Research**

In the past decade, the world has experienced an increasing proliferation of renewable energy resources (RES) due to the rapidly booming advances in renewable energy technology and industries, e.g. solar photovoltaic (PV) panels and wind turbines. The global total installed capacity of distributed energy resources has achieved the biggest increase ever, by more than 256 gigawatts (GW), and reached 1668 GW by the end of 2020 [1]. Among various distributed energy resources, e.g. bioenergy, ocean power, geothermal power, wind power, etc. solar power and wind power are growing rapidly in many countries. The total capacity of solar PV and wind power have reached 760 GW and 743 GW, respectively. More than nine countries generated more than 20% of their electricity from solar photovoltaics panels and wind turbines in 2020. Favorable economics have boosted interest in distributed rooftop solar PV systems. South Australia achieved one of the world's highest levels of solar penetration in 2020. Wind power accounted for a substantial share of electricity generation in Denmark with the highest penetration in the world at over 58% at the end of 2020 [1].

The widespread adoption of RES may bring about lots of economic, environmental, and technical benefits, such as reducing greenhouse gas emissions, reducing transmission losses, and so on. However, the proliferation of RES is dramatically changing the operation states of the distribution systems. Traditionally,

the distribution systems deliver the electricity from the large-scale central power plants to the consumers via the radial distribution networks. On contrary, the electricity is generated locally and consumed by customers directly. Since the outputs of RES are inherently intermittent and uncertain, the increasing proliferation of RES in the distribution system brings new challenges. For example, the uncertain fluctuation of RES will introduce significant uncertainty to the system state monitoring functions, such as topology identification, state estimation, and probabilistic power flow [2]. More specifically, the conventional transmission system enjoys redundant measurements to obtain the system topology via logic signal data processing, while the topologies change frequently in distribution systems where the measurement is limited due to expensive installation costs. This phenomenon arouses a significant requirement for advanced approaches to distribution system topology identification. Besides, instead of a few RES plants in a transmission system that covers a large area and are assumed independent, the RESs in the distribution system are close to each other, leading to dependent power output. The traditional state estimation for transmission systems cannot adapt to this change efficiently so that more appropriate methods are needed. Furthermore, the forceful fluctuations brought by the high penetration of RES introduce inevitable uncertainties in system states. To timely acquire the system states uncertainties also call for new techniques. Therefore, the changes in the distribution system diagram spur the progressive approaches for timely monitoring of the distribution system states.

Recently, thanks to the widespread adoption of measurement devices, such as smart meters, massive historical data are available [3]. To better leverage such measurement data to facilitate the monitoring of the system states, data-driven deep learning methods are proposed. This is because the deep learning approach has the powered capability of data mining and high latitude fitting. However, the application of the traditional system state monitoring functions is limited to specific scenarios, which rarely take full advantage of abundant measurement data. The

inappropriate implementation of these functions may give incorrect distribution system states, misleading the operators and thus threatening the normal operation of the distribution system. Therefore, it is crucially important to utilize deep learning to fully facilitate the system state monitoring functions.

The network topology gives the connection of the components of the power system to maintain its aim of power transmission and distribution to the consumers. In comparison with the transmission network that is structured in looped with full observability, the distribution network is operated in radial. The architecture of the distribution system consists of several feeders delivering the electricity from the substation that connects with the transmission system to the consumers. It is deployed with limited measurement devices and thus with poor observability. This is because unlike transmission systems that enjoy a high level of data redundancy, it is too limited financial budget to deploy enough digital (on/off) measurement devices which represent the state of the line breaker in such a large-scale distribution system [4]. Besides, the aged and outdated monitoring devices also result in inadequate monitoring and measurement of electrical distribution systems. Based on these reasons, it is rare to ensure a full-scale monitoring distribution system. However, challenged by the uncertainty due to the proliferation of RES, to maintain the reliable and economical operation of the distribution system, the topology changes frequently in practice, sometimes several times an hour [5]. It requires timely identify the distribution system topology to facilitate the decision-makers. Considering aforementioned factors, alternative topology identification approaches such as the data-driven ones can be investigated to handle limited measurements.

The distribution system state estimation provides the system states to the operators according to the available measurements. The increasing adaption of RES in the distribution system brings significant challenges to the state estimation due to complex correlations in nodal power injections. Different from the transmission system covering a wide area, the distribution system covers a relatively small area

but with more feeder lines and switching devices, and limited measurement devices as well as integrated with dispersed RES. In this situation, the renewable energy deployment sites are close to each other, sharing similar weather and environment, leading to similar power profiles generated by these renewable energy units [6]. Thus, there are significant complex correlations among nodal power injections, resulting in hidden spatiotemporal correlations among nodal measurements of distribution systems [7]. These correlations pose inevitable bias in the system state estimation calculations and thus render a significant challenge to be addressed [2]. However, many previous works in DSSE ignore this gap due to no immature approach available to fill it. Thus, it is an extremely urgent task to explore a data-based approach to DSSE and state forecasting with the proliferation of RES.

The probabilistic power flow qualifies the uncertainties of nodal power injections spread to the nodal voltage, line power flow, and other system operation states. The growing adoption of RES in distribution systems introduces unmissable fluctuations in system states due to the dramatic stochastic and intermittent nature of RES. This uncertainty will propagate from the nodal power injections to all system states during the operation of the distribution system. In addition, high penetrated wind power and solar panels resources are deployed in one low voltage distribution system which generally covers a relatively smaller area geographically. In such a situation, the wind and solar power generation profiles tend to share similar patterns of power output features and thus represent high complex significant correlation [6]. Since the distinguished ability of the PPF allows the uncertain distribution of the system states to be represented, it is an important basic analysis function for reliable power system management. The high-penetrated RES brings complexity to the PPF calculation and thus leading the errors, which may further mislead the correct decision-making of operators [8]. Thus, these challenges require a data-enabled PPF calculation tool to handle the dense uncertainties and complex correlations.

## 1.2 Motivation and Purpose

At present, the state monitoring functions of distribution system states, e.g. topology identification, state estimation, as well as the system planning basis function the probabilistic power flow, have attracted much academic attention. There is a great number of approaches being proposed to solve these problems. As mentioned above, the challenges introduced by the low observability and the proliferation of RES give the considerable complexity in the monitoring of distribution system states. It is crucially important to overcome this complexity, otherwise, it would give error information on distribution system states and thus mislead the operators to make unreliable management decisions. However, the conventional approaches for transmission systems cannot directly fit these issues in the distribution system. Thus, advanced approaches are required to be developed for distribution system state monitoring.

This thesis aims at proposing effective approaches for the monitoring of distribution system states under the limited measurements and uncertainty brought by the proliferation of RES. Specifically, by combining the graphical nature of physical distribution systems, several graph-oriented learning models are developed, including the grid topological generative adversarial network for topology identification, the spatiotemporal graph convolutional network for state estimation and forecasting, the graph attention enabled convolutional network for probabilistic power flow calculation. These approaches can also be applied to other specific operational applications.

As for the grid topological generative adversarial network for topology identification, it aims at the topology identification problem in distribution systems with very limited measurements. As mentioned above, unlike transmission systems that enjoy a high level of data redundancy, the distribution systems are generally with poor observability due to the too expensive cost to fully deploy measurement device in such a large-scale distribution system. Besides, with the proliferation of

RES, the distribution system's topology varies frequently and sometimes even several times per hour in practice. The topology identification utilizes the available measurements to infer the distribution system's topology. So far, several related studies have been carried out, including the Markov Random Field (MRF) based method [9], mutual information method [10], graphical learning approach [11], and so on. However, most related research leverages the conditional independence among measurements to infer the topology through enumerating the dependence among every possible node pair which is time-consuming. Besides, these methods are restricted to identifying the limited topologies. Therefore, timely identifying numerous topologies of the distribution system with limited measurements becomes a formidable gap. To fill up this gap in supporting the reliable operation of the system, in this work, a grid topological generative adversarial network is proposed. It aims at identifying the topology of the distribution system in an effective and efficient manner with very limited measurement data collected from the smart meters, given disturbing factors such as missing or bad data etc.

As for the unrolled spatiotemporal graph convolutional network for state estimation and forecasting, it aims at the state estimation and state forecasting synchronously considering the spatial and temporal correlations among nodal power injections. With the proliferation of the RES, it is well known that the distribution system is integrated with various renewable energy sources, while it covers a relatively small area geographically. This phenomenon results in renewable energy deployment sites being close to each other and thus they are generally shared similar weather and environment. Similar profiles generated by these renewable energy units lead to complex correlations among nodal power injections, leading to significant state estimation error in traditional approaches, e.g. WLS, LAV. Thus, the impact of the nonlinear spatiotemporal correlations on state estimation remains a significant gap. These challenges inspire related approaches in some research, such as WLS based method [12], [13], that linearizes the correlations. Some machine learning-based methods [14], [15] try to characterize

the complex spatiotemporal correlations while full knowledge on correlations' distribution is required. As such, advanced approaches should be put forward to deal with this problem. With the boom in measurement data collected from the smart meters, numerous measurements can be leveraged to extract the distribution between the measurements and the system states. This inspires the proposed spatiotemporal graph convolutional network model to address the state estimation and state forecasting ahead of time, avoiding being disturbed by the complex correlations.

Although the state estimation provides the system states according to the measurement data, the distributions of the system states are also needed to be investigated. The probabilistic power flow calculation qualifies the uncertainty of the RES and spreading to the system states. As mentioned above, the correlations among RES exist, which also leads to a significant error in the probabilistic power flow calculation according to the traditional methods, like PEM. Although some methods, e.g. Nataf transformation [16] and polynomial chaos [17], can remove the correlation's impact to a certain degree, they cannot adapt to the distribution system with high penetration of RES featuring complex correlations. The impact of uncertainty and correlations of RES on probabilistic power flow remains a gap. The graph attention enabled convolutional network is proposed for the probabilistic power flow calculation problem, aiming at representing the distributions of the system states introduced by the uncertainty of the RES. In this way, the novel graph attention enabled convolutional network is developed to approximate the power injections and the system states with a low computational error and high efficiency.

### **1.3 Objectives and Primary Contributions**

As discussed previously, the research on artificial intelligence-based power system states perception is at its early stage. Moreover, with the rapidly increasing

integration of renewable energy, the traditional approaches are unsuitable to address the challenges. More advanced approaches are required to be precisely aware of the operational states of the distribution system to provide more accurate information to operators for reliable management. To fill up these research gaps, progressive methods are developed for graph-oriented awareness of distribution system states in the following four aspects. The main contributions are summarized below.

- 1) The Gridtopo-GAN is innovatively proposed to adapt to the topology identification problem in the distribution system under the challenges of limited measurements and with radial/meshed topologies. To this end, firstly, by proposing the topological Markov Blanket, both the radial and meshed topologies can be inferred effectively using the measurements within the given window size. Then, by leveraging the graph embedding techniques with the topological Markov Blanket, the TPNE model is proposed, which compresses the numerous changing topologies in a low dimension space to achieve the high topology identification accuracy in a large-scale system. Furthermore, the GAN model with an auxiliary classifier allows the partial observation up to the extent of half bad or missing nodal voltage measurements based on the case study. Different from the traditional topology identification approaches that require full observability and numerous training parameters, the proposed model compresses numerous topologies using limited measurements to improve the identification accuracy.
- 2) The unrolled spatiotemporal graph convolutional network model has been proposed in this work to be better aware of the distribution system states that are exposed to complex correlations among the renewable power outputs. Instead of considering the spatial and temporal correlations separately in the previous works, three aspects of spatiotemporal correlations are captured simultaneously by the proposed unrolled spatiotemporal graph model that leveraged the splicing of the spatial graphs across adjacent time steps. In this way, ameliorated forecasting accuracy and computational efficiency can be achieved. Then, the node embedding is

leveraged to construct the dependence described inside the unrolled spatiotemporal graph to learn the nonlinear spatiotemporal correlations automatically instead of utilizing the linear correlation coefficient matrix. Further, the ahead-of-horizon state forecasting is achieved effectively by the multi-module layers that capture the long-range spatiotemporal correlations.

- 3) A graph-aware deep learning network model is innovatively leveraged to handle the probabilistic power flow (PPF) that is exposed to complex dependence among renewable power outputs. Specifically, to fully capture the mapping from the fluctuated power injections and the uncertain system states, the convolutional operation is introduced to aggregate the correlations among renewable power outputs according to their geographical locations to facilitate the PPF. In this way, the deviation pattern of the system state variables can be well learned. Thus, the proposed GADLN can learn the implicit distribution of the correlation of renewable power autonomously according to the historical data, and thus the improved effectiveness and speed-up calculation can be achieved.
- 4) Graph attention enabled convolutional network is proposed to calculate PPF considering the hidden correlation of the wind and solar power resource injections. This model utilizes the graph attention enabled convolutional layer to fully extract and leverage the correlations of the nodal power injections to improve the PPF calculation accuracy. Besides, within this layer, different from only utilizing the Euclidean structure adapted convolutional neural network, the node embedding technique is integrated into the graph convolutional operation to extract the non-Euclidean structure of the power system. In this way, instead of using the linearized correlations of nodal power injections from historical data, the self-adaptive graph convolutional operation based on the graph embedding technique can capture the complex correlations automatically. Furthermore, the convolutional neural network operation is followed to handle the violent fluctuations of the outputs of renewable energy.

## **1.4 Outlines of the Thesis**

The thesis is organized as follows. In Chapter 2, a novel grid topological generative adversarial network model is presented to identify the topology of the distribution system with limited measurements. In Chapter 3, an unrolled spatiotemporal graph convolutional network is developed for distribution system state estimation and forecasting with spatiotemporal correlations on RES outputs considered. In Chapter 4, a graph-aware deep learning network model is built to calculate the probabilistic power flow considering the unknown correlation distribution pattern between the wind and solar power generation. In Chapter 5, the graph attention enabled convolutional network model is innovatively proposed to approximate probabilistic power flow to consider the complex correlations among nodal power injections. In Chapter 6, the conclusion of this thesis is given, and the future works are introduced.

# *Chapter 2      Gridtopo-GAN for Distribution System Topology Identification*

## **2.1 Introduction**

The control of the distribution system is threatened by the proliferation of renewable energy sources, energy storage, and electric vehicles. In this situation, the deepening of the complexity of power grid operation depends on more timely monitoring information than ever before. Such monitoring information includes the topology status [18], which changes frequently in practice, and sometimes even several times an hour [19]. Thus, quick distribution topology identification becomes a more important than ever task to facilitate the reliable operation of the distribution system [20].

Transmission systems are designed to carry large volumes of electricity from generators via the high voltage power lines to the distribution system sides, so strict monitoring is required to ensure the safety of system operation. On the contrary, traditional power distribution systems typically operate in radial configurations, delivering power from the substation connected to the transmission grid to individual customers without strict monitoring, for the following reasons. The distribution network often covers a relatively small area geographically while with a large number of load demand nodes, feeder lines and switching devices. In this situation, installing

a large number of measurement devices to ensure complete monitoring is extremely expensive and rarely practical. Thus, there is the limited measurements in distribution systems [21]. Even though this problem can be alleviated by pseudo-measurement that is generated according to the historical data, the obtained measurements are not accurate enough to make sure the topology is identified with high accuracy [22]. In addition, aging and outdated monitoring equipment also lead to insufficient monitoring of power distribution systems [23]. Therefore, comprehensive monitoring systems for distribution networks are uncommon, which becomes a huge challenge for distribution system topology identification. Moreover, in some urban areas with high power requirements, the topology here is set as meshed with loops. This poses another threat by introducing high computational complexities in the topology identification problem[24]. Therefore, it is an urgent task to develop a progressive approach to obtain a timely topology both in radial and meshed distribution systems [25].

In the previous work, a lot of approaches are employed to focus on the distribution system topology identification problem. The additional variables that describe the branch status are leveraged in the Weighted Least Square (WLS) function on [26]. Maximum-a-Posteriori (MAP) estimation leveraging the line impedance parameters is utilized in [27] for topology identification according to the historical active power injections. Since the complete line impedance information which is rarely available is the premise, these methods are unsuitable in practice. To mitigate this challenge, a lasso regression model that linearizes WLS is introduced in [28] and [29] for topology identification. However, these approaches are highly sensitive to measurement errors due to nonlinear power flow equations being ignored, which results in unsatisfactory accuracy.

To release the dependence on the system parameters and improve the ability against errors, the graph features on the distribution system states are leveraged for topology identification by inferring the topology according to the measurements. A Markov Random Field (MRF) that describes the nonlinear dependence among nodal voltages is proposed in [9] to estimate the topology. Since the topology configuration change is

frequent, the MRF is also leveraged in [30] to learn the dependence of all nodes represented by the feature function. The graphical learning approach is developed in [11] to identify topology by conditional independence tests, requiring complete line impedance information which is not available in practice. By employing the graphical features hidden in the system states, the principal component analysis is leveraged in [31] to identify a distribution system with a radial structured topology. The Chow-Liu algorithm is leveraged in [32] to guarantee that the pairwise mutual information among phasors can be used to depict the radial topology so that the topology identification can be achieved. Based on the work for non-radial topology identification in [32], [10] detects a loop by searching the maximum mutual information of all nodes with two possible parent nodes which involve high-order time complexity of the calculation, and then applies the spanning minimum tree algorithm to identify the radial topology. The inverse covariance matrix is utilized on [24] to search the neighborhood of each node to search the meshed topology, but also full knowledge about line impedance is required. These methods are with heavy computational burden especially in the large-scale distribution system due to it traverses all possible dependence between all node pairs.

To ameliorate the computational efficiency, a support vector machine approach is pre-trained in [33] so as to achieve real-time topology identification. However, this method needs too much measurement data to be applied in a large-scale system. The Kullback-Leibler divergence is leveraged in [34] to guide the indicator to be trained for topology identification. Compressive sensing (CS) and graph theory are utilized in [35], [36] to release the reliance on numerous measurement collections. However, these approaches need full nodal observation via full-scale system monitoring which is not yet common. To remove the full observation dependence, the meter placement strategies are leveraged in [37] and [38] that optimize the limited sensors' locations to ameliorate the accuracy of topology identification. By leveraging the measurements at the terminal or leaf nodes, a learning scheme is leveraged in [39] iteratively identifying the dependencies between the leaf and the root node. This learning scheme is extended

by [40] via the linear power flow function when some nodal measurements are unavailable. However, the locations of measurement devices are limited to specific places, such as the end of a feeder, which only adapts to limited configurations rather than frequent changing topologies and thus it is not applicable. The CS method is leveraged in [41] to identify topology based on partial observation but this method is highly sensitive to the measurement noise. Therefore, these approaches cannot be applied in distribution system topology identification under limited measurements with noise.

To identify distribution systems under the challenges of limited measurements and with radial/meshed topologies, this section proposed a novel Gridtopo-GAN model to adapt to the topology identification problem in the distribution system directly and accurately in real-time [42]. Specifically, by proposing the topological Markov Blanket, both the radial and meshed topologies can be inferred effectively using the measurements within the window. Then, by leveraging the graph embedding techniques, the Topology Preserved Node Embedding (TPNE) is able to utilize the local nodal measurements rather than the entire measurement data to predict the status of the edges. In this way, numerous changing topologies are compacted into a low dimension space so as to achieve high topology identification accuracy in a large-scale system. Moreover, the Gridtopo-GAN allows the generator to complement missing data so that the proposed model can be achieved under limited measurement situations. Furthermore, the impact of the measurement locations is excluded by the simulation results. The effectiveness of the Gridtopo-GAN model is verified by conducting the IEEE 33-node, 118-node, and 415-node distribution test systems. In summary, the contributions are summarized as follows.

- 1) To adapt to a distribution system with radial/meshed topologies, the topological Markov Blanket model is proposed to characterize mapping from the measurements and the topologies to infer both the radial and meshed topologies directly.

- 2) Instead of a conventional neural network with numerous training parameters, the TPNE leveraging the graph embedding techniques is proposed to compact numerous

changing topologies into a low dimension space to achieve high topology identification accuracy in a large-scale system.

3) Different from the few missing measurement assumptions in the previous work, the Gridtopo-GAN model allows the generator to complement missing data so that this model can be applied to the missing data situation.

The remaining sections are organized as follows. The TPNE-based topology identification in the mathematical format is proposed in Section 2.2. The novel Gridtopo-GAN model as well as its procedure are developed in Section 2.3. The experimental results and the corresponding discussions are introduced in Section 2.4. Finally, the works of this section are concluded in Section 2.5.

## **2.2 Topology Preserved Node Embedding**

Firstly, the MB model is introduced to describe the basic logic in the topology identification of the proposed model. Then, the TPNE is proposed to represent the relationship between the measurements and the topology so that the compact format with low dimension can be achieved, which will later be integrated into the GAN.

### **2.2.1 Topological Markov Blanket Model**

In the previous works, the graph theory is leveraged to explore the topology identification problem [11]. This remains a gap is that the number of topologies that can be identified by these approaches is very limited because the computational complexity increases exponentially with system size. Due to the frequent changing topologies in the modern distribution system, it is essential to develop a new model to identify as possible as many topologies. This inspires the introduction of the Markov Blanket concept to describe the topology identification problem. This is because the distribution system can be defined as a Bayesian Network  $\mathcal{G}=(\mathcal{V},\mathcal{E})$ , where  $\mathcal{V}$  denotes a

set of random variables  $\mathcal{V}=\{v_1,\dots,v_i,\dots,v_N\}$  which also indicates the nodes of the system,  $\mathcal{E}=\{s_1,\dots,s_i,\dots,s_L\}$  presents conditional dependencies between these random variables  $\mathcal{V}$  which also presents edges of the system [10].

For node  $v_i$  in  $\mathcal{G}$ , the Markov Blanket of  $v_i$ ,  $MB(v_i)$ , is defined as any subset of  $\mathcal{V}$ , conditioned on which other variables are independent with  $v_i$ . That is  $\{v_i\} \perp \mathcal{V} - (MB(v_i) \cup \{v_i\}) \mid MB(v_i)$ . It represents the fact that the information in  $MB(v_i)$  can be utilized to infer the  $v_i$  rather than the information of  $\mathcal{V} - (MB(v_i) \cup \{v_i\})$ . Based on this assumption, the topology Markov Blanket concept is proposed to build the corresponding topology identification model.

When the nodal voltage measurement data set  $v$  is available, a subset  $V_t$  of  $v$  is defined as the window nodes of edge  $s_i$  within a window,  $w$ , in a feeder. The status (on/off) of each edge  $s_i$  can be inferred by the expression  $P_u(s_i | V_t) = P_u(s_i | \{v_{1-w}, \dots, v_1, v_2, \dots, v_{2-w}\})$ . An example to illustrate this concept is depicted in Figure 2.1. When the window size of  $w$  is set 3, this definition can be rewritten according to the above definition as: (a)  $P_u(s_i | V_t) = P_u(s_i | \{v_{1-3}, v_{1-2}, v_{1-1}, v_1, v_2, v_{2-1}, v_{2-2}, v_{2-3}, v_{3-1}, v_{3-2}, v_{3-3}\})$ ; (b)  $P_u(s_i | V_t) = P_u(s_i | \{v_{1-3}, v_{1-2}, v_{1-1}, v_1, v_2, v_{2-1}, v_{2-2}, v_{2-3}\})$  and (c)  $P_u(s_i | V_t) = P_u(s_i | \{v_{1-3}, v_{1-2}, v_{1-1}, v_1, v_2, v_{2-1}, v_{2-2}, v_{2-3}, v_{3-1}, v_{3-2}, v_{3-3}\})$ . Note that the meshed topology with the loop is contained in case (c). Thus, the problem of topology identification focused here is to use historical voltage magnitudes  $V_t$  and known topology  $s_i$  to obtain the  $\arg \max P(s_i | V_t)$  so that the timely unknown topology can be inferred according to the maximum a-posteriori probability  $P(s_i | V_t)$ . To this end, the Kullback-Leibler (KL) divergence is leveraged to calculate the closeness between two distributions  $P(s_i | V_t)$  and  $P_{True}(s_i | V_t)$ .

$$D_{KL}(P_{True}(s_i | V_t) \| P(s_i | V_t)) = \sum P_{True}(s_i | V_t) \log \frac{P_{True}(s_i | V_t)}{P(s_i | V_t)} \quad (2.1)$$

The distribution  $P(s_t | \mathbf{V}_t)$  is equivalent to maximizing the KL divergence as follows.

$$\arg \min D_{KL}(P_{True}(s_t | \mathbf{V}_t) \| P(s_t | \mathbf{V}_t)) = \arg \min \mathbb{E}[\log(P(s_t | \mathbf{V}_t))] \quad (2.2)$$

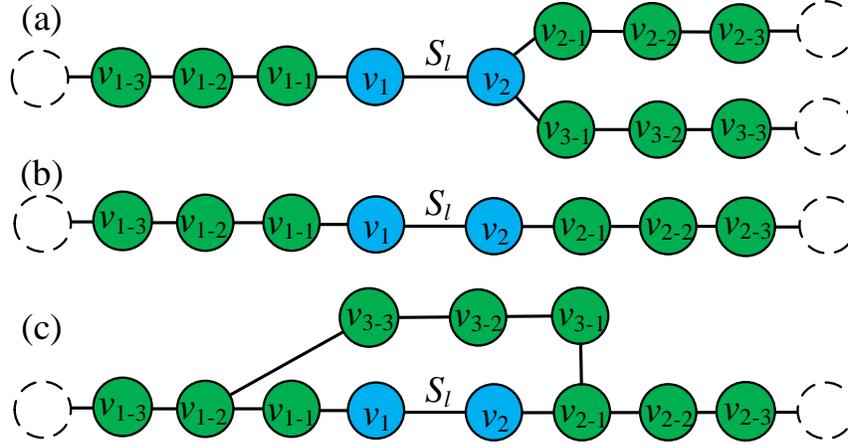


Figure 2.1. Topological Markov Blanket model.

### 2.2.2 Topology Preserved Node Embedding Architecture

Because the scale of the distribution system is generally large and the topology changes greatly, the parameters of the general neural network model for such a situation are massive, which is difficult to train with extreme computing resources and obtain a stable trained model. Based on the introduction of the topological Markov Blanket model, the graphical nature of the distribution system inspires the leverage of the graph embedding (GB) topologies, which is popular with its graph representation capability from high dimension to low dimension. In this way, the various topologies can be compressed into low dimension numerical matrix.

More specifically, the classic GB uses a numerical vector, commonly defined as node embedding, to describe each node in a graph. These vectors can compact the exact dependence among nodes to a low- dimensional space [43]. On the contrary, in the topology identification problem, the nodal dependences are unknown such that the topology identification cannot be modeled directly by the GB. Despite this, the

thoughts of GB can be learned. This is why the topological MB is introduced. Based on the topological MB, the system states and the implicit topology information can be compressed in an efficient manner. Thus, a new neural network combining the topological MB and GB techniques is innovatively represented, defined as Topology Preserved Node Embedding (TPNE).

The advantages of the proposed TPNE are obvious. It fully leverages the merits of the non-Euclidean data processing ability from the graph neural network (GNN) technique [44]. Specifically, different from the general graph embedding technique that utilizes the nodal pair dot product to describe the dependence of the nodal pair, the TPNE utilizes the concatenation of the node embedding within the topological MB to represent the implicit connectivity among nodes. After this process, the TPNE ignores redundant nodal information, which contributes little to the topology identification while introducing a significant computational burden and modal complexity. Therefore, the model's efficiency and effectiveness can be enhanced.

To achieve the topology identification model, there innovatively introduces two categories of graph embedding parameters. The first one is the node embedding, defined as  $\theta_1: \mathcal{V} \rightarrow \mathbb{R}^{|\mathcal{V}| \times k}$ , which is a matrix with the size of  $|\mathcal{V}| \times k$ , where each row of this matrix represents a node embedding vector. The second one is the edge embedding, defined as  $\theta_2: \mathcal{E} \rightarrow \mathbb{R}^{|\mathcal{E}| \times k}$ , where each row of this matrix represents an edge embedding vector of the distribution system.  $k$  is the dimension of the node embedding and the edge embedding. Thus, by leveraging the neural network model to approximate the topology identification model in (2.2), the probability distribution of the edges can be characterized.

$$f^l = (\text{ReLU}(V_l G_l^T \theta_1 + b_1)) G_l^h \theta_2 + b_2 \quad (2.3)$$

In (2.3),  $V_l$  is the measurements on nodal voltage magnitudes, whose size is defined as  $\text{Batch} \times |\mathcal{V}_l|$ , where the batch denotes the sampling number, and  $\mathcal{V}_l$  is the window size for edge  $s_l$ .  $\text{ReLU}(\cdot) = \max(0, x)$ , rectified linear unit (ReLU) function, which is defined as the

activate function. The bias parameters  $b_1$  and  $b_2$  are in the vector formats with the sizes  $k$  and  $|\mathcal{E}_i|$ , respectively, where  $\mathcal{E}_i$  denotes the edge  $s_i$ . Although more than one matrix can be utilized as graph embedding parameters, two matrices are representing the graph embedding parameters in the proposed model architecture, which is determined by optimal hyperparameters selection as well as the ReLU activation function.

Then, to select the node within the topological MB, the embedding extraction matrices  $G_i^n$  and  $G_i^b$  are introduced, which are parameterized with  $|\mathcal{V}| \times |\mathcal{V}|$  and  $|\mathcal{E}_i| \times |\mathcal{E}|$ . In this way, the node embedding and the edge embedding within the topological MB corresponding to the edge  $s_i$  can be selected.

$$G_i^n = [g_1^n; \dots; g_i^n; \dots; g_{|\mathcal{V}|}^n], \quad i \in \mathcal{V} \quad (2.4)$$

In (2.4),  $g_i^n$  is the  $i^{\text{th}}$  row of  $G_i^n$ , which is a  $|\mathcal{V}|$  dimensional one-hot vector with  $j^{\text{th}}$  element as 1, where  $j$  is the  $j^{\text{th}}$  node in  $\mathcal{V}$ . Similarly, the edge embedding parameters for each area can be defined as follows.

$$G_i^b = [g_1^b; \dots; g_i^b; \dots; g_{|\mathcal{E}_i|}^b], \quad i \in \mathcal{E}^h \quad (2.5)$$

In (2.5),  $g_i^b$  is the  $i^{\text{th}}$  row of  $G_i^b$ , which also denotes a  $|\mathcal{E}|$  one-hot vector with  $j^{\text{th}}$  element as 1, where  $j$  is the  $j^{\text{th}}$  edge in  $\mathcal{E}$ .

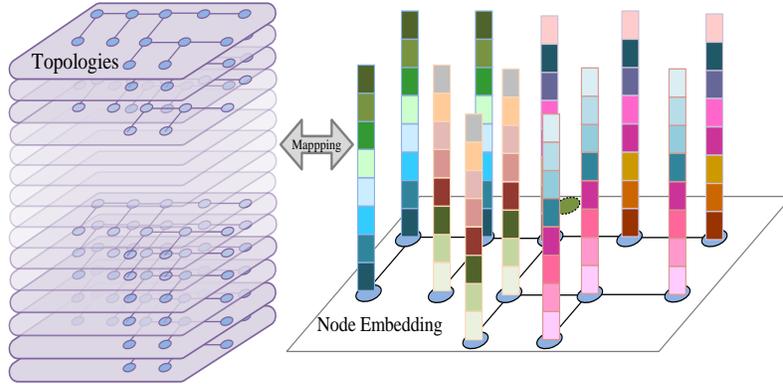


Figure 2.2. The structure of the TPNE.

The voltage magnitude is commonly collected by smart meters in the distribution system so that it is utilized as the input of the proposed model. The parameters in TPNE are shared. However, in the training process, only the node/edge embedding parameters within a topological MB are updated together under the guidance of the objective function. In this way, the topology identification model conditioning on the measurements is compacted in the TPNE, intuitively shown as the bars in Figure 2.2. Thus, this model releases the rely on the massive parameters learning so that enhanced efficiency can be achieved. Later, to handle the lack of measurement data problem, the TPNE is served as the architecture block inside the GAN framework.

## 2.3 Grid Topological Generative Adversarial Network Model

Firstly, the topology identification based on the Gridtopo-GAN is introduced, including the discriminator, the generator, and the auxiliary classifier. Secondly, the formulations of the entire framework are represented. Finally, the procedure of this framework is listed.

### 2.3.1 Gridtopo-GAN Framework

Since the full observability is rarely in the distribution system, GAN is developed to address the missing measurement issue in this section. Based on the game theory, the GAN is introduced by Goodfellow et al. [45]. GAN is famous for its generative ability which has been applied in many tasks with sound effects [46]. This is because GAN leverages two neural networks, generator and discriminator, to compete with each other so that the generative can learn the nearly true distribution of data to generative samples that can fool the discriminator. Specifically, the generator is defined as  $G$ , whose input is the random noise  $z$ , and output is the fake samples  $V^{fake} = G(z)$ . The discriminator is defined as  $D$ , whose inputs are the real samples and the generated

samples from the G, and outputs are the probability of samples, indicating their real or not. D is trained to maximize the log-probability as:

$$Loss = E[\log P(C = real | V^{real})] + E[\log P(C = fake | V^{fake})] \quad (2.6)$$

In (2.6), the second term is the objective of G.

Conventionally GAN utilizes the convolutional neural network as the internal structure of GAN. Different from the conventional GAN, the TPNE proposed in the above section is introduced to construct the GAN to adapt the GAN in the topology identification model. Moreover, in other perspectives, the problem of missing measurements is considered by leveraging an auxiliary classifier. The advantage of introducing the auxiliary classifier is that the Gridtopo-GAN infers the branch status directly according to the measurement data instead of the traditional GAN's single function of distinguishing whether the measurement is real or not.

In the G of Gridtopo-GAN, the inputs consist of the measurement data  $V_l^{av}$  and the noise  $z$ . The measurement data is the available voltage magnitudes while the noise is the randomly generated normal distribution value. Thus, the proposed GAN is conditional. G is utilized to generate voltage magnitude  $V_l^{fake} = G(V_l^{av}, z)$ . The output of D provides a probability  $P(C_l | V_l)$  that distinguishes the real or the generated measurements from the inputs. Besides, the auxiliary classifier provides a probability  $P(s_l | V_l)$  that indicates the status of the edge. The probability closing to 1 means a high possibility on connectivity. More specifically, the purposes of Gridtopo-GAN are to obtain the models as listed.

**Generator:**  $G(V_l^{av}, z)$  tries to learn the implicit distribution on the relationship between the nodal voltages and the branch status.

**Discriminator:**  $D_1(C_l | V_l) = P(C_l | V_l)$  aims to find out the false measurements.

**Auxiliary classifier:**  $D_2(s_l | V_l) = P(s_l | V_l)$  is to infer the status of the branch according to the generated measurements from the G.

The loss function of the D is to maximize the log-probability as shown in the following.

$$\max L_c = \sum_{l=1}^L (\mathbb{E}[\log P(C_l = \text{real} | \mathbf{V}_l^{\text{real}})] + \mathbb{E}[\log P(C_l = \text{fake} | \mathbf{V}_l^{\text{fake}})]) \quad (2.7)$$

The AC's loss function is to maximize the log-probability according to the guidance of the branch status labels.

$$\max L_s = \sum_{l=1}^L (\mathbb{E}[\log P(s_l = s | \mathbf{V}_l^{\text{real}})] + \mathbb{E}[\log P(C_l = s | \mathbf{V}_l^{\text{fake}})]) \quad (2.8)$$

where  $s$  is the status of the branches labels by 1/0 indicating on/off.

The G's loss function is to minimize the  $L_s - L_c$ .

$$\min L_g = L_s - L_c \quad (2.9)$$

According to the mentioned equations, by alternately maximizing and minimizing the loss function  $L_c$ , the parameters of the G and the D can be updated. In addition, by alternately maximizing and minimizing the loss function  $L_s$ , the parameters of the auxiliary classifier are also updated. In other words, the competing game in G and D allows G to ameliorate its generative effectiveness to fool the D. The game stops when the generated measurement from G is near the real one that can confuse D. The alternate learning process in the G and auxiliary classifier can ameliorate the ability of AC to outputs the status of branches.

### 2.3.2 Gridtopo-GAN's Learning procedure

#### (1) Discriminator

In Gridtopo-GAN, the sigmoid function  $\sigma(\cdot)$  is leveraged in the TPNE to define the last layer of the D.

$$D_l(C_l | \mathbf{V}_l) = \sigma(f^l) = \frac{1}{1 + \exp(-f^l)} \quad (2.10)$$

where  $f^l$  is D's structure.

The log probability as the output in the loss function can indicate the real or false of the input so that the D is to maximize this loss function. In the proposed GAN, by receiving the available nodal voltage magnitudes as the conditional inputs and the generated measurements, the D updates its parameters by the cross-entropy loss function as follows.

$$\begin{aligned}\theta_{D1}^* &= \arg \max L_c \\ &= \arg \max \sum_{l=1}^L (\log D(C_l = real | \mathbf{V}_l^{real}) \\ &\quad + (1 - \log D(C_l = fake | \mathbf{V}_l^{fake})))\end{aligned}\quad (2.11)$$

The Adam gradient descent is leveraged to search for the optimal parameters  $\theta_{D1}$  of D according to the cross-entropy loss function.

$$\nabla_{\theta_{D1}} L_c = \begin{cases} \nabla_{\theta_{D1}} \log D(C_l = real | \mathbf{V}_l^{real}), & \text{if } C_l = real \\ \nabla_{\theta_{D1}} (1 - \log D(C_l = fake | \mathbf{V}_l^{fake})), & \text{if } C_l = fake \end{cases}\quad (2.12)$$

## (2) Auxiliary classifier

Different from the D which aims to distinguish whether the measurement data is the real ones or the generated measurements from the G, the AC provides the direct inferred results on the status of the branches. Since the number of branches in a distribution system is not small, and each branch has two labels, the cross-entropy loss function is also utilized in training the auxiliary classifier. Besides, the Adam gradient descent is also utilized to optimize the parameters  $\theta_{D2}$  of the AC.

$$\begin{aligned}\theta_{D2}^* &= \arg \max L_s \\ &= \sum_{l=1}^L (\log D(s = 1 | \mathbf{V}_l^{real}) + (1 - \log D(s = 0 | \mathbf{V}_l^{real}))) \\ &\quad + \log D(s = 1 | \mathbf{V}_l^{fake}) + (1 - \log D(s = 0 | \mathbf{V}_l^{fake})))\end{aligned}\quad (2.13)$$

## (3) Generator

Different from the D, the loss function of the G consists of two parts. The first part is the minimization of the log-probability conducted by the log-probability from the D fed with samples generated by G. The second part is the maximization of the log-

probability conducted by the edges with correct labels. When training the G, the parameters of the D and the auxiliary classifier are constant. The G is updated by the following equation. Note that the parameters are also updated by the Adam gradient descent.

$$\theta_G^* = \operatorname{argmin} L_g = \operatorname{argmin} L_s - L_c \quad (2.14)$$

#### (4) Regularization

Due to the limitation on the training samples, the over-fitting problem is unavoidable. To alleviate this problem, the regularization technique is introduced by adding the penalty function to the loss function [47]. In this part, the  $L_2$  norm is leveraged as the penalty item of the loss functions.

$$r(\theta) = \lambda \|\theta\|^2 \quad (2.15)$$

where  $\lambda$  is the coefficient indicating the degree of penalty on model parameters. The  $\lambda$  value controls the balance between the loss function and the objective function.

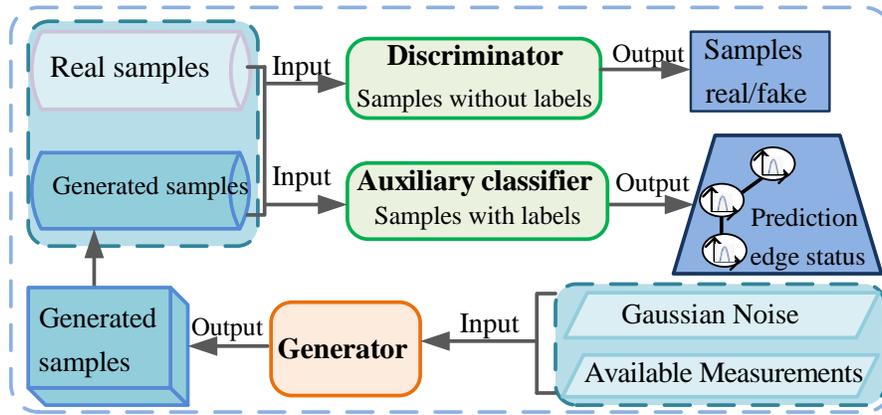


Figure 2.3. Gridtopo-GAN's procedure.

### 2.3.3 Training Procedure of Gridtopo-GAN

The training details of the training process represented in Figure 2.3 are introduced by Algorithm 1 listed as follows.

---

**Algorithm 1** Gridtopo-GAN model

---

1. Initialize: Size of batch  $m$ , dimension of embedding  $k$ ; the node embedding matrices and the edge embedding matrices of parameters of generators and discriminator with normal distribution  $N(0,1)$ ; the noises for missing data nodes with normal distribution  $N(0, 0.001)$ .

2. While Gridtopo-GAN does not converge do:

3. For D-steps do:

5. Samples  $B_S$  batch data from real data

6. Update  $\theta_{D_1}$  by running Adam algorithm by  $M_d$  times:

$$\begin{aligned}\theta_{D_1} &\leftarrow \nabla_{\theta_{D_1}} \left[ -\frac{1}{B_S} \sum_{m=1}^{B_S} \sum_{l=1}^L [\log D(C_l = \text{real} | \mathbf{V}_l^{\text{real}}) + \right. \\ &\quad \left. + (1 - D(C_l = \text{fake} | \mathbf{V}_l^{\text{fake}}))] + r(\theta) \right] \\ \theta_{D_1} &\leftarrow \theta_{D_1} - \alpha \cdot \text{Adam}(\theta_{D_1}, \mathbf{g}_{\theta_{D_1}})\end{aligned}$$

7. End for

8. For AC-steps do:

9. Samples  $B_S$  batch data from real data and  $G(\mathbf{V}_i^{\text{mv}}, z)$

10. Update  $\theta_{D_2}$  by running Adam algorithm by  $M_d$  times:

$$\begin{aligned}\theta_{D_2} &\leftarrow \nabla_{\theta_{D_2}} \left[ -\frac{1}{B_S} \sum_{m=1}^{B_S} \sum_{l=1}^L [\log D(s=1 | \mathbf{V}_l^{\text{real}}) + (1 - \log D(s=0 | \mathbf{V}_l^{\text{real}})) \right. \\ &\quad \left. + \log D(s=1 | \mathbf{V}_l^{\text{fake}}) + (1 - \log D(s=0 | \mathbf{V}_l^{\text{fake}}))] + r(\theta) \right] \\ \theta_{D_2} &\leftarrow \theta_{D_2} - \alpha \cdot \text{Adam}(\theta_{D_2}, \mathbf{g}_{\theta_{D_2}})\end{aligned}$$

11. End for

12. For G-steps do:

13. Samples  $B_S$  batch data by noise generating.

14. Update  $\theta_{G_1}$  and  $\theta_{G_2}$  according to by running Adam algorithm by  $M_g$  times:

$$\begin{aligned}\theta_G &\leftarrow \nabla_{\theta_G} \frac{1}{B_S} \sum_{m=1}^{B_S} \sum_{l=1}^L (L_s - L_c) + r(\theta) \\ \theta_G &\leftarrow \theta_G - \alpha \cdot \text{Adam}(\theta_G, \mathbf{g}_{\theta_G})\end{aligned}$$

15. End for

16. End while

---

## 2.4 Case study

The IEEE 33-node system, the 118-node system and the large 415-node distribution system are employed as the test systems to explore the performance of the Gridtopo-

GAN model [48]. The training data consists of the voltage magnitudes as the measurements and the corresponding topologies. This training data is generated by the following two steps. In the first step, the Monte Carlo simulation is introduced to randomly generate the topologies [5]. The island topologies are not included. The labels of 0/1 are utilized to denote the status off/on of the branches. For each topology, a vector consists of 0/1 as its elements are introduced as the labels of a topology. In the second step, to generate the nodal voltage magnitudes as the measurements, the power flow is calculated for each topology. Then, the IID Gaussian noises following the normal distribution as  $V_{noise} \sim N(0, \sigma_v^2)$  are introduced in the nodal voltage magnitudes. Note that the standard deviation  $\sigma_v^2$  is set in different cases in the following sections. In addition, the load demand is set according to the uniform distribution  $U(0.9, 1.1)$  for different topologies. Note that not all topologies can converge or satisfy the constraints. More than 80% of generated topologies are unqualified that are excluded. After the training data are ready, the data set sizes of the 33/118/415 are 20K, 30K, and 40K, respectively. The data sets are separated with 90% being the training data and 10% being the testing data.

#### (1) Normalization

To enhance the training effect, the normalization technique is leveraged as follows.

$$\mathbf{v}' = \frac{\mathbf{v} - \text{mean}(\mathbf{v})}{\sigma(\mathbf{v})} \quad (2.16)$$

where  $\mathbf{v}$  is the nodal voltage magnitude vector.

#### (2) Evaluation index

The F1 score is introduced as the index to evaluate the performance of the GridtopoGAN in topology identification. The F1 score is from the confusion matrix [49], which can represent the accuracy of the classification tasks. This matrix is conducted by four indexes: true positive, false positive, false negative, true negative, as shown in Table 2.1. There are two comprehensive indexes, Recall (R) and Precision (P), employed to characterize the classifier's capability on recognizing positive examples.

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \quad (2.17)$$

Table 2.1. Confusion Matrix.

Predict value	Actual value	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

It is a fact that 13.5% of the branches tend to be in off status, leading to a high P, such that P is unable to be applied. For this reason, the comprehensive index F1 score is employed to be the main index to evaluate the performance of the proposed model. A higher F1 score indicates better identification accuracy.

$$F1 = \frac{2 * P * R}{P + R} \quad (2.18)$$

(3) Hyperparameters setting:

The same learning rate setting as  $10^{-5}$  is conducted in G, D, and AC, where there are shared with the same structure. The batch size is set as 100 and the window size is set as 4. The regularization coefficient is set  $10^{-12}$ . The details of the parameters  $\theta_1$  and  $\theta_2$  are summarized in Table 2.2. These parameters are selected by optimization. The test system with the proposed model is implemented by python and performed in a computer with Intel(R) Xeon(R) CPU E5-2650 v4 @ 4.4 GHz processor and 64 GB RAM, running Windows 10.

Table 2.2. The Sizes of the Parameters.

IEEE 33- node		118-node		415-node	
$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
33 x 100	37x100	118x500	132x500	415x800	480x800

### 2.4.1 The Performance of the TPNE architecture

The D is structured by the TPNE architecture, whose performance is the basis of the proposed Gridtopo-GAN model. To explore the classification accuracy of the TPNE architecture, a classifier with three full connected neural network layers, defined as onenet and the graph convolutional network (GCN) [29]. The last layer of these classifiers is the sigmoid function that indicates the 1/0 of the branches. Note that the branch status is available in the training data set. 50 dimensions in node embedding are applied in the TPNE architecture. 50, 300 neurons are set in the middle layer of the onenet for 33/118-node systems, respectively. Besides, the 50, 300 neurons are also set in the GCN in 33/118-node systems, respectively.

Figure 2.4 shows the F1 scores from onenet and TPNE with different parameter sizes. With 50 neurons in the hidden layer, the onenet and GCN achieve F1 scores of 0.9866 and 0.9908, respectively. With 300 neurons, two corresponding models obtain 0.9932 and 0.9943 F1 scores, respectively. The TPNE obtains a better F1 score of 0.9973. This indicates that TPNE achieves higher accuracy in topology identification. The learning convergence speed is also represented in Figure 2.4. It shows that the convergence speed of TPNE is faster than onenet and GCN models even when TPNE has a smaller size on learning parameters. Besides, with the close F1 score, the sizes of learning parameters in onenet and GCN models are nearly 10 times the number of the distribution system's nodes. This fact demonstrates that a larger size of parameters is needed in the traditional classifier, especially in a large-scale system. Therefore, the proposed TPNE outperforms onenet and GCN models with smaller parameters and faster convergence speed.

Since the training data set has a significant effect on the evaluation results, it is essential to explore the impact of the training data's size on the topology identification accuracy. Thus, there are several experimental cases set with several sizes of data in 33 and 118-node systems. For the 33-node distribution system, the scales of the data set are assigned at 10K, 20K, 30K, and 40K. For the 118-node distribution system, the

scales of the data set are assigned at 10K, 20K, 30K, and 40K. Figure 2.5 presents the learning process curves. It shows that with the increase in the training data set, the topology identification F1 scores improve slightly. This phenomenon indicates that the proposed sizes in the training data set are representative enough to be used to evaluate the effectiveness of the proposed model.

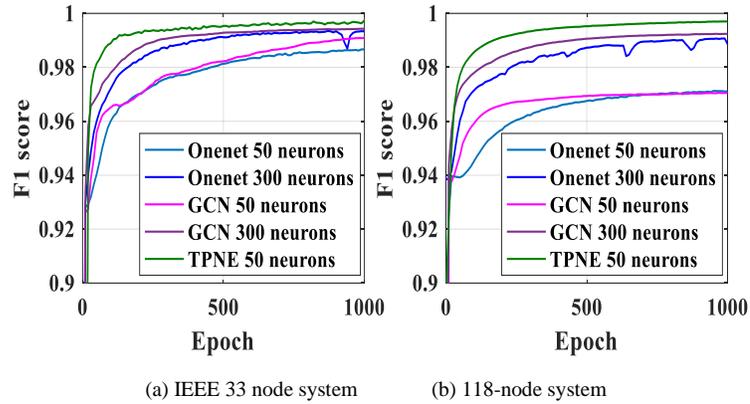


Figure 2.4. Convergence curves on onenet, GCN, and TPNE.

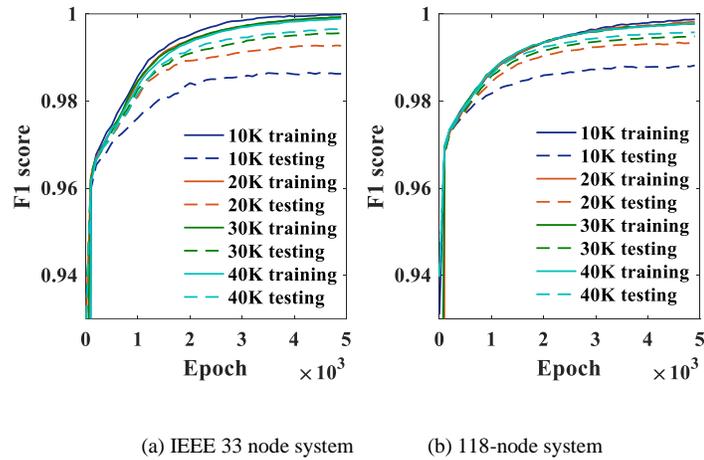


Figure 2.5. Results of different sizes of data sets.

## 2.4.2 The Performance of the Gridtopo-GAN Model

The above section explores the TPNE's effectiveness, which is the basic structure of the proposed model Gridtopo-GAN, which will be explored in the next step. The D and G play a game with each other to prompt the G to achieve the purpose in data supplement to fool the D's judgment.

In order to evaluate the Gridtopo-GAN's topology identification results with different levels of missing, five levels of data missing are set, as shown in Table 2.3 and Table 2.4 for 33-node and 118-node distribution systems, respectively. The numbers of missing data in the five cases are 6, 10, 16, 21, 26 for the IEEE 33 node system. The numbers of missing data in the five cases are 23, 43, 59, 83, 94 for the 118-node system.

(1) In comparison with conventional methods

The state-of-the-art topology identification models, Mutual Informatics (MI) [10], and the group Lasso (GL) [50], are utilized as the comparison to evaluate the Gridtopo-GAN's performance. In these methods, the inputs where the data is missing are set at zero value. Both the MI and GL are simulated 100 times. The mean values of F1 scores obtained from the MI, GL, and Gridtopo-GAN for both the 33/118-node distribution systems are listed in Figure 2.6. The F1 scores obtained from the MI and GL are less than 0.9 for all cases. This demonstrates that the MI and GL cannot be applied to the topology identification under the situation of missing data. In comparison with the MI and GL methods, the Gridtopo-GAN outperforms by at least 11%. This is due to the powerfully generative capability of the G in Gridtopo-GAN that can learn the implicit joint distribution of the measurements in the historical data so that to achieve generate the missing data conditioning on the available measurements. This demonstrates that the Gridtopo-GAN is superior in supplementing missing data for improving the topology identification accuracy.

Table 2.3. Missing Data Numbers on the IEEE 33-node System.

Measurement cases	Measurement installation node	No. of missing measurement
m6	1,2,3,5,6,7,9,10,11,13,14,15,17,18,19,20,21,23,24,25,27,28,29,30,31,32,33	6 (18.18%)
m10	1,2,3,5,7,9,10,11,13,15,17,18,20,21,23,24,25,27,28,30,31,32,33	10 (30.30%)
m16	1,3,5,7,9,11,13,15,17,18,20,21,23,25,28,30,32	16 (48.48%)
m21	1,3,5,9,11,13,17,19,23,25,29,33	21 (63.64%)
m26	1,5,11,19,23,29,33	26 (78.79%)

Table 2.4. Missing Data Numbers on the 118-node System

Measurement cases	m23	m43	m59	m83	M94
No. of missing measurement	23 (19.49%)	43 (36.44%)	59 (50.00%)	83 (70.34%)	94 (79.66%)

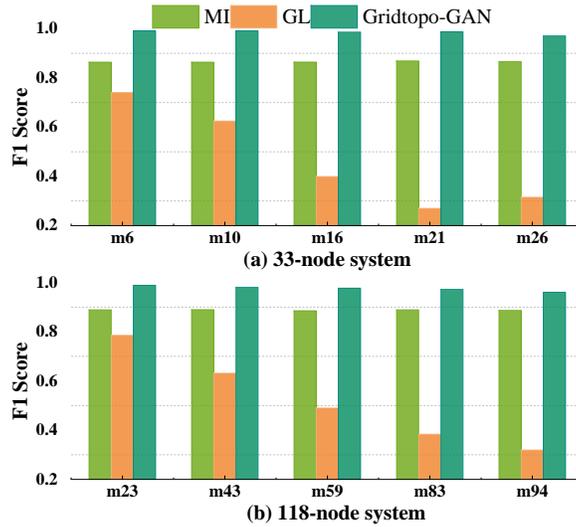


Figure 2.6. F1 scores of MI, GL, and Gridtopo-GAN.

(2) In comparison with other learning-based methods

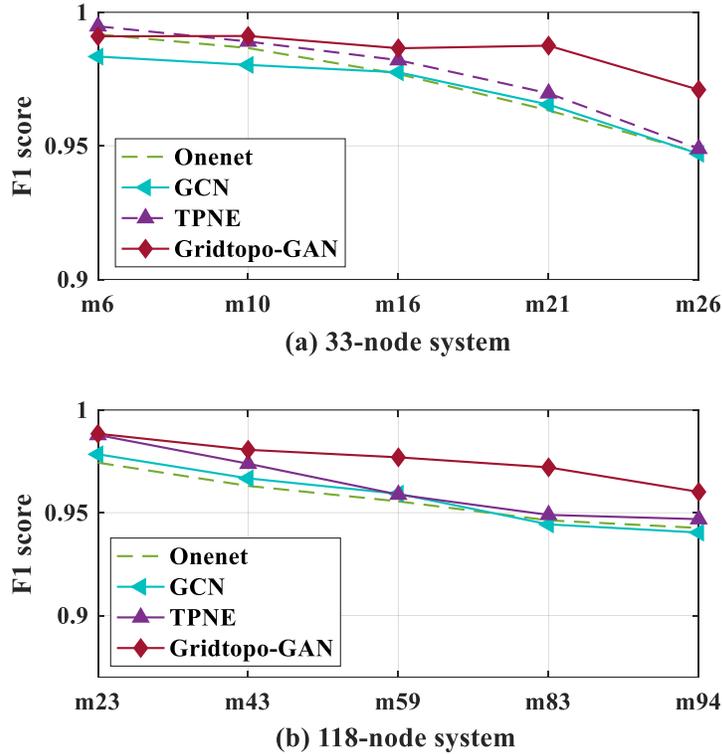


Figure 2.7. F1 scores of different learning methods with different cases of missing data.

The F1 scores obtained by Onenet, GCN, TPNE, and Gridtopo-GAN are shown in Figure 2.7 for IEEE 33 and 118-node distribution systems, respectively. In these cases, except for the missing data, 5% noise is added to the measurements of each case. In comparison with the MI in Figure 2.6, the F1 scores obtained by onenet, GCN, and TPNE are more than 0.9 for all missing data cases. However, with missing data increases, the F1 scores of onenet, GCN, and TPNE degrade rapidly to less than 0.95, especially when missing data reaches three-quarters. This demonstrates that the classifiers, onenet, GCN and TPNE, cannot adapt to topology identification well with missing data. For all situations of missing data, F1 scores of the Gridtopo-GAN model are still above 0.97 for IEEE 33 system and 0.96 for the 118-node system, respectively. This result indicates that Gridtopo-GAN is highly outperforming other classifiers in

identifying topology with incomplete measurements.

(3) With the different locations of measurements

The location of the measurements can pose an impact on the topology identification results. To evaluate this factor in the proposed model, different locations of measurements are set. There are ten cases with 26 missing measurements and 5% noise added. Note that the locations of the available measurements are randomly selected. The F1 scores obtained by Gridtopo-GAN with different combinations of measurement locations are shown in Figure 2.8. It shows that the F1 scores are 0.97 for all cases. This indicates that the Gridtopo-GAN has little impact on the locations of measurements.

Furthermore, the calculating time for topology identification in different learning methods is shown in Table 2.5. It shows that the time consumptions of Gridtopo-GAN are 0.17ms and 5.68ms for 33/118-node systems, respectively. The MI requires 8.56s and 222.02s for these two systems, which is time-consuming due to tremendous enumerating behavior. The onenet has a less computation time than Gridtopo-GAN due to only one neural network being required while Gridtopo-GAN requires several levels of neural networks. However, the Gridtopo-GAN achieves higher topology identification accuracy and the computation time of close to a few microseconds can be neglected in practice. Thus, the Gridtopo-GAN achieves a more prominent performance in topology identification under missing data situations.

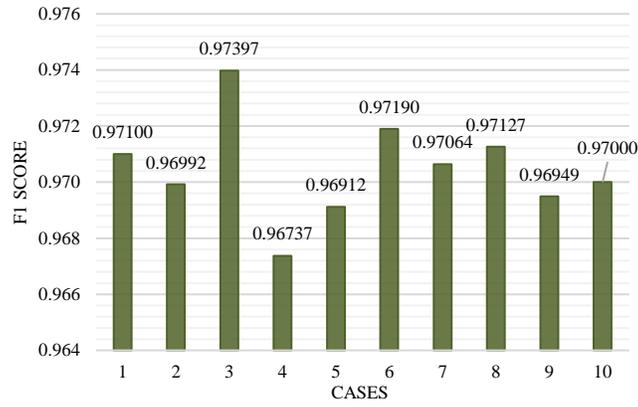


Figure 2.8. The F1 scores from ten cases of the measurement locations.

Table 2.5. Computation Time.

Methods	MI	Onenet	Gridtopo-GAN
<b>IEEE 33</b>	8.56s	0.04633ms	0.1658ms
<b>118-node</b>	222.02s	0.1431ms	5.684ms

### 2.4.3 Topology Identifications with Loops

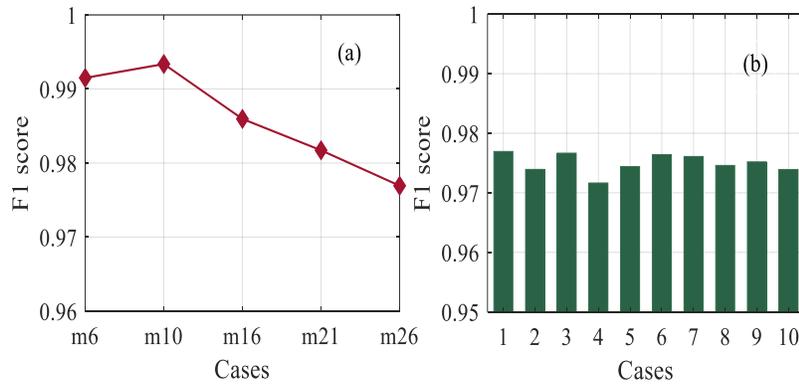


Figure 2.9. (a) F1 scores of different missing data situations with mixed topology. (b) The F1 scores of different measurement locations with mixed topology.

The distribution system topology in the cases discussed above is set radial. However, in some distribution systems in the urban area, the topology is not only radial but also consists of loops, defined as meshed topology. Since the meshed topology cannot be ignored in the modern distribution systems, the proposed Gridtopo-GAN is also conducted in such a topology.

There are up to five loops in the IEEE 33-node distribution system. According to the training data generation steps mentioned above, 17000 meshed topologies are randomly generated. The loop's number in these topologies is from one to five. Then, the radial topologies are added to these meshed topologies. Finally, a 20000 training data set is obtained, defined as mixed topology. In the 33-node system, the minimum window nodes' number in a loop is 7. The window size is set 4 in this section. Thus, the maximum window nodes' number in the loops within the mixed topology is 9. In this setting, the range of the window nodes contains loops so that this setting can be utilized to verify the effectiveness of the Gridtopo-GAN in the mixed topology identification. The F1 score obtained by Gridtopo-GAN with complete measurements is 0.993. This indicates that the Gridtopo-GAN can adapt to the distribution system topology identification problem with loops.

The F1 scores obtained by Gridtopo-GAN with different missing data cases and measurement locations using the mixed topology for the IEEE 33-node system are depicted in Figure 2.9. It shows that Gridtopo-GAN achieves high F1 scores in different levels of missing data and different measurement locations with mixed topology. This indicates that the distribution system structure with loops brings little impact in the proposed Gridtopo-GAN.

#### **2.4.4 Results with Different Levels of Noises**

Because the measurement errors are unavoidable, there are five levels of noises set as the cases as  $\pm 0.1\%$ ,  $\pm 0.5\%$ ,  $\pm 1\%$ ,  $\pm 2.5\%$ , and  $\pm 5\%$  [51]. For different cases, these noises are added so that there are five cases to be utilized to verify the effectiveness of

the proposed model.

As mentioned above, the TPNE is the basic structure of the Gridtopo-GAN, its learning performance plays an important role in the topology identification results of the proposed model. Thus, the TPNE is firstly conducted with complete data under different levels of noise and is listed in Table 2.6 and Figure 2.10(a). Then, with the missing data cases m26's setting, the performance of the Gridtopo-GAN is verified and the corresponding F1 scores are depicted in Figure 2.10.

As it is shown in Table 2.6 and Figure 2.10(a), F1 scores obtained by the TPNE are more than 0.99 with all cases in different levels of noise. This result indicates that the TPNE can infer the status of the branches with high accuracy and at the same time hedge against noises. Since the TPNE obtains a significant topology identification accuracy against noise, the Gridtopo-GAN's performance with different noise under the case m26 is further verified in Figure 2.10(b). The F1 scores obtained from the Gridtopo-GAN are more than 0.97 for all noise settings. This indicates that the Gridtopo-GAN can handle the noise measurements even with limited available measurements.

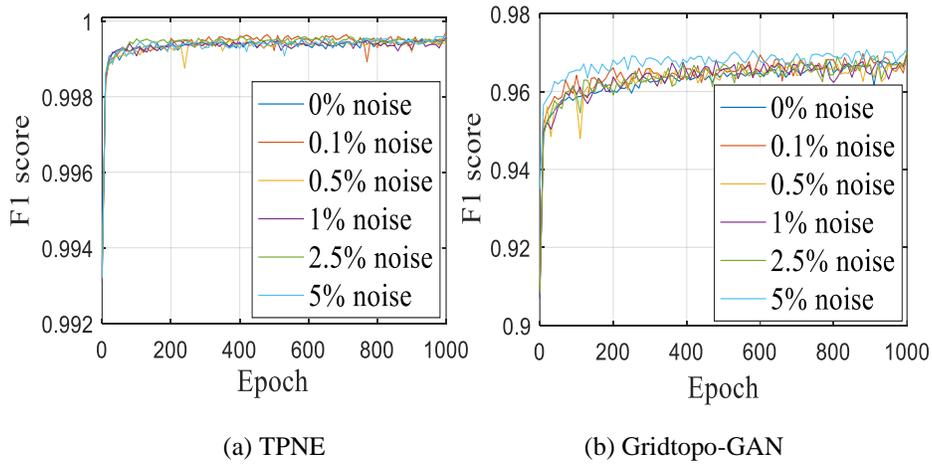


Figure 2.10. The F1 scores from TPNE and Gridtopo-GAN under different noises.

Table 2.6. F1 Scores of TPNE with complete measurements and different noises.

Noise	0%	0.1%	0.5%	1%	2.5%	5%
F1 Score						
TPNE	0.9996	0.9997	0.9996	0.9996	0.9996	0.9996

### 2.4.5 Applying Gridtopo-GAN in Large Scale Distribution System

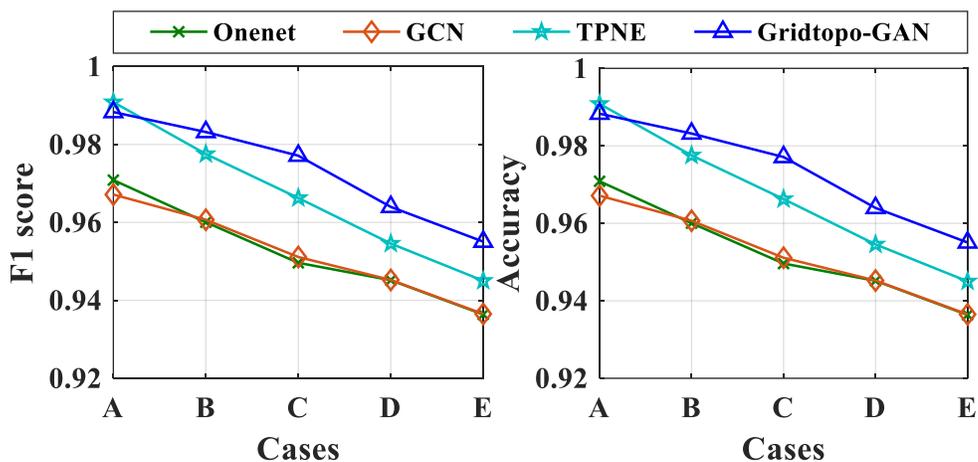


Figure 2.11. The F1 score and accuracy from the 415-node system with different levels of missing data.

Since the scale of the modern distribution system is relatively large with the boost of load demand, the Gridtopo-GAN is conducted in a 415-node distribution system [48], where 415 nodes and 480 lines with 65 tie lines are included. The number of missing measurements is set to be 20%, 35%, 50%, 65%, and 80%, which is defined as the cases as Case A, Case B, Case C, Case D, and Case E. The F1 score and accuracy obtained by Gridtopo-GAN are shown in Figure 2.11. This picture shows that the proposed Gridtopo-GAN outperforms other methods for all cases. This is because the TPNE architecture is designed to be able to extensibility to large-scale systems. In this way, the Gridtopo-GAN can still achieve a higher F1 score and accuracy in a large-

scale distribution system effectively without the impact of the system scales.

## **2.5 Summary**

The Gridtopo-GAN is innovatively proposed to adapt to the topology identification problem in the distribution system under the challenges of limited measurements and with radial/meshed topologies. To this end, firstly, by proposing the topological Markov Blanket, both the radial and meshed topologies can be inferred effectively using the measurements within the window. Then, by leveraging the graph embedding techniques with the topological Markov Blanket, the TPNE model is proposed, which compacts the numerous changing topologies in a low dimension space so as to achieve the high topology identification accuracy in a large-scale system. Finally, the Gridtopo-GAN allows the generator to complement missing data so that nearly more than 10% in improved topology identification F1 score can be achieved under the limited measurement situations. Furthermore, the impact of the measurement locations is excluded by the simulation results. In a word, the Gridtopo-GAN has a significant performance in topology identification for distribution systems with efficiency, and robustness against data scarcity.

# *Chapter 3 Unrolled Spatiotemporal Graph Convolutional Network for Distribution System State Estimation and Forecasting*

## **3.1 Introduction**

The proliferation of renewable energy brings significant uncertainty to active distribution systems due to its intermittent and stochastic nature. Such a new scene in the distribution system unavoidably introduces challenges to the reliable operation of power systems. To consider the impact of uncertainty on distribution system operation, the DSSE gains many academic eyes due to it plays a distinguishing role in the real-time awareness of system states for delivering enhanced monitoring, control, and management functionalities [52], [53]. Thus, it arouses a new model in DSSE to achieve more effective and efficient state estimation and extend the application in modern distribution systems.

Generally, an active distribution system covers a relatively small area but more integrated dispersed RES and limited measurements. In this situation, the renewable energy deployment sites are close to each other, sharing similar weather and environment, leading to similar profiles generated by these renewable energy units [6].

Thus, there are significant complex correlations among nodal power injections, resulting in hidden spatiotemporal correlations among nodal measurements of distribution systems [7]. More specifically, these correlations consist of three aspects as introduced in the following. The first one is the spatial correlation. It means that at the same time slot, the measurements in a node are affected by its neighbor nodes. The second one is the temporal correlation. It represents a fact that the measurements of a node are influenced by itself across different time periods. The third one is the spatiotemporal correlation. It means a node is affected by its neighbor nodes across time slots. These correlations pose inevitable bias in the system state estimation calculations and thus remain a significant challenge to be addressed [2]. However, many previous works in DSSE ignore such a gap due to no immature is available to fill it. Thus, it is an extremely urgent task to explore a suitable approach to DSSE and state forecasting with the proliferation of RES.

The weighted least square (WLS) is the most classic solution in DSSE. It is a nonconvex optimization problem that minimizes the residual between the measurement and the estimated states, where the optimal searching process is based on the Gauss-Newton algorithm [54]. In this method, the correlation among different measurements is ignored and thus its application is limited to the scenarios without correlations. However, since the correlations are unavoidable in the distribution system, the WLS-based method will introduce considerable error in DSSE [4]. To address this issue, the WLS is ameliorated by replacing the weighting matrix with the correlation coefficient matrix, leading to improved accuracy [12]. Similarly, the covariance matrix is leveraged in [55] to degrade the impact of the correlation. The correlation among nodal measurements is employed to construct the weighting matrix in [56] for ameliorating the estimation accuracy. However, these correlations are assumed linear and time-invariant. To remove this assumption, an artificial neural network is introduced in [13] to get the correlation coefficient matrix in the WLS for microgrid state estimation. However, the temporal correlations are not taken into account in these works. Without spatial correlations considered, the violent fluctuations on RES and their impact on

DSSE are neglected, leading to unsatisfactory accuracy.

According to the review above, the temporal correlations are also important in facilitating the DSSE. Inspired by his, the valuable information hidden in the historical data is leveraged to be extracted to ameliorate the conventional DSSE methods. The RES's temporal correlations are utilized to improve the accuracy of the EKF-based DSSE method in [57]. Similarly, the Complex Kalman filter (CKF) in [58] is leveraged to characterize the changes in time-series measurements so that the state estimation errors can be reduced. However, only a single type of correlation, either spatial or temporal correlation on measurements is considered in these works. These approaches limit their applications in DSSE with spatiotemporal correlations virtually existing.

There are few works presented to deal with this problem. The Conditional multivariate complex Gaussian distribution is leveraged in [14] to describe the spatiotemporal correlations among RES unbalance loads to ameliorate the DSSE accuracy, requiring comprehensive information on the correlations which is rarely common in practice. A vector auto-regressive approach is introduced in [15] to describe the spatial and temporal correlation between the demands and RES to facilitate the DSSE to achieve enhanced forecasting accuracy. However, the spatial and temporal correlations are considered individually, which is inappropriate since the coupled spatiotemporal correlations must be considered. The vector autoregressive is also introduced in [59] to forecast the voltage phase angles according to the assumption of linear correlations. However, it cannot take the nonlinear dynamic characters of system states into account, and thus leading to insufficient precision results. Besides, these methods are introduced on top of the traditional WLS method, which is computationally cumbersome.

To ameliorate these challenges, the learning-based approaches are introduced to the DSSE and state forecasting in recent years. The physics-guided neural network is developed in [60] to conduct a new state estimation model with temporal correlations considered. Similarly, the physics-inspired deep learning model is introduced in [61] to address the real-time state estimation on transmission systems, leading to enhanced

accuracy. Besides, a physics-aware neural network is leveraged in [62] to speed up the state estimation. A deep learning model is introduced in [63] to degrade the computational burden. However, these methods do not consider the spatiotemporal correlations which cannot adapt to a distribution system with violent fluctuations on RES. With the proliferation of smart meters, rich measurement data is available in the modern distribution system. Based on this opportunity, the potentially powerful learning capability of deep-learning techniques can be leveraged to extract useful information from numerous measurements so as to enhance the accuracy of the DSSE with spatiotemporal correlations. The spatial-temporal graph convolutional network (STGCN) has attracted more and more attention in different fields, such as traffic forecasting [64], social pedestrian behavior prediction [65], skeleton-based action recognition [66], and so on. They capture the correlations from the data to achieve enhanced performance. However, these works treat the spatial and temporal correlations separately, which ignore the spatiotemporal correlations and may lead to unsatisfactory results. To address this issue, a spatial-temporal synchronous graph convolutional network is proposed for network data forecasting to consider the spatiotemporal simultaneously [67]. The graphical structure properties of the power system and the spatial-temporal graph convolutional network techniques inspires a new way to consider the correlations' impact on distribution system state estimation. Therefore, it is unprecedentedly essential to leverage the spatiotemporal correlations to enhance the real-time state estimation performance with simultaneous spatiotemporal correlations considered. In this perspective, the spatiotemporal correlations can serve as helpful information rather than the challenges in the conventional DSSE.

To solve these problems, the unrolled spatiotemporal graph convolutional network model is proposed in this work to be better characterize of the distribution system states that are exposed to complex correlations among the renewable power outputs. Instead of considering the spatial and temporal correlations separately in the previous works, three aspects of spatiotemporal correlations are captured simultaneously by the proposed unrolled spatiotemporal graph model that leveraged the splicing of the spatial

graphs across adjacent time steps. In this way, ameliorated forecasting accuracy and computational efficiency can be achieved. Then, the node embedding is leveraged to construct the dependence on the unrolled spatiotemporal graph to learn the nonlinear spatiotemporal correlations automatically instead of utilizing the linear correlation coefficient matrix. Further, the ahead-of-horizon state forecasting is achieved effectively by the multi-module layers that capture the long-range spatiotemporal correlations. The experiments are conducted on a 118-node distribution system and a 1746-node distribution system.

The remaining sections are organized as follows. The state estimation considering the spatiotemporal correlations is developed in Section 3.2. The USGCN model is introduced both for DSSE and state forecasting in Section 3.4. The experimental results and the corresponding discussions are introduced in Section 3.5. Finally, the works of this section are concluded in Section 3.6.

### 3.2 State Estimation with Spatiotemporal Correlations

The basic equation for DSSE is introduced as follows.

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

In (3.1),  $z$  denotes the measurements that are collected by the devices, such as smart meters.  $h$  represents the measurement function that maps the actual state variables  $x$  to the measurements.  $e$  is the errors between actual measurements and the estimated measurements. The traditional state estimation can be built based on the WLS optimization.

$$\hat{x} = \arg \min_x (z - \mathbf{h}(x))^T \mathbf{W} (z - \mathbf{h}(x)) \tag{3.2}$$

In (3.2),  $\hat{x}$  is the estimated state variables.  $T$  denotes the matrix transposition operation.  $w$  is the weight matrix. This model can only adapt to the state estimation with independent Gaussian measurements. Thus, it is inappropriate in the modern

distribution system when implicit correlations are considered [68]. Motivated by this gap, without the prior Knowledge on correlations, the Bayesian rule is leveraged to model the posterior distribution for estimating states conditioning on measurements.

$$f(\mathbf{x}/\mathbf{z}) = \frac{f(\mathbf{z}/\mathbf{x})f(\mathbf{x})}{f(\mathbf{z})} \quad (3.3)$$

In (3.3),  $f(\mathbf{x}/\mathbf{z})$  denotes the posterior probability density function (PDF) of the state variables.  $f(\mathbf{z}/\mathbf{x})$  denotes conditional PDF of the measurements, which is generally commutated by the maximum likelihood function of the state variables under the given measurements set.  $f(\mathbf{x})$  represents the prior distribution of the state variables.  $f(\mathbf{z})$  represents the distribution of the measurements. In this model, the  $f(\mathbf{z}/\mathbf{x})$  is difficult to obtain due to the complex distribution for RES outputs brought by their uncertain and intermittent nature. Moreover, obtaining  $f(\mathbf{z})$  needs massive historical measurement data, which is rarely commonly available due to the low observability of the distribution system. These challenges motivate the practice of posterior conditional distribution-based estimation states. Specifically, conditioning on the measurements  $\mathbf{z}$ , the estimation state  $\hat{\mathbf{x}}$  is equivalent to the expectation of the posterior conditional distribution  $f(\mathbf{x}/\mathbf{z})$ .

$$\hat{\mathbf{x}} = E[\mathbf{x}/\mathbf{z}] = \int \boldsymbol{\alpha} f_{\boldsymbol{\alpha}|\mathbf{z}}(\boldsymbol{\alpha}|\mathbf{z}) d\boldsymbol{\alpha} \quad (3.4)$$

In (3.4), the integral operation is performed in the whole state space.  $\boldsymbol{\alpha}$  denotes the integration variable that can span all the values  $\mathbf{x}$ . In this way, the expectation of this posterior distribution can be utilized to replace the state estimation, which is formulated as follows.

$$\hat{\mathbf{x}} = E[\mathbf{x}/\mathbf{z}] = \mathbf{h}(\mathbf{z}) \quad (3.5)$$

In (3.5),  $\mathbf{h}$  denotes the several layers  $[\mathbf{h}^0, \dots, \mathbf{h}^l, \dots, \mathbf{h}^{out}]$  of the proposed USGCN model.

### 3.3 Unrolled Spatiotemporal Graph Convolution Networks

#### 3.3.1 Unrolled Spatiotemporal Graph

Conventionally, the previous models introduced the convolutional neural network (CNN) into the recurrent neural network (RNN) to capture the spatial and temporal correlations across time steps, while CNN can only feature the Euclidian data structure [69]. The graph neural network (RNN) is combined with the RNN to improve the non-Euclidian data feature extraction [70]. However, either the spatial or temporal correlations are singly considered, resulting in neglecting the correlations across time between different nodes. Thus, an unrolled spatiotemporal graph model is developed to directly extract the spatial and temporal correlations simultaneously. Specifically, the effect from the neighbors on each node within the current and the adjacent time steps can be perceived synchronously, which can be illustrated in Figure 3.1.

To characterize the correlations of the nodal measurements in a graphical format, the bus in the distribution system is defined as nodes, and the correlations among these nodes are defined as edges. In this way, a graph  $\mathcal{G}=(\mathcal{V},\mathcal{E})$  can denote the spatial correlations among nodes, where  $\mathcal{V}=\{v_1,\dots,v_i,\dots,v_N\}$  is the set of all nodes with the number of  $N$  and  $\mathcal{E}$  is the set of edges with the number of  $L$ .

To characterize this graph quantitatively, the adjacent matrix  $A \in \mathbf{R}^{N \times N}$  is developed.

$$A_{ij} = \begin{cases} 1, & \text{if } v_i, v_j \in \mathcal{V}, \text{ and } (v_i, v_j) \in \mathcal{E} \\ 0, & \text{else} \end{cases} \quad (3.6)$$

Since the aim is to model spatial and temporal correlations at the same time across different time slots, the graphs in different time slots are connected chronologically, as shown in Figure 3.1(b). By connecting the graph of all the nodes belonging to the previous, current, and next time slots, a spatiotemporal graph is obtained, which is

denoted by the adjacent matrix  $\tilde{A} \in \mathbb{R}^{3N \times 3N}$ . For each node  $i$  in the spatial graph, its new index can be calculated by  $(t-1)N+i$ , where  $t$  ( $1 \leq t \leq 3$ ) indicates the time step number in the spatiotemporal graph model. In Figure 3.1(b), the elements  $A_{ij}$  in the  $A_{adp}^t$  present the spatial correlation between node  $i$  and node  $j$  in time step  $t_1$ . The diagonal elements  $A_{ii}$  in  $A_{adp}^{(t_1-t_2)}$  denote the temporal correlation of node  $i$  across the time steps  $t_1$  and  $t_2$ . The off-diagonal elements  $A_{ij}$  in  $A_{adp}^{(t_1-t_2)}$  denote the spatiotemporal correlation of node  $i$  and node  $j$  across the time steps  $t_1$  and  $t_2$ . In a word, the unrolled spatiotemporal graph across three continuous time steps. The diagonal adjacency matrices denote the spatial correlation of each node with others in the current time step. The diagonal elements of the off-diagonal adjacency matrix represent the spatial correlation of each node with itself. The non-diagonal elements of the off-diagonal adjacency matrix represent the spatiotemporal correlation of each node with other nodes across time steps. In this model, the spatiotemporal correlations between neighbors can be extracted directly.

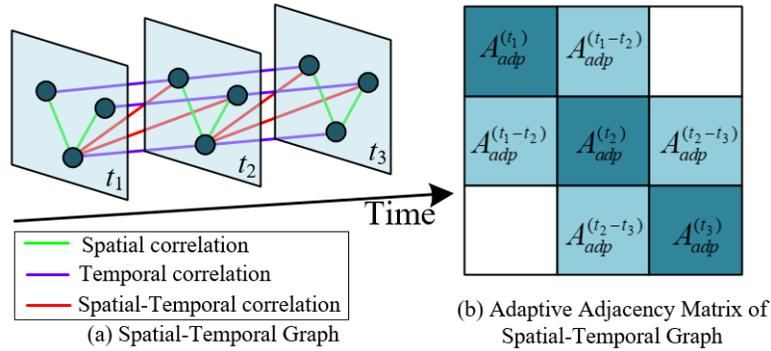


Figure 3.1. Unrolled Spatiotemporal Graph Structure.

### 3.3.2 Self-adaptive Unrolled Spatiotemporal Graph

It is important to build up a graph model to describe the adjacent matrix that

represents the correlations. Traditionally, the historical nodal measurements can be used to model the deterministic spatial correlations by using the static correlation coefficient matrix. However, the spatial correlations and the temporal correlations are nonlinear and vary over time. It is not appropriate to utilize deterministic values to describe these correlations. Therefore, it is an urgent task to develop a new model in DSSE to achieve more effectively and efficiently in modern distribution systems with spatiotemporal correlations.

Traditionally, the fixed spatial correlation is characterized by the adjacency matrix  $A \in \mathbb{R}^{N \times N}$ . Inspired by the node embedding technique that is powerful in graph representation, it is leveraged in this work to feature the spatiotemporal dependence among nodes. To this end, a self-adaptive adjacency matrix  $\tilde{A}_{adp} \in \mathbb{R}^{3N \times 3N}$  is proposed for each time step without the requirement of any prior knowledge. The hidden complex spatial correlations can be discovered automatically through this model during the training process. This self-adaptive adjacency matrix is defined by the dot-product by two node embeddings parameterized via  $E_1, E_2 \in \mathbb{R}^{3N \times k}$ , where  $k$  denotes the dimension of embedding.

$$\tilde{A}_{adp} = SoftMax(ReLU(E_1 E_2^T)) \quad (3.7)$$

In (3.7),  $E_1$  and  $E_2$  present the parameters in the corresponding two nodes. The dot-product of  $E_1$  and  $E_2$  represents the spatial dependence between two nodes, where the weak connectivity is removed by the active function ReLU. The activation function SoftMax is leveraged to normalize the self-adaptive adjacency matrix.

### 3.3.3 Unrolled Spatiotemporal Graph Convolution Operation

The general convolutional neural network can only adapt to the data with Euclidean pixel structure. Different from this, the graph convolutional neural network can be utilized in the data with a non-Euclidean graph structure [71]. Specifically, this kind of neural network structure can aggregate the neighbor's information for any node so that

efficiency can be achieved [72].  $\tilde{\mathbf{A}}^{(n)} \in \mathbb{R}^{3N \times 3N}$  is defined to represent the adjacency of the spatiotemporal graph model.  $\mathbf{h}^{(l-1)} \in \mathbb{R}^{3N \times D}$  denotes the inputs of the  $l$ -th layer in the proposed model, where  $D$  is the feature size of the input.  $\mathbf{h}^{(l)} \in \mathbb{R}^{3N \times D}$  denotes the output of the  $l$ -th layer in the proposed model.  $\mathbf{W} \in \mathbb{R}^{D \times D}$  is the learnable parameters.  $\mathbf{b} \in \mathbb{R}^D$  is the bias.  $\sigma$  is the activation function. The proposed graph convolutional operation is formulated as follows.

$$\mathbf{h}^{(l)} = USGCN(\mathbf{h}^{(l-1)}) = \sigma(\tilde{\mathbf{A}}^{(n)}\mathbf{h}^{(l-1)}\mathbf{W} + \mathbf{b}) \in \mathbb{R}^{3N \times D} \quad (3.8)$$

The diffusion convolutional operation in [73] is popular for its brilliant power in spatial modeling. Generally, the diffusion convolutional operation is formulated as:

$$\mathbf{h}^{(l)} = \sum_{m=0}^M \mathbf{P}^m \mathbf{h}^{(l-1)} \mathbf{W}_m \in \mathbb{R}^{3N \times D} \quad (3.9)$$

In (3.9),  $\mathbf{P}^m$  denotes the power series of the transition matrix. Since the nodal measurement data can be modeled as an undirected graph, such that  $\mathbf{P}^m = \tilde{\mathbf{A}} / \text{rowsum}(\tilde{\mathbf{A}})$ , where  $\text{rowsum}$  compute column sums across rows of  $\tilde{\mathbf{A}}$ . By combining the above self-adaptive spatiotemporal graph model and the diffusion convolutional operation, the innovative spatiotemporal graph convolutional layer can be defined as follows.

$$\mathbf{h}^{(l)} = \sum_{m=0}^M \mathbf{P}^m \mathbf{h}^{(l-1)} \mathbf{W}_{m1} + \tilde{\mathbf{A}}_{\text{adp}}^{(n)} \mathbf{h}^{(l-1)} \mathbf{W}_{m2} \in \mathbb{R}^{3N \times D} \quad (3.10)$$

Note that in the situation where the graph is unknown, only the second term remains in the spatiotemporal graph convolutional layer.

$$\mathbf{h}^{(l)} = \sum_{m=0}^M \tilde{\mathbf{A}}_{\text{adp}}^{(n)} \mathbf{h}^{(l-1)} \mathbf{W}_{m2} \in \mathbb{R}^{3N \times D} \quad (3.11)$$

### 3.3.4 Gated Operation

The gated linear unit (GLU) is a representative operation in the deep learning methods. It is popular for its powerful ability to control whether the information is

being flowed into the next layer or not [74]. This operation is combined with the proposed spatiotemporal graph convolutional operation, and it is formulated as follows.

$$\begin{aligned} \mathbf{h}^{(l)} &= \left( \sum_{m=0}^M \mathbf{P}^m \mathbf{h}^{(l-1)} \mathbf{W}_{m1} + \tilde{\mathbf{A}}_{adp}^{(n)} \mathbf{h}^{(l-1)} \mathbf{W}_{m2} + \mathbf{b}_1 \right) \\ &\otimes \text{sigmoid} \left( \sum_{m=0}^M \mathbf{P}^m \mathbf{h}^{(l-1)} \mathbf{W}_{m1} + \tilde{\mathbf{A}}_{adp}^{(n)} \mathbf{h}^{(l-1)} \mathbf{W}_{m2} + \mathbf{b}_2 \right) \in \mathbb{R}^{3N \times D} \end{aligned} \quad (3.12)$$

In (3.12),  $\mathbf{W}_1 \in \mathbb{R}^{D \times D}$  and  $\mathbf{W}_2 \in \mathbb{R}^{D \times D}$  are the learnable parameters.  $\mathbf{b}_1 \in \mathbb{R}^D$  and  $\mathbf{b}_2 \in \mathbb{R}^D$  are the bias parameters. *sigmoid* denotes the sigmoid activation function ( $\text{sigmoid}(x) = 1/(1 + e^{-x})$ ).  $\otimes$  represents the element-wise product. Note that by multiply stacking the proposed spatiotemporal convolutional layers, the receptive range of the time-series data can be achieved.

## 3.4 USGCN Model

### 3.4.1 USGCN for State Estimation

The historical measurement data with spatial and temporal correlations can be used to facilitate the DSSE. The proposed USGCN is utilized to learn the expectation of the posterior conditional distribution (Eq. (3.5)) that accounts for the spatial and temporal correlations hidden in the measurement data. To achieve this goal, the proposed unrolled spatiotemporal graph convolution operation and the gated operation are utilized to construct the USGCN for state estimation and state forecasting.

The framework of USGCN for state estimation is shown in Figure 3.2(a). This framework consists of a USGCN layer, three CNN layers, and a fully connected neural network layer as the output layer. The USGCN layer is utilized to capture the spatiotemporal correlations at the beginning. The CNN layers are employed to further capture the hidden features within the measurement data.

The measurement data includes the active and reactive power injections, denoted

by  $\mathbf{Z} \in \mathbb{R}^{N \times C \times T}$ , where  $N$  is the number of nodes.  $T$  is the sequence length.  $C$  is the hidden dimension, which presents the measurement data in each node, e.g. four dimensions include the nodal active/reactive power injection and active/reactive line flow. Note that the measurement data includes the active/reactive power injections and the active/reactive line power flows, which are available from smart meters. Since the estimated states are the voltage magnitudes and the phase angles, the number of the measurements are satisfied with the observability without full deploying instruments in all nodes, e.g. above 50% of the system node number. Generally, the branch number in the distribution system is less than the node number due to the radial structure so that the first dimension  $N$  is enough for nodal and line measurements. Note that the nodal and line input without the measurements are replaced with zero.

The measurement data  $\mathbf{Z} \in \mathbb{R}^{N \times C \times T}$  is then utilized to extract the slices of unrolled spatiotemporal graphs' series as  $\mathbf{h}_s^0 \in \mathbb{R}^{N \times C \times 3}$  which can be reshaped  $\mathbf{h}_s^0 \in \mathbb{R}^{3N \times C \times 1}$ . Instead of utilizing the fixed value-based adjacency matrix as the aggregate weight between nodes in the conventional graph convolutional operation, the node embedding product is leveraged to be the aggregate function. This aggregate function features the dependence among nodes according to the information extracted from the measurements. After this setting, the slices of T-2 unrolled spatiotemporal graphs across different time slots can be obtained. By splicing these spatiotemporal graphs by row, the input  $\mathbf{h}^0 \in \mathbb{R}^{3N \times C \times (T-2)}$  of USGCN can be obtained.

Since the input and the USGCN are defined above,

Based on the definition of the input and the USGCN, the unrolled spatiotemporal graph convolutional layer's output is defined as  $\mathbf{h}^l \in \mathbb{R}^{3N \times C' \times (T-2)}$ , where  $C'$  is the hidden dimension of the layer  $l$ . To obtain the target channel dimensional of features in each node, the 2-d convolutional neural network layers are also utilized to transfer the  $\mathbf{h}^l \in \mathbb{R}^{3N \times C' \times (T-2)}$  into  $\mathbf{h}^l \in \mathbb{R}^{3N \times C'' \times (T-2)}$ . In addition, the fully connected network is employed to transfer the  $3N$  dimensional  $N_{out}$  so that the target output dimensional can be achieved. Therefore, the output of the proposed USGCN is  $\mathbf{h}^{out} \in \mathbb{R}^{N_{out} \times C'' \times (T-2)}$ .  $C''$  is the

hidden dimension of the layer, e.g.  $C''=2$  means two dimensions of channels, presenting the voltage magnitude and the voltage phase angle.  $N_{out}$  is the number of voltage magnitudes or phase angles.

Due to the less sensitive ability of the Huber loss [75], which is leveraged to be the loss function.

$$Loss(\hat{\mathbf{h}}, \mathbf{h}) = \frac{1}{TNC} \sum_{i=1}^N \sum_{j=4}^{j=T} \sum_{k=1}^{k=C} \begin{cases} 0.5(\hat{\mathbf{h}}_{ik}^{(t+j)} - \mathbf{h}_{ik}^{(t+j)})^2, \\ \delta |\hat{\mathbf{h}}_{ik}^{(t+j)} - \mathbf{h}_{ik}^{(t+j)}| - 0.5\delta^2 \end{cases} \quad (3.13)$$

In (3.13),  $\hat{\mathbf{h}}$  is the outputs of the model.  $\mathbf{h}$  denotes the corresponding actual values.  $\delta$  represents the threshold that controls the range of squared error loss.

### 3.4.2 USGCN for State Forecasting

Generally, the predicted states are corrected relying on complete real-time measurements. Then the corrected states are utilized to predict the next single-step system states [61], such that the predicted states rely on the accuracy of the previously estimated states. Instead of utilizing the previously estimated states, the proposed method can directly forecast the system states based on the previous measurements. In this way, the proposed method can fully leverage the features of the spatial and temporal correlation in measurements to forecast the states directly with available measurements so that the improved awareness of system states ahead of time can be achieved.

The framework of USGCN for state forecasting is shown in Figure 3.2(b). Different from the USGCN for DSSE, it consists of two stacked spatiotemporal graph convolutional layers, three CNN layers, and a fully connected neural network layer. When the several spatiotemporal graph convolutional layers are stacked, the proposed USGCN can perceive spatiotemporal correlations across adjacent time steps as well as more than adjacent time steps.

The previous works in [69] require  $T$  time steps to generate the time-series  $\mathbf{h}'$ .

Different from this, the proposed model directly generates the entire  $\mathbf{h}^{(t+1):(t+T)}$ . Specifically, to this end, the temporal dimension in the spatiotemporal graph convolutional layers is set at  $T-3$ . In this way, the input of the USGCN for state forecasting is  $\mathbf{h}^0 \in \mathbb{R}^{3N \times C \times (T-3)}$  and the output is  $\mathbf{h}^t \in \mathbb{R}^{3N \times C \times (T-3)}$ .

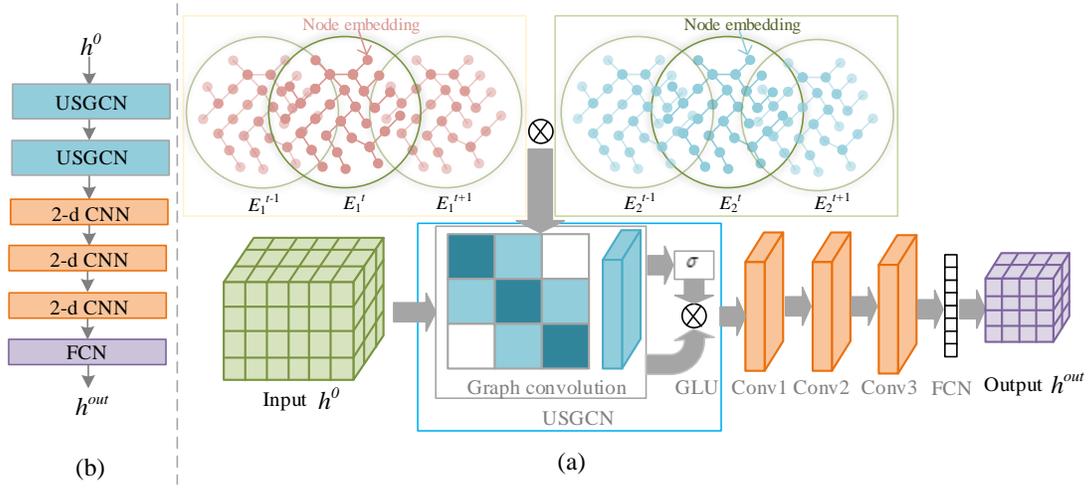


Figure 3.2. (a) The structure of the USGCN for DSSE. (b) The structure of the USGCN for state forecasting.

### 3.5 Case study

To verify the effectiveness of the proposed model, an 11kV 118-node distribution system [76], and a 130.8kV 1746-node distribution system [77] are introduced.

Because the historical measurements of the real power systems are rarely available, the outputs of the RES are collected from the 2012 Global Energy Forecasting Competition. The training and testing datasets are generated based on this. The data generation process is based on the following two steps. The first one is to prepare the outputs of RES and the load demand and then calculate the nodal power injections.

The second one is to obtain the system states by leveraging the AC power flow

procedure in the Matpower to generate the voltage magnitudes ( $V_m$ ) and the voltage phase angles ( $V_a$ ). Thus, the measurements are obtained, including the active/reactive power injections and the active/reactive line power flows, which are available from smart meters. The number of the measurement instruments is set at 55% of the system node number. These measurement data are the input of the proposed model, and the voltage magnitudes and phase angles are the output. The measurement noise distribution is the gaussian with zero mean and 1% of the expected measurements as the standard deviation. Specifically, the noises for each measurement data in different time steps are new error values. The noises for different measurement data in a time step are also randomly added. When the data set is ready, there are near 1500K and 20000K for the 118-node distribution system and 1746-node distribution system, respectively. 90% of this data is set as training data, and 10% of them is set as testing data.

(1) USGCN's Performance index

Several evaluation indexes are leveraged to verify the proposed model.

1) Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N_{sam}} \sum_{i=1}^{n_{sam}} |\hat{h}_i - h_i| \quad (3.14)$$

2) Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{N_{sam}} \sum_{i=1}^{n_{sam}} \left| \frac{\hat{h}_i - h_i}{h_i} \right| \times 100\% \quad (3.15)$$

3) Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_{sam}} \sum_{i=1}^{n_{sam}} (\hat{h}_i - h_i)^2} \quad (3.16)$$

In (3.16),  $\hat{h}_i$  represents the estimated values outputted by the model.  $h_i$  denotes the actual corresponding state variable values.

## (2) Hyperparameters setting

The channels of the 2-D CNN are set 4, 8, and 2, and the corresponding kernel sizes are (5,1). The dimension of the node embedding is set at 10. The same learning rate setting as 0.01. The batch size is set at 1000. The regularization coefficient is set  $10^{-12}$ . These parameters are selected by optimization. The training process is conducted on a windows 10 computer with an NVIDIA GeForce GTX 3090 with 24 GB RAM. The language used is python.

## (3) Renewable Energy Sources (RES) setting

The outputs of wind turbines and photovoltaic panels according to the historical data are depicted in Figure 3.3. Note that it is the average output of a location with wind turbines or photovoltaics. The total rated power from the wind turbines at one site is set 0.06 MW. The rated power from the photovoltaic is set 0.08 MW. The locations of these RES employment sites in the test systems are shown in Table 3.1. The RES's average correlation coefficient matrix is depicted as heatmap in Figure 3.4.

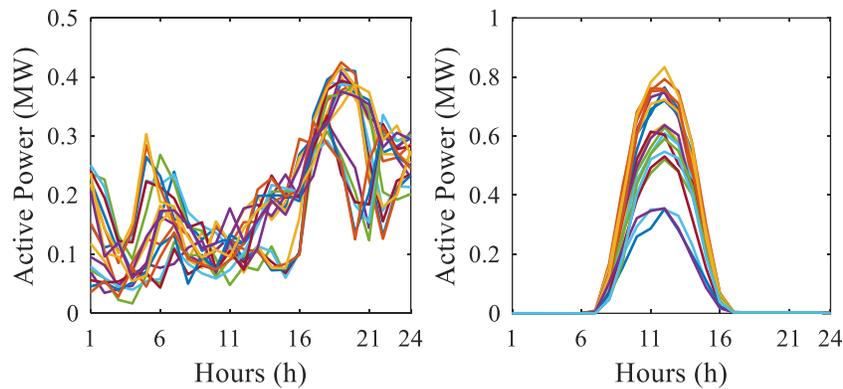


Figure 3.3. RES power output with each curve depicting one day.

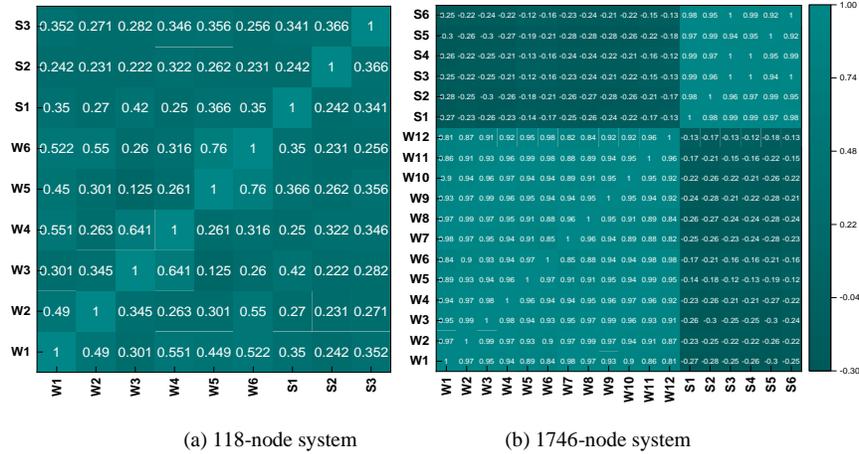


Figure 3.4. RES output's average correlation coefficient.

Table 3.1. The locations of the wind turbines and solar panels

118-node	Location	13	32	42	22	25	74	101	86	54
	Renewable	W1	W2	W3	W4	W5	W6	S1	S2	S3
1747-node	Location	34	73	172	191	1277	1387	1517	260	377
	Renewable	W1	W2	W3	W4	W5	W6	W7	W8	W9
	Location	434	587	615	803	948	1050	1064	1116	1719
	Renewable	W10	W11	W12	S1	S2	S3	S4	S5	S6

### 3.5.1 USGCN in State Estimation

(1) Comparison to Traditional Methods:

The conventional methods for DSSE, including weight (Weighted Least Squares) WLS [78], (Least Absolute Value) LAV [79] are conducted as the comparison methods to verify the effectiveness of the proposed model. These conventional methods are optimized by the Gaussian-Newton algorithm.

Since the historical data of the wind and solar power outputs shown in Figure 3.4 are available with correlations, to investigate the state estimation on traditional methods and the proposed USGCN with and without correlations. The normal distribution is

leveraged to generate the RES's outputs, where the mean is the rated power, and the standard deviation is set as 15% of the mean. In this way, the independent RES outputs can be obtained. Then, the power flow is conducted to generate the state variables, where the additional 1% noise is added. The procedures of WLS and LAV are performed for 480 scenarios selected from the testing data set. The average evaluation indexes, including the MAE, MAPE, and RMSE from the WLS and LAV methods are summarized in Table 3.2.

From Table 3.2, it can be seen that without correlations, in comparison with the proposed model, the WLS and LAV can provide a lower accuracy result in state estimation in terms of the MAE, MAPE, and RMSE. The evaluation indexes obtained by the USGCN almost are very close to that of the WLS and LAV. This indicates that the proposed model has effectiveness in the DSSE. When the correlations are considered, the evaluation indexes obtained by the USGCN are lower than that obtained by the traditional methods. Specifically, in the 118-node distribution system with correlations, voltage magnitudes' MAE obtained by USGCN are degraded by 91.77% and 91.79% compared with the WLS and LAV, respectively. The voltage magnitudes' MAPE are degraded by corresponding 91.82% and 91.82%, respectively. In addition, for the 1746-node distribution system, voltage magnitudes' MAE obtained by USGCN are also reduced by 85.64%, 85.64% and the corresponding MAPE are 85.21%, 85.21%, respectively. The voltage magnitudes' MAPE and RMSE obtained by USGCN are also lower than the traditional methods. The MAE and RMSE in voltage phase angles of the USGCN state estimation results considered either with or without correlation are near that of the traditional methods, while the MAPE is smaller. Since the MAPE index is much more sensitive to errors, the smaller MAPE demonstrates that the proposed can give acceptable state estimation results. This is due to the USGCN's outstanding ability in extracting information from complex spatiotemporal correlations. Thus, there comes the conclusion that the USGCN gives high accuracy state estimation results with hidden spatiotemporal correlations considered.

Table 3.2. The evaluation index of USGCN compared with traditional methods with and without correlations

System node	States	Methods	Without Correlation			With Correlation		
			WLS	LAV	USGCN	WLS	LAV	USGCN
118	$V_m$	MAE	7.67E-03	7.67E-03	7.13E-04	7.53E-03	7.55E-03	6.20E-04
		MAPE (%)	7.92E-01	7.92E-01	7.36E-02	7.87E-01	7.88E-01	6.44E-02
		RMSE	9.68E-03	9.68E-03	8.93E-04	9.49E-03	9.51E-03	8.39E-04
	$V_a$	MAE	2.11E-03	2.11E-03	4.07E-03	2.67E-03	2.68E-03	4.68E-03
		MAPE (%)	7.95E-01	7.95E-01	4.18E-01	7.91E-01	7.90E-01	2.36E-01
		RMSE	8.56E-03	8.56E-03	6.73E-03	8.78E-03	8.78E-03	7.24E-03
1746	$V_m$	MAE	7.87E-03	7.87E-03	1.14E-03	7.88E-03	7.88E-03	1.01E-03
		MAPE (%)	7.98E-01	7.98E-01	1.18E-01	7.98E-01	7.98E-01	1.09E-01
		RMSE	9.87E-03	9.87E-03	1.44E-03	9.87E-03	9.87E-03	1.12E-03
	$V_a$	MAE	5.23E-03	5.23E-03	4.84E-03	5.24E-03	5.24E-03	4.24E-03
		MAPE (%)	7.97E-01	7.97E-01	2.44E-01	7.97E-01	7.97E-01	2.21E-01
		RMSE	5.62E-03	5.62E-03	4.46E-03	5.64E-03	5.64E-03	4.15E-03

The voltage magnitudes and phase angles estimated at 121 in the 1746-node distribution system during different time periods are depicted in Figure 3.5. From this picture, it can be seen that for the conventional methods, WLS and LAV, the bias between the estimated values and the actual values with correlation are much larger than that without correlations. This intuitively indicates a fact that RES's correlations bring a significant effect on the distribution system state estimation. This will further lead to negative operational decisions on the management of the power system. Besides, the state estimated voltage magnitudes and phase angles are also depicted in Figure 3.5. It intuitively shows that the voltage magnitudes and phase angles from the proposed USGCN have a smaller bias with the corresponding real value in comparison with the traditional WLS and LAV methods.

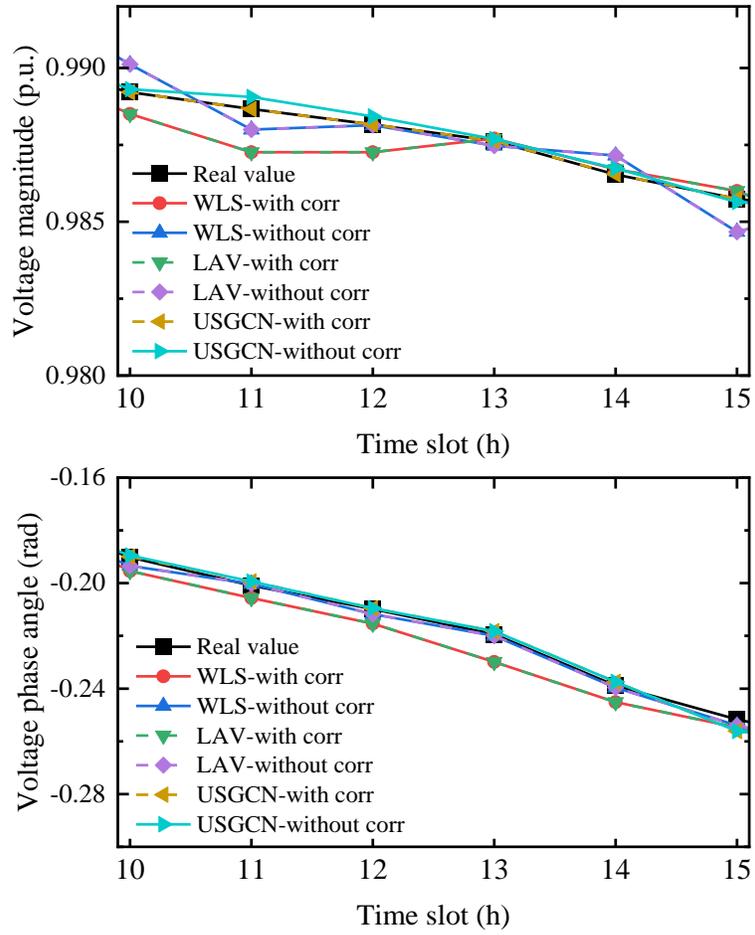


Figure 3.5. The estimated value of voltage magnitudes and phase angles of node 121 of the 1747-node distribution system.

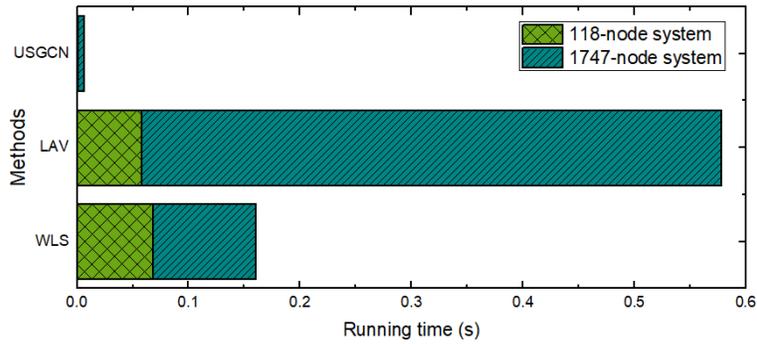


Figure 3.6. The estimation time of USGCN and traditional methods.

Moreover, to verify the efficiency of the USGCN, the computational times of USGCN, WLS, and LAV for 118/1746-node systems are depicted in Figure 3.6. Specifically, the calculating times for the 118-node system from WLS, LAV, and USGCN are 0.1758, 0.0624, and 0.0006s respectively. The corresponding times for 1746-node system are 0.0928, 0.5201, 0.0064s, respectively. It is obvious that USGCN's computational time is less than 1ms, but the traditional method needs more than 60ms. This fact demonstrates that the USGCN can save about 100 times calculating time compared with the traditional methods. This implies the potential applicational ability of the USGCN in real-time DSSE. This is because the proposed USGCN can learn the expectation of the posterior conditional distribution that describes the complex relationship between the measurements and the system states. In this way, the USGCN directly performs state estimation different from traditional methods that involve many time-consuming iteration processes so that reducing computational time can be achieved. The USGCN can be adapted to the distribution system state estimation efficiently.

## (2) Comparison to Data-driven Methods

To verify the USGCN's effectiveness in comparison with other learning models, the Full Connected Network (FCN), Convolutional neural network (CNN), Graph Convolutional Network (GCN) [71], and Spatiotemporal graph convolutional network (SGCN) are conducted as the comparison methods. There are five fully connected network layers with 1000, 2000, 1000, 500, and 118 neurons in each layer. The adjacency matrix in the GCN is formed by the topology of the distribution system. There are two convolutional graph layers in GCN. There are three 2-d convolutional layers in CNN where each kernel size of the CNN layer is 5, the channel sizes are 8, 16, and 8, respectively, and three fully connected network layers with 1000, 500, and 118 neurons. Note that the distribution system states are estimated based on the measurement at the current time step and the previous two time steps. Thus, the time step setting in USGCN for distribution system state estimation USGCN is 22.

Table 3.3. The Evaluation Index in State Estimation of USGCN in Comparison with Data-driven Methods with Correlations

System node	States	Methods	FCN	GCN	CNN	SGCN	USGCN
118	$V_m$	MAE	3.48E-03	8.25E-03	2.20E-03	2.50E-03	6.20E-04
		MAPE (%)	3.64E-01	8.67E-01	2.29E-01	2.61E-01	6.44E-02
		RMSE	5.32E-03	1.24E-02	3.30E-03	3.89E-03	8.39E-04
	$V_a$	MAE	1.87E-02	3.24E-02	1.20E-02	2.41E-02	4.68E-03
		MAPE (%)	9.37E-01	1.61E+00	6.04E-01	1.20E+00	2.36E-01
		RMSE	3.15E-02	5.38E-02	2.01E-02	4.12E-02	7.24E-03
1747	$V_m$	MAE	3.98E-03	2.95E-03	2.12E-02	1.25E-03	1.01E-03
		MAPE (%)	4.12E-01	3.04E-01	2.15E+00	1.29E-01	1.09E-01
		RMSE	7.26E-03	4.91E-03	2.66E-02	2.06E-03	1.12E-03
	$V_a$	MAE	8.55E-01	8.86E-02	6.04E-02	1.36E-02	4.24E-03
		MAPE (%)	4.62E+01	4.81E+00	3.26E+00	7.68E-01	2.21E-01
		RMSE	8.60E-01	1.11E-01	7.66E-02	2.17E-02	4.15E-03

The state estimation results are listed in Table 3.3. For the IEEE 118-node system, it shows that the proposed USGCN gains improved 82.18%, 92.48%, 71.82%, and 75.20% in voltage magnitude MAE in comparison with FCN, GCN, CNN, and SGCN, respectively. Besides, USGCN also enhances voltage magnitude RMSE at 82.23%, 32.34%, 74.58%, and 78.43% in comparison with FCN, GCN, CNN, and SGCN, respectively. Moreover, for the 1747-node system, the USGCN reduces voltage phase angle MAE at 74.62%, 96.58%, 95.24%, and 68.82% in comparison with FCN, GCN, CNN, and SGCN, respectively. This indicates that the proposed USGCN can achieve a better accurate state estimation. The state estimation of voltage magnitudes and phase angles in nodes 5-15 of the 118-node system is depicted in Figure 3.7. Curves illustrate that the proposed USGCN model outperforms other data-driven models in state estimation.

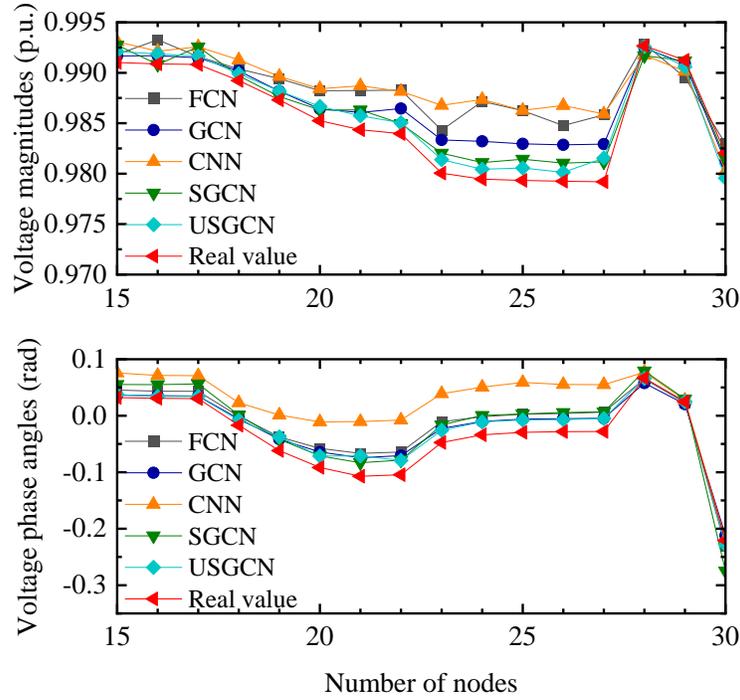


Figure 3.7. The state estimation of voltage magnitudes and phase angles of the 118-node distribution system at hour 12.

Furthermore, to inventively deliver the correlations learned by the USGCN, the parameter in the self-adaptive unrolled spatiotemporal graph is depicted. Each element in this graph is calculated by the inner product of the corresponding node embeddings. Figure 3.8(a) denotes the correlations of RES. Figure 3.8(b) denotes the adjacent matrix of topology. Figure 3.8(c) denotes the correlations learned by the self-adaptive unrolled spatiotemporal graph. In comparison with the heatmaps in (a) and (b), (c) presents more complex correlations in unrolled spatiotemporal graph models between nodes across time steps. This indicates that the correlation coefficient matrix and the adjacent matrix of topology cannot fully represent the complex correlations between nodes. However, the proposed node embedding can capture the correlations from the measurement data automatically. In this way, the state estimation accuracy can be improved.

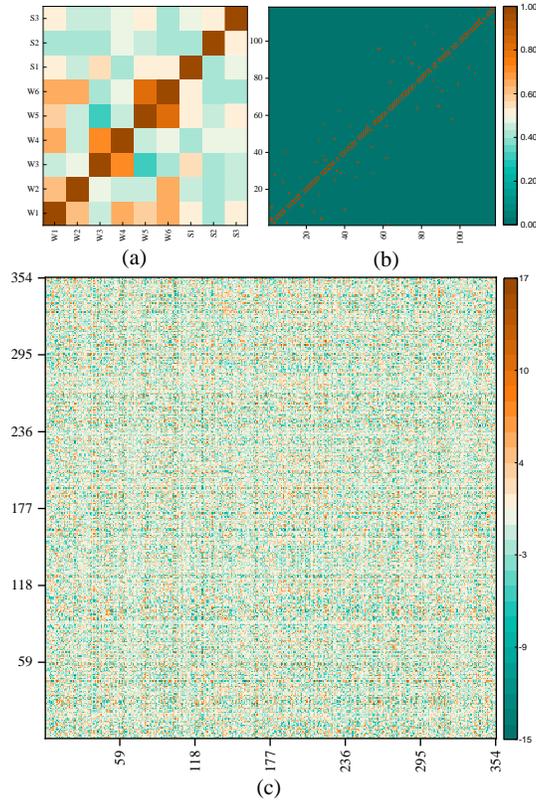


Figure 3. 8. (a) Correlations of RES. (b) Adjacent matrix of topology. (c) Parameters in the unrolled spatiotemporal graph.

Since the number of the convolutional layers affects the state estimation results, a different number of CNN layers are set to verify their effect on the performance of the USGCN. Figure 3.9 depicts the MAE and MAPE of state estimation for USGCN with different numbers of CNN layers. The biases between the training data and the testing data increase with the increase of the CNN layers. This indicates that the overfitting rises with eh increasing number of CNN. Besides, the three CNN layers achieve the lowest MAE and MAPE so three CNN layers are chosen in the proposed model.

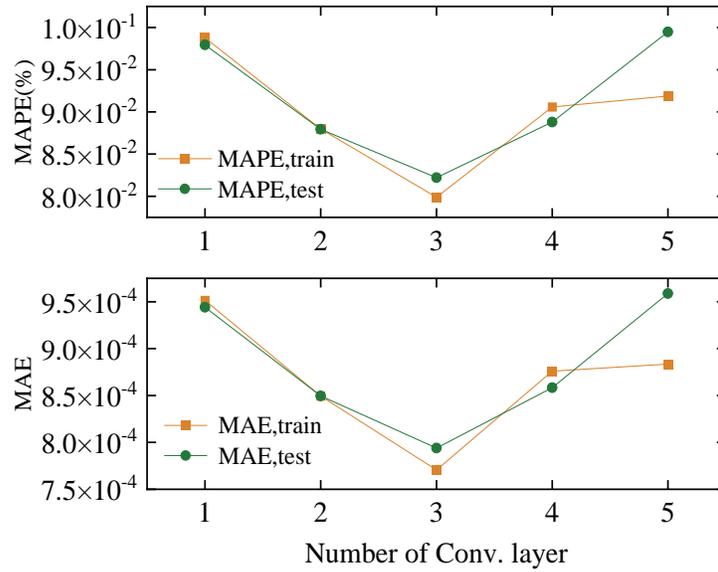


Figure 3.9. The state estimation for a different number of CNN layers in USGCN.

### 3.5.2 USGCN in State Forecasting

To verify the USGCN's effectiveness in comparison with other learning models, the Full Connected Network (FCN), Convolutional neural network (CNN), Graph Convolutional Network (GCN), Long Short-Term Memory (LSTM), GCN-LSTM [69], and SGCN are introduced as the comparison methods. The FCN, GCN, and CNN have the same setting as mentioned in the above section. There are three layers in LSTM. Then, the CNN is combined with LSTM to form the CNN-LSTM method. The voltage magnitudes and phase angles' forecasting results are summarized in Table 3.4. Note that state forecasting refers to the single step following the last observed time step. The time step used in SGCN, USGCN and LSTM is set at 21.

As shown in Table 3.4, voltage magnitudes' MAE, MAPE, RMSE obtained by USGCN are lower than the comparison methods. This demonstrates that the USGCN achieves better forecasting system states than other data-driven methods. Specifically, the voltage magnitudes' MAE obtained by the USGCN are degraded by 11.7%, 27.3%,

56.9%, 77.8%, 92.2%, and 41.28% in comparison with the FCN, GCN, CNN, LSTM, CNN-LSTM, and SGCN methods, respectively. In addition, the voltage phase angles' MAPE obtained by the USGCN are also degraded by 51.8%, 59.3%, 51.6%, 76.4%, 82.8%, and 70.59%, respectively.

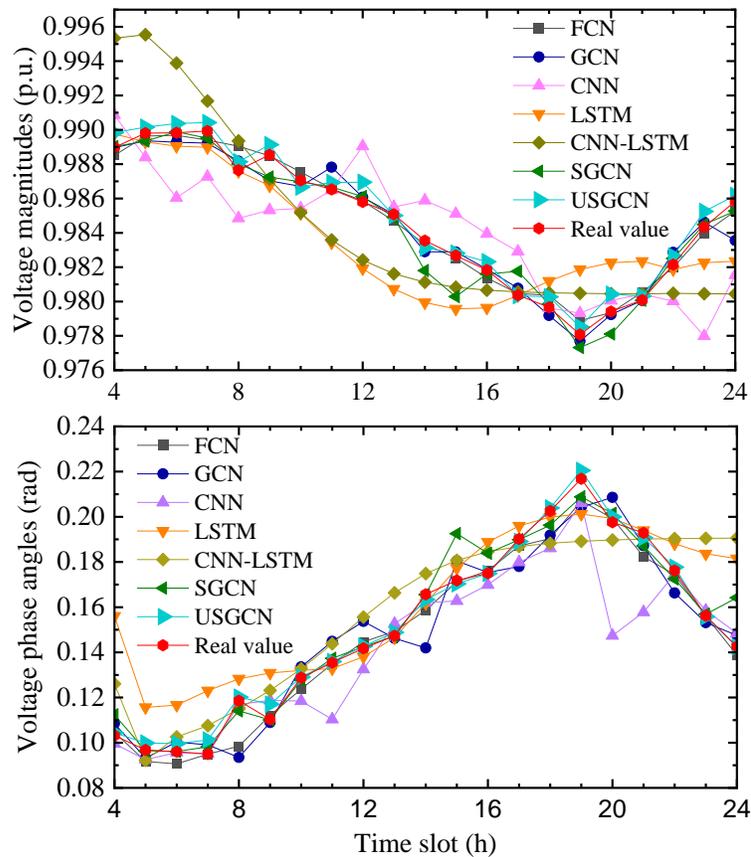


Figure 3.10. The prediction results of voltage magnitudes and phase angles of node 55 in the 118-node distribution system.

Table 3.4. The evaluation index in State forecasting of USGCN compared with deep learning methods with correlations

States	Methods	FCN	GCN	CNN	LSTM	CNN-LSTM	SGCN	USGCN
$V_m$	MAE	1.45E-03	1.76E-03	2.93E-03	5.78E-03	1.65E-02	2.18E-03	1.28E-03
	MAPE (%)	1.51E-01	1.84E-01	3.06E-01	6.00E-01	1.69E+00	2.28E-01	1.33E-01
	RMSE	2.23E-03	2.63E-03	4.03E-03	1.02E-02	4.05E-02	3.53E-03	1.68E-03
$V_a$	MAE	1.25E-02	1.50E-02	1.32E-02	2.64E-02	3.55E-02	1.89E-02	6.06E-03
	MAPE (%)	6.37E-01	7.54E-01	6.65E-01	1.30E+00	1.79E+00	9.48E-01	3.07E-01
	RMSE	1.98E-02	2.33E-02	2.59E-02	4.40E-02	5.78E-02	3.09E-02	9.62E-03

The predicted voltage magnitudes and phase angles in node 55 on the 118-node system are depicted in 3.10. From this figure, it can be seen that the state variables obtained by USGCN have a more similar trend to the actual values than the comparison methods. More specifically, LSTM and CNN-LSTM show a relatively low accuracy in state forecasting due to their limited learning ability in the power systems' data. Besides, the SGCN captures the spatial and temporal correlations separately, leading to low accurate state forecasting results. This indicates the distinguished prediction ability of the proposed model. This is because the spatiotemporal correlations are extracted by the unrolled spatiotemporal graph model so that enhanced accuracy can be achieved. Furthermore, the predicted voltage magnitudes and phase angles in different time periods are depicted in Figure 11. This figure intuitively shows that the USGCN provides more accurate predicted values than the comparison methods. This is because both the temporal and spatial correlations are fully captured and thus the ahead-of-time DSSE can be achieved effectively.

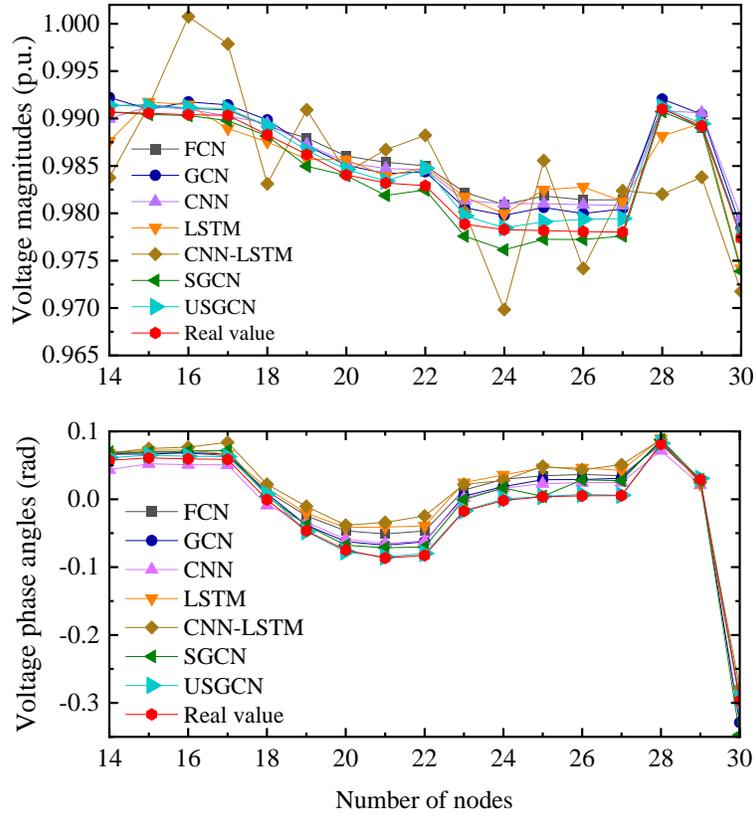


Figure 3.11. The prediction results of voltage magnitudes and phase angles of node 14-30 of the 118-node distribution system at hour 12.

### 3.5.3 USGCN in Large-scale Distribution System

To verify the USGCN's performance in the large-scale distribution system, a 1746-node distribution system is conducted as the test system, where 18 locations are deployed with RES units. Due to the data scale being too large to be learned by the FCN, GCN, and LSTM methods. the CNN is maintained to be the baseline method.

The predicted voltage magnitudes and phase angles obtained by USGCN and the comparison methods are depicted in Table 3.5. It can be seen that the voltage magnitudes' MAE, MAPE, and RMSE obtained by USGCN are degraded by 86.3%,

86.1%, and 63.3%, respectively. The phase angles' MAE, MAPE, and RMSE obtained by USGCN are degraded by 59.3%, 80.4%, and 7.3%, respectively. Besides, the MAE, MAPE, and RMSE of voltage phase angles forecasted by the USGCN also outperform SGCN by 76.92%, 53.87%, and 51.64%, respectively. Thus, the USGCN achieves high accuracy in state forecasting for large-scale distribution systems in terms of accuracy and efficiency.

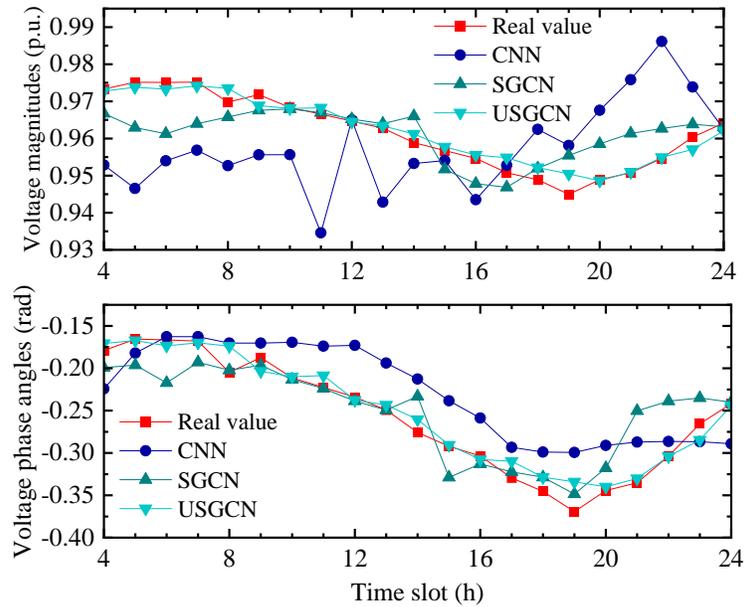


Figure 3.12. The prediction results of voltage magnitudes and phase angles for bus 500 of the 1746-node distribution system.

The forecasting voltage magnitudes and phase angles at node 500 in the 1746-node distribution system are shown in Figure 3.12. It is intuitively that biases between the predicted values and the actual values obtained by the USGCN are much smaller than that obtained by the comparison method. It intuitively shows that USGCN can archives the state forecasting in a large-scale distribution system with a high accuracy so that the ahead of time system operational decisions can be made by the operators.

Table 3.5. The evaluation index of USGCN of 1746-node distribution system

States	Methods	CNN	SGCN	USGCN
$V_m$	MAE	1.23E-02	7.10E-03	1.69E-03
	MAPE (%)	1.25E+00	7.19E-01	1.74E-01
	RMSE	1.55E-02	9.16E-03	5.68E-03
$V_a$	MAE	3.54E-02	1.56E-02	1.44E-02
	MAPE (%)	1.94E+00	8.26E-01	3.81E-01
	RMSE	3.98E-02	7.63E-02	3.69E-02

### 3.6 Summary

The unrolled spatiotemporal graph convolutional network model is innovatively proposed in this work to timely estimate and forecast distribution system states with high-penetrated renewable power energy that exhibits complex correlations temporarily and spatially. More than 16.42% improvement in the evaluation indexes of estimated states is superiorly achieved over the traditional methods due to the complex correlations being captured by the proposed unrolled spatiotemporal graph model effectively. Besides, the proposed USGCN also speeds up the computational time by more than 100 times. Moreover, the ahead-of-time forecasting states are achieved by the proposed USGCN due to multiple stacked of the unrolled spatiotemporal graph model that allows a large field of correlations to be perceived. The proposed USGCN represents the ability in DSSE and state forecasting in terms of accuracy and efficiency.

# *Chapter 4 Probabilistic Power Flow of Distribution System Based on a Graph-Aware Deep Learning Network*

## **4.1 Introduction**

The proliferation of renewable energy brings significant uncertainty to active distribution systems due to its intermittent and stochastic nature. Such a new scene in the distribution system unavoidably introduces challenges to the reliable operation of power systems. To represent the impact of uncertainty on distribution system planning and operation, the PPF computation gains many academic eyes due to its distinguishing ability on quantifying the uncertain impact on system states induced by the fluctuated power injections [80]. Thus, it is of great importance to develop a more effective model for PPF.

The probabilistic power flow provides the distribution features of the system states for the power system planning to hedge against uncertainty. It is unprecedentedly important to take the complex uncertainty of RES into account in the PPF. It is well known that the distribution system is integrated with various renewable energy sources while it covers a relatively small area geographically. This phenomenon results in renewable energy deployment sites being close to each other and thus they are generally sharing similar weather and environment. Similar profiles generated by these

renewable energy units lead to complex correlations among nodal power injections. It poses an inevitable bias in PPF calculations [81]. Thus, there remains a problem to handle PPF while taking the complex correlation characteristic of RES into account.

Many works reported in literatures have been focused on the PPF calculation. The Monte Carlo (MC) simulation method repeatedly computes the deterministic power flow for as many as possible scenarios of the wind and solar power injection to obtain the statistical characteristic of the system state variables. In this way, the PPF can achieve accurate results and is often utilized as the baseline [82], [83]. However, the MC-based PPF suffers from a heavy computational burden. To alleviate this issue, many approaches are utilized to attempt to reduce the number of power injection scenarios via generating representative samples, including the stratified Latin Hypercube Sampling [84], the Latin Supercube sampling [85], and so on. However, a great many samples may still be required in these numerical methods to represent the uncertain patterns of state variables accurately.

To further alleviate the calculation burden in PPF, some analytical methods are introduced in the previous work. The cumulant-based method [86], [87] is introduced to compute the PPF in a linear manner on the power flow, which is not very adaptive to distribution system PPF with high nonlinearity. However, it is not suitable for a distribution system, whose power flow equation is difficult to linearize due to the high impedance ratio. The generalized polynomial chaos is proposed in [88] for reducing the computational complexity of PPF. Point estimation methods (PEM) [89], [90] calculate the nonlinear power flow at the moment of the system condition so that the approximated density function of the system states can be obtained. Thus, by solving the deterministic power flow with limited classical scenarios, the PPF computational efficiency can be ameliorated. However, the non-gaussian distribution on nodal power injections brings bias to classical PEM. In [91], the Cornish-Fisher expansion approach is introduced for fitting complex data features extracted from the renewable power injections, including wind and solar power. Limited deterministic power flow calculations are required in these methods, resulting in computationally tractable.

However, these approaches are only fit to PPF with power injections with dependence.

To handle the correlations on the PPF problem, the common practice is to utilize the transformation methods to decouple the correlations and then apply the PPF calculation reported above. An approach is introduced in [89] to transform the input random variables set into a non-physical uncorrelated reference frame, then the PEM is employed to PPF. Similarly, in [16] the Nataf transformation is leveraged to transfer the estimation point with correlations into independence before PEM application. In [92] and [93], the Nataf transformation is combined with the Latin Hypercube Sampling to handle the correlations before the PPF calculation and thus enhance the accuracy and efficiency. A Hierarchical Adaptive Polynomial Chaos-ANOVA Method is proposed in [17] to extend the application of the polynomial chaos method [88] on PPF with correlations. The temperature-related errors are taken into account in the temperature-augmented model proposed in [94]. However, these reported methods assumed that the prior distribution of the correlations is known. Then this is pre-processed by the mathematical transform methods and combined with either the numerical methods or the analytical methods to calculate the PPF. This relies on the full knowledge of the correlations and it is not practical in realistic PPF implementation.

Smart meters are greatly deployed in the distribution system, recent works try to develop data-driven methods to deal with the PPF problem. Inspired by this, a Model-Based Deep Learning Approach is proposed in [95] to remove the computational burden, which ignores the correlations brought by the renewable energy units. A convolutional neural network is proposed in [96] to speed up PPF calculation. However, this training process involves plenty of learnable parameters. Thus, a more efficient learning-based PPF model is essentially required to be studied.

In this Chapter, a Graph-Aware Deep Learning Network (GADLN) model is leveraged to handle the probabilistic power flow (PPF) that is exposed to complex dependence among renewable power outputs. Specifically, to fully capture the mapping from the fluctuated power injections and the uncertain system states, the convolutional operation is introduced to aggregate the correlations among renewables

outputs according to their geographical locations to facilitate the PPF. In this way, the deviation pattern of the system state variables can be well learned. Therefore, by extracting the implicit distribution features in the measurements, the GADLN allows the graph-aware learning model to map the power injections into the system states autonomously while considering the complex correlations among renewable power outputs so that the improved effectiveness and speed-up calculation can be achieved. Moreover, the numerical results show the superior of the GADLN over the state-of-the-arts with accurate and effective manners in the IEEE 33-node distribution system.

The remaining sections are organized as follows. The definition of the probabilistic power flow with the corresponding mathematical format is proposed in Section 4.2. The novel graph-aware learning model is developed in Section 4.3. The experimental results and the corresponding discussions are introduced in Section 4.4. Finally, the works of this section are concluded in Section 4.5.

## **4.2 Probabilistic Power Flow Problem**

In this section, the PPF formulation with uncertainty as well as the correlations on the renewable energy outputs is represented, which will later be considered by the proposed model.

### **4.2.1 Deterministic Power Flow**

Among the functions to maintain the reliable monitoring and management of power system operation in the modern distribution system, the probabilistic power flow is the basic tool to perceive the system states. It delivers the uncertain impact of renewable energy outputs on the distribution pattern of the system states directly.

A set of equations that are restricted to Kirchhoff's law is employed to describe the operation of the power system. Such equations represent the mapping from the nodal

power injections to the voltage.

$$\begin{aligned} P_i &= v_i \sum_{j=1}^n v_j s_{ij} (g_{ij} \cos(\theta_i - \theta_j) + b_{ij} \sin(\theta_i - \theta_j)) \\ Q_i &= v_i \sum_{j=1}^n v_j s_{ij} (g_{ij} \sin(\theta_i - \theta_j) - b_{ij} \cos(\theta_i - \theta_j)) \end{aligned} \quad (4.1)$$

In (4.1),  $P_i$  is the net active power injection at node  $i$ .  $Q_i$  denotes the net reactive power injection at node  $i$ .  $v_i$  is the voltage magnitude at node  $i$ .  $\theta_i$  denotes the voltage phase angle at node  $i$ .  $g_{ij}$  is the conductance at branch  $s_{ij}$ .  $b_{ij}$  is the susceptance at branch  $s_{ij}$ .

$$\begin{aligned} P_i &= \sum_{g=1}^{G_i} P_i^G - \sum_{d=1}^{D_i} P_i^D \\ Q_i &= \sum_{g=1}^{G_i} Q_i^G - \sum_{d=1}^{D_i} Q_i^D \end{aligned} \quad (4.2)$$

In (4.2),  $P_i^G$  is the active RES output at node  $i$ .  $P_i^D$  denotes the active load demand at node  $i$ .  $Q_i^G$  is the reactive RES output at node  $i$ .  $Q_i^D$  denotes reactive load demand at node  $i$ . Note that for the node without RES or load demand, the corresponding  $P_i^G$ ,  $P_i^D$ ,  $Q_i^G$  and  $Q_i^D$  are set zeros.

#### 4.2.2 Probabilistic Power Flow

The outputs of the RES are uncertain due to their stochastic and intermittent nature. Such uncertain impact on the system states can be calculated and represented by PPF, which is formulated by a function of system conditions. The relationship between the system states and the system conditions can be expressed by:

$$Y = f(X) \quad (4.3)$$

The input of the PPF can be formulated as a vector.

$$X = [P, Q] \quad (4.4)$$

The input of the PPF consists of the network conditions, the load demand, and the

renewable power outputs. The outputs of the PPF can be expressed as:

$$Y=[V,\theta] \quad (4.5)$$

In (4.5), it is obvious that the system states as the outputs are determined by the inputs.

### 4.2.3 The Uncertainty of the RES Power Output

The distribution of the wind speed is generally described by the Weibull probabilistic distribution function.

$$f(v) = \frac{k}{\lambda} \left(\frac{v}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{v}{\lambda}\right)^k\right] \quad (4.6)$$

In (4.6),  $v$  is the actual wind speed.  $k$  is the shape parameter in this function.  $\lambda$  is the scale parameter in this function. After the wind speed distribution is modeled, the power outputs of wind turbines can be obtained by the wind speed-power curve.

$$P_w = \begin{cases} k_1 v + k_2, & v_i \leq v \leq v_r \\ P_r, & v_r \leq v \leq v_o \\ 0, & \text{otherwise} \end{cases} \quad (4.7)$$

In (4.7), the  $k_1$  and  $k_2$  can be obtained by the following equations,  $k_1 = P_r / (v_r - v_i)$ ,  $k_2 = -k_1 v_i$ .  $v_r$  denotes the rated wind speed of wind turbines.  $v_i$  is the cut-in wind speed.  $v_o$  represents the cut-off wind speed.  $P_r$  is the rated wind power.  $P_w$  denotes actual active wind power output.

The Beta probabilistic distribution function is generally leveraged to characterize solar irradiance.

$$f(E) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} E^{\alpha-1} (1-E)^{\beta-1} \quad (4.8)$$

where  $E$  is the solar irradiance ( $\text{W}/\text{m}^2$ ).

Similarly, after the solar irradiance is obtained, the power outputs of the panels can be commutated by the solar radiation-power curve.

$$P^{PV} = E\eta^{PV}S^{PV} \quad (4.9)$$

In (4.9),  $S^{PV}$  is the Photovoltaic panel area (m<sup>2</sup>).  $\eta^{PV}$  denotes the solar power conversion efficiency.  $P^{PV}$  is the actual PV power output.

#### 4.2.4 The Correlation Among RES

Conventionally, the correlation coefficient matrix  $C$  is employed to model the correlation among RES.

$$C = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1g} \\ \rho_{21} & 1 & \cdots & \rho_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{g1} & \rho_{g2} & \cdots & 1 \end{bmatrix} \quad (4.10)$$

where  $g$  is the correlation nodal injections' number. The element of  $C$  is obtained by the following equation.

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma_i \sigma_j} \quad (4.11)$$

where  $X_i$  is the variable presenting the nodal injection in node  $i$ ;  $\text{cov}(X_i, X_j)$  denotes the covariance of variables  $X_i$  and  $X_j$ ; and  $\sigma_i$  is the standard deviation of variables  $X_i$ . Since the matrix  $C$  is a symmetric positive definite matrix, it can be decomposed into

$$C = LL^T \quad (4.12)$$

where  $L$  is an inferior triangular matrix. The inverse  $L^{-1}$  is generally leveraged to decouple the dependent variables so that the independent variables can be obtained by  $X_{in} = L^{-1}X$ . However, due to the violent fluctuation in renewable energy outputs, the corresponding correlations among them usually represent nonlinearity and complexity. The linear correlation coefficient matrix can no longer be utilized to describe such features. Thus, it is the motivation of the proposed GADLN to explore this issue.

### 4.3 Graph-aware deep learning Network

Firstly, the graph-aware scheme is introduced to fully capture the features of the nodal power injections and then the entire structure of the proposed GADLN model is formulated in detail.

#### 4.3.1 Graph-Aware Scheme

The convolutional operation is one of the technologies of the neural network. To adapt the nodal power injections vector of the power system, the one-dimensional convolutional operation is proposed to extract the features of the nodal power injections instead of a three-dimensional convolutional operation which requires too many parameters. The one-dimensional convolutional operation is parameterized by the following equation with the input vector  $X \in R^{B \times N \times 1}$ .

$$Y(B, c, N - ks + 1) = b + \sum_{c=0}^{C-1} W(C, c) \otimes X(B, c, N) \quad (4.13)$$

In (4.13),  $\otimes$  is the dot-product.  $c$  represents the number of channels;  $b$  denotes the bias;  $W$  is the kernel parameters;  $ks$  is the kernel's size;  $B$  denotes the batch size;  $N$  indicates the length of the power injection sequence. During the training process, the kernel can aggregate the data features within the window size.

To make sure the nodal power injections vector is compatible with the one-dimensional convolutional operation while with graph nature preserved, the graph-aware scheme-based inputs are defined as  $X = [P_i, Q_i]$ .

$$\begin{aligned} P &= [P_{fh1}, P_{fh2}, \dots, P_{fhk}, \dots, P_n] \\ Q &= [Q_{fh1}, Q_{fh2}, \dots, Q_{fhk}, \dots, Q_n] \end{aligned} \quad (4.14)$$

where  $fh$  denotes feeder  $h$ ;  $k$  denotes the number of lines in feeder  $h$ . This definition of the data inputs can facilitate the data being captured by the model graphically. This is because the nodes in a feeder are more related to each other

geographically. Their data can be aggregated by the convolutional operation more directly. Furthermore, the nearby wind and solar generation outputs can be aggregated after several convolutional operations. In this way, the complex correlations are extracted by several convolutional operations.

### 4.3.2 Structure of GADNL Model

The GADNL’s architecture is characterized in Figure 4.1. It shows that the data pass through six hidden layers from the input layer to the output layer. These six hidden layers are conducted by three convolutional neural network layers and three fully connected neural network layers. The kernel size and the neurons’ sizes are also marked in the picture. Only three convolutional neural network layers are conducted is to alleviate the overfitting problem as well as avoid numerous parameters training.

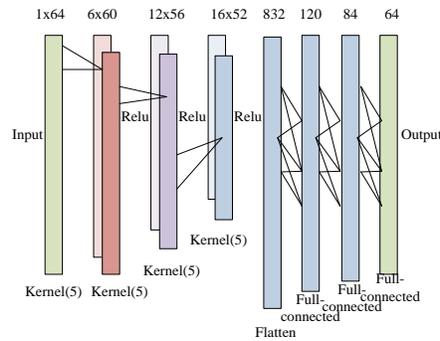


Figure 4.1. The structure of the GADNL.

## 4.4 Case study

To verify the performance of the GADNL, the IEEE 33-node system [2] is introduced to be the test system where renewable generation is integrated, including PV and wind turbines. Before the experimental simulation, the training data set is required to be generated. The generated training data set consists of the net active and

reactive power injections as the inputs of the model, the voltage magnitudes and voltage phase angles, and the active line flow and the reactive line flow as the outputs of the model. This data set can be achieved by the following steps. Firstly, the outputs of PV, wind turbines are generated according to the Weibull and Beta functions based on the Monte Carlo (MC) methods. The load demand is generated according to the normal distribution. Secondly, AC power flow is leveraged to obtain the system states taking the generated injections above as the inputs. This data generation process is conducted in Matlab. After the training data set is ready, 100K samples are obtained, where 90% of samples as the training data and 10% of samples as the testing data.

(1) Error metrics for voltage magnitudes and phase angles

In order to provide the indexes in evaluating the performance of the GADNL, the mean error of voltage magnitudes and phase angles is quantitatively formulated.

$$\varepsilon_{\mu}^m = \left| \frac{\mu_x - \mu_{MC}}{\mu_{MC}} \right| \times 100\% \quad (4.15)$$

In (4.15),  $x$  represents the PPF approaches apart from the MC methods.  $MC$  represents the MC methods in PPF.  $\mu$  denotes the average value for each nodal system state, voltage magnitudes and phase angles, active line flow, and the reactive line flow.  $m$  represents the mean or maximum errors.

(2) GADNL's Performance index

In order to verify the learning effectiveness of the GADNL, the Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) indexes are leveraged.

$$\text{MSE} = \frac{1}{N_{sam}} \sum_{i=1}^{n_{sam}} |\hat{y}_i - y_i| \quad (4.16)$$

$$\text{MAPE} = \frac{1}{N_{sam}} \sum_{i=1}^{n_{sam}} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (4.17)$$

In (4.16) and (4.17),  $\hat{y}_i$  is the direct outputs of the GADNL.  $y_i$  denotes the actual values corresponding to the outputs.

(3) Hyperparameters setting

During the training process, the learning rates of GADNL and the employed baselines methods are set at 0.01. The size of the batch is set at 1000. The regularization coefficient is set at  $10^{-12}$ . These parameters are selected by optimization. The test system conducted by the GADNL is applied by python and performance in a computer with the windows 10 environment and Intel(R) Xeon(R) CPU E5-2650 v4 @ 2020GHz, and the graphics card of NVIDIA GeForce GTX 3090 24G.

(4) The parameters of the wind and solar power models

Table 4.1. Wind Speed and Solar Irradiance Distribution Functions' Parameters

RES	PDF types	Parameters	
		<i>a</i>	<i>b</i>
W1	Weibull	6.00	2.00
W2	Weibull	6.21	1.98
W3	Weibull	6.01	2.17
W4	Weibull	5.89	2.05
W5	Weibull	6.08	2.14
S1	Beta	2.06	2.5
S2	Beta	2.12	2.8
S3	Beta	2.17	3.0
S4	Beta	2.08	2.6

Table 4.2. Correlation Coefficient Matrix of the RES

	W1	W2	W3	W4	W5	S1	S2	S3	S4
W1	1	0.49	0.301	0.551	0.450	0.522	0.350	0.242	0.352
W2	0.490	1	0.345	0.263	0.301	0.550	0.270	0.231	0.271
W3	0.301	0.345	1	0.641	0.125	0.260	0.420	0.222	0.282
W4	0.551	0.263	0.641	1	0.261	0.316	0.250	0.322	0.346
W5	0.449	0.301	0.125	0.261	1	0.760	0.366	0.262	0.356
S1	0.522	0.550	0.260	0.316	0.76	1	0.350	0.231	0.256
S2	0.350	0.270	0.420	0.250	0.366	0.350	1	0.242	0.341
S3	0.242	0.231	0.222	0.322	0.262	0.231	0.242	1	0.366
S4	0.352	0.271	0.282	0.346	0.356	0.256	0.341	0.366	1

The wind speed and solar irradiance distribution functions' parameters are listed in Table 4.1 [97]. The correlation coefficient matrix of the RES is shown in Table 4.2 [98]. The parameters wind-power curve are  $v_i = 4$  m/s,  $v_r = 14$  m/s and  $v_o = 25$  m/s, respectively.

#### 4.4.1 Considering Independent RES Power Outputs

In order to verify the GADLN's performance in PPF calculation with independent renewable energy. Two traditional approaches, MC and PEM (Three-point) are introduced as the comparison methods. The distribution of the voltage magnitudes and active line power flow is characterized in Figure 4.2, where the nodes 7 and 17, lines (9,10) and (17,18) are selected. Taking the results of MC methods as the baseline, the results of PEM and GADLN almost coincide with that of MC. This indicates that the proposed GADLN can calculate PPF with high accuracy under the independent renewable energy injections.

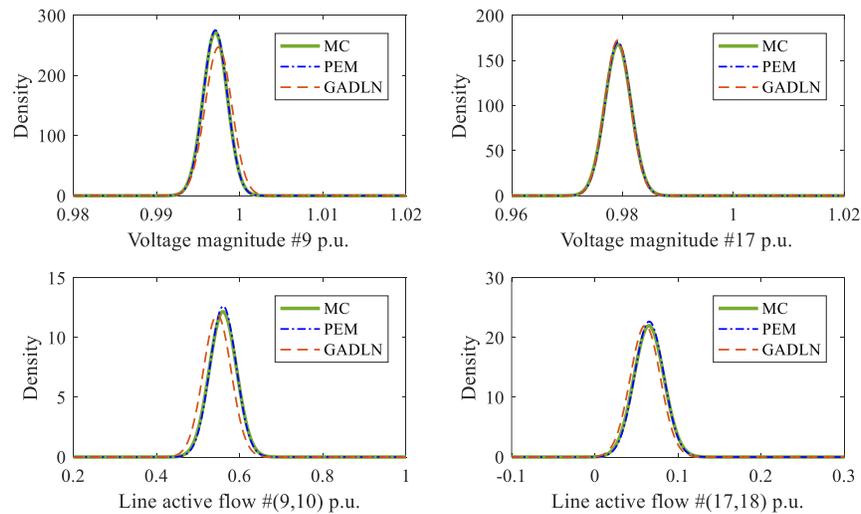


Figure 4.2. The voltage magnitudes and line active power flow.

Table 4.3. The mean error of comparison of different methods without correlation

Method	Error (%)	$V$	$\theta$	$P$	$Q$
PEM	$\varepsilon_{\mu}^{mean}$	0.0012	0.0210	0.1274	0.0769
	$\varepsilon_{\mu}^{max}$	0.0028	0.0909	0.8111	0.4983
GADLN	$\varepsilon_{\mu}^{mean}$	0.0085	0.0406	0.0502	0.0792
	$\varepsilon_{\mu}^{max}$	0.0326	0.0854	0.4061	0.4635

To represent the total simulation results of the GADLN, in comparison with the PEM, the relative error with MC on voltage magnitudes, phase angles, active line flow, and reactive line flow are summarized in Table 4.3. The relative errors are all less than 0.5%. These results further demonstrate the effectiveness of the proposed GADLN in PPF under the independent renewable energy injections.

#### 4.4.2 Considering Correlation on RES Units

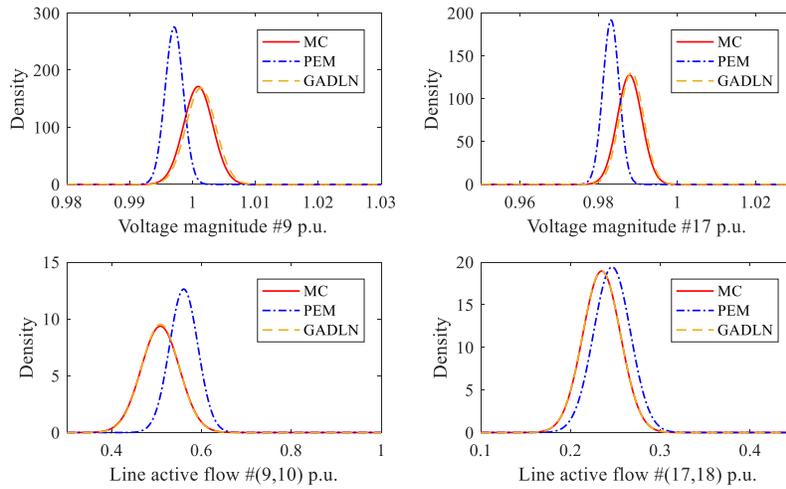


Figure 4.3. The voltage magnitudes and line active power flow.

In order to verify the GADLN's performance in PPF with renewable energy correlations, the correlation coefficient parameters are set in Table 4.3. The distribution of the voltage magnitudes and active line power flow is characterized in Figure 4.3, where the nodes 7 and 17, lines (9,10) and (17,18) are selected. It indicates that the comparison method PEM is unable to calculate the PPF with correlations considered. The GADLN achieves accurate enough results with MC.

Table 4.4. The mean error of comparison of different methods with correlation

Method	Error (%)	$V$	$\theta$	$P$	$Q$
PEM	$\mathcal{E}_{\mu}^{mean}$	0.2927	2.0810	15.6586	6.1703
	$\mathcal{E}_{\mu}^{max}$	0.5116	4.9379	237.9131	44.2972
GADLN	$\mathcal{E}_{\mu}^{mean}$	0.0071	0.0918	0.0423	0.0758
	$\mathcal{E}_{\mu}^{max}$	0.0206	0.1716	0.3452	0.5382

The relative error with MC on voltage magnitudes, phase angles, active line flow, and reactive line flow is summarized in Table 4.4. The errors obtained by GADLN are all less than 0.1%. It indicates that the proposed GADLN can achieve high effectiveness in PPF calculation with correlations considered. Furthermore, if the realistic data is utilized in the proposed model, the efficiency can also be guaranteed this is because of the strong graphical learning ability of the proposed model.

#### 4.4.3 The performance of the proposed model

To verify the GADLN's convergence, a fully connected neural network (NN) with five layers and the graph convolutional neural network (GCN) with three layers [71] are introduced as the comparison methods. The performance evaluation index MAE and MAPE are selected as mentioned above. The convergence curves of MAE and MAPE on GCN, NN, and GADLN are shown in Figure 4.4. It shows that the convergence values obtained by GADLN are higher than that of NN and GCN. This

demonstrates that the proposed GADLN can achieve higher accuracy. Moreover, the MAE and MAPE curves of GCN and NN are below that of GADLN, which means that the convergence speed of GCN and NN are much lower than GADLN. This is because the GADLN can fully capture the correlation between power injections and the system states efficiently. Therefore, the GADLN can give a high accuracy in PPF calculation as well as with a fast training speed.

The computational time for MC, PEM, and GADLN are listed in Table 4.5. Note that the number of samples utilized in MC and GADLN is set at 5000. In comparison with MC, the GADLN provides a smaller calculation time, which indicates the efficiency of the GADLN.

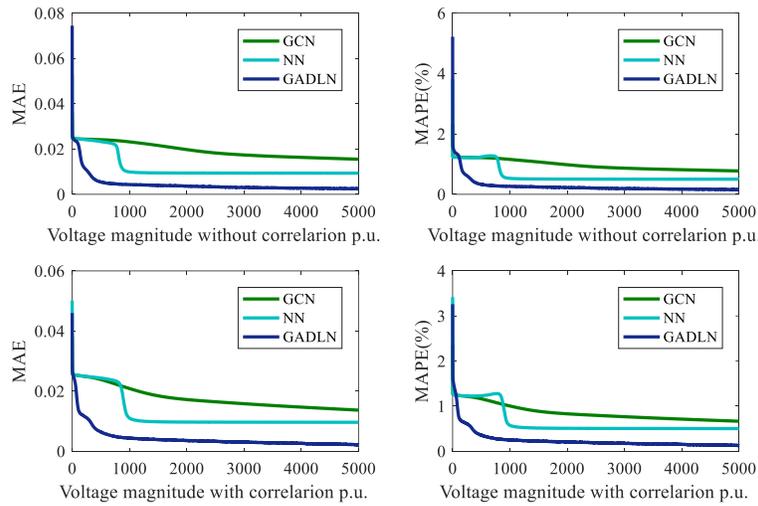


Figure 4.4. The training performance of the GADLN.

Table 4.5. The calculation time of comparison of different methods with correlation

Methods	MC	PEM	GADLN
Time (s)	11.5831	1.0487	0.0220

## **4.5 Summary**

A GADLN model is employed to calculate the PPF considering the complex correlations among nodal power injections. The high effectiveness and efficiency of the GADLN model are demonstrated by the accurate simulation results. This is due to the hidden distribution pattern on correlations among wind and solar power generation outputs being well captured by the proposed model. Furthermore, the proposed model also represents the capability of saving computational time.

# *Chapter 5 Graph Attention Enabled Convolutional Network for Distribution System Probabilistic Power Flow*

## **5.1 Introduction**

In previous Chapter 4, by using the merits of CNN's powerful feature extracting ability, the GADNL is proposed to [99] to characterize the correlations between solar and wind power units and results in more accurate results in the mean value of distribution system state variables in PPF calculation. However, GADNL neglects the high-dimensional statistical features of the system states. Thus, more efficient, and accurate data-driven PPF calculation methods are still needed to be studied.

In this Chapter, a graph attention enabled convolutional network (GAECN) is proposed to calculate PPF considering the hidden correlation of the wind and solar power resource injections [100]. This model utilizes the graph attention enabled convolutional layer to aggregate the neighboring information to improve the accuracy. Within this layer, different from only utilizing the Euclidean structure adapted convolutional neural network, the graph convolutional network is employed to extract the non-Euclidean structure of the power system. Besides, instead of using the linearized correlations of nodal power injections from historical data, the self-adaptive graph convolutional operation based on the graph embedding technique can capture the

complex correlations automatically. Then, the convolutional neural network operation is followed to handle the violent fluctuation of the outputs of renewable energy. The performance of the proposed GAECN is investigated through extensive case studies on the IEEE 33-node, PG&E 69-node, and 118-node distribution systems. The contributions are summarized as follows.

(1) The graph convolutional operation is employed to fully extract and leverage the correlations of the nodal power injections to improve the PPF calculation accuracy.

(2) The node embedding technique is integrated into the graph convolutional operation to automatically capture the nonlinear nature of the correlations so that the lower errors in the standard deviation of the system states can be achieved.

(3) The convolutional neural network is employed to adapt to the stochastic fluctuations to achieve higher accuracy with high penetration of RES.

The remaining sections are organized as follows. The definition of the probabilistic power flow with the corresponding mathematical format is proposed in Section 5.2. The GAECN model and its framework are proposed in Section 5.3. The experimental results and the corresponding discussions are introduced in Section 5.4. Finally, the works of this section are concluded in Section 5.5.

## **5.2 Graph Attention Enabled Convolutional Network**

The conventional PPF calculation model introduced in Section II depends on the specific prior knowledge, including the information on the correlation coefficient matrix. At the same time, PPF is a time-sensitive task that requires quick computation results. To overcome these difficulties, the GAECN model is introduced to solve the probabilistic power flow problem efficiently. Firstly, the graph convolutional operation combining the node embedding technique is proposed to handle the implicit correlations automatically. Then, the framework of the GAECN model is introduced following.

### 5.2.1 Graph Convolutional Operation

Generally, the physical topology of the power system is designed in a non-Euclidean structure. The correlations among different nodal power injections are also represented as a non-Euclidean structure. This data structure is described as a graph structure rather than a grid or line. It is this natural essence that inspires the choice of graph convolutional network (GCN). By representing the data in the form of a graph, GCN [71] is popularly implemented on plenty of tasks to deal with the non-Euclidean structure data, such as node classification, and traffic forecasting [101].

The measurement sites of the distribution system can be seen as nodes, and the correlations among these nodes can be depicted as edges. A graph  $\mathcal{G}=(\mathcal{V},\mathcal{E})$  is utilized to represent the graphical structure among nodes, where  $\mathcal{V}=\{v_1,\dots,v_i,\dots,v_N\}$  is the set of all nodes and  $\mathcal{E}$  is the set of edges. To describe this graph structure quantitatively, the adjacent matrix  $A \in \mathbb{R}^{N \times N}$  is introduced such that:

$$A_{ij} = \begin{cases} 1, & \text{if } v_i, v_j \in \mathcal{V}, \text{ and } (v_i, v_j) \in \mathcal{E} \\ 0, & \text{else} \end{cases} \quad (5.1)$$

Besides, the Laplacian matrix  $L$  is defined as  $L=D-A$ , where  $D_{ii} = \sum_j A_{ij}$  is the degree matrix. The normalized Laplacian matrix  $L$  can be calculated as:

$$L = D^{-1/2}(D-A)D^{-1/2} = I_N - D^{-1/2}AD^{-1/2} \quad (5.2)$$

The nodal power injection data can be handled as the features of each node, denoted by  $X \in \mathbb{R}^{N \times d}$ , where  $d$  is the dimension of features. The neighbor features can be aggregated by the aggregation scheme in the graph convolutional operation, which is the core part of GCN. This aggregation scheme is a function over the nodal features, such as an add or means, which can be divided into two categories: the spectral domain and the spatial domain. Spectral-based methods update the node's representation by multiplying its Fourier transform with its neighbor's Fourier transform. Spatial-based methods update the node's representation by convolutional itself directly with its neighbor's representation, which obtains more attention due to their high effectiveness.

The latter one is used to construct the graph convolutional operation of the proposed model. The spectral graph convolutional operation is defined as the input  $X \in \mathbb{R}^{N \times d}$  and the filter  $g_\theta(L)$  parameterized by  $\theta$ :

$$g_\theta(L) * X = U g_\theta(U^T X) \quad (5.3)$$

where  $U$  is the eigenvector of normalized Laplacian matrix  $L$  and  $U^T X$  is the Fourier transform of  $X$ . Due to the heavy computational burden of this Fourier transform calculator, a simplified graph convolutional operation is proposed [71], shown as:

$$H^{(l)} = \sigma(\widehat{A} H^{(l-1)} W^l) \quad (5.4)$$

where  $H^{(l)}$  is the output of layer  $l$  and  $H^{(0)}$  denotes the input;  $W^{(l)}$  is the trainable parameters of the layer  $l$ ;  $\sigma$  denotes the active function;  $\widehat{A}$  is the normalized format of the adjacency matrix plus self-loop connections  $\tilde{A}$ ,  $\widehat{A} = D^{-1/2} \tilde{A} D^{-1/2} = D^{-1/2} (A + I_N) D^{-1/2}$ .  $I_N$  is the identity matrix.

### 5.2.2 Self-adaptive Graph Convolutional Operation

The most important task in the graph convolutional layer is to determine the adjacency matrix  $A$ . Since the physical distribution power system can be seen as a graph, the adjacent matrix corresponding to the branch information of the distribution system can be extracted to form  $A$  in the GAECN. In this way, the neighbor information can be aggregated via the connections inside  $A$ , so that enhanced accuracy can be achieved. This process treats the correlations as a fixed matrix according to the natural connection of the power system. However, it cannot represent the implicit correlations among power injections brought by the dramatical uncertain nature of RES. To address this issue, the node embedding [102] that can represent the graph structure adaptively is employed to facilitate the graphical representation ability of the adjacency matrix  $A$ . To achieve this goal, the self-adaptive adjacency matrix  $\tilde{A}_{adp}$  is defined as:

$$\tilde{\mathbf{A}}_{adp} = \varphi \hat{\mathbf{A}} + (1-\varphi) \text{SoftMax}(\text{EeLU}(\mathbf{E}_1 \mathbf{E}_2^T)) \quad (5.5)$$

In (5.17),  $\varphi \in [0,1]$  are the weights of the  $\hat{\mathbf{A}}$ , which reflects the importance of the information of branches. Node embeddings  $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{N \times k}$  are the learnable parameters; The elements of  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are randomly initialized and are updated during the learning process;  $k$  is the dimension of embedding.  $\mathbf{E}_1$  and  $\mathbf{E}_2$  present two nodes, respectively. The multiplied value of them is used to denote the weights of correlation between the corresponding two nodes. During the training process, the implicit correlations can be learned and preserved  $\tilde{\mathbf{A}}_{adp}$  automatically. The ReLU activation function is employed to eliminate weak connections, while the SoftMax activation function is employed to normalize the self-adaptive adjacency matrix. Note that when  $\tilde{\mathbf{A}}_{adp} = \hat{\mathbf{A}}$  equivalent to  $\varphi=1$  with no node embedding, the model is defined as Non-adp-GAECN. Besides, when  $\varphi \neq 0$ , the corresponding model is defined as Bi-adp-GAECN. Most importantly, if  $\varphi=0$ , the self-adaptive graph convolutional operation is fully adaptive without any information of adjacent matrix, which is defined as Full-adp-GAECN. This setting allows the self-adaptive graph convolutional operation to fully capture the implicit correlations from the data itself. The performance of these models is represented later.

### 5.2.3 Convolutional Neural Network

The neural network obtains more and more academic attention due to its powerful ability of approximation. Conventionally, the fully connected neural network [103] treats the input information indiscriminately and ignores the uncertain influence among neighboring nodal features of the data. However, it is proved that the uncertain features of data can be learned via the convolutional operation layer by layer to improve the learning performance such as the image data [104]. Thus, the convolutional neural network is also utilized to handle the uncertainty of the power injections to facilitate

the approximation of the PPF.

The convolutional operation is one of the technologies of the neural network. To adapt the nodal power injections vector of the power system, the one-dimensional convolutional operation is proposed to extract the features of the nodal power injections instead of a three-dimensional convolutional operation which requires too many parameters. The one-dimensional convolutional operation is parameterized by the following equation with the input vector  $X \in R^{B \times N \times 1}$ .

$$Y(B, c, N - ks + 1) = \mathbf{b} + \sum_{c=0}^{C-1} \mathbf{W}(C, c) \otimes X(B, c, N) \quad (5.6)$$

In (5.6),  $\otimes$  the dot-product.  $c$  represents the number of channels.  $\mathbf{b}$  denotes the bias;  $\mathbf{W}$  is the kernel parameters.  $ks$  is the kernel's size.  $B$  denotes the batch size.  $N$  is the number of elements in the power injection sequence. During the training process, the features inside the window size of the kernel can be aggregated by a kernel convolution operation.

Then, the graph-aware scheme is introduced to capture the uncertain influence of the neighboring power injections, which is realized by defining the structure of the input of the proposed model. As it is defined and emphasized above,  $X=[P,Q]$  is the input of the proposed model. To fully utilize the neighboring information among nodes including the uncertain correlations of power injections due to the dependence of the output of wind, solar generations, and load demands, the sequence of the elements  $P,Q$  is defined as follows.

$$\begin{aligned} \mathbf{P} &= [P_{fh1}, P_{fh2}, \dots, P_{fhk}, \dots, P_n] \\ \mathbf{Q} &= [Q_{fh1}, Q_{fh2}, \dots, Q_{fhk}, \dots, Q_n] \end{aligned} \quad (5.7)$$

where  $fh$  denotes feeder  $h \in \{1, \dots, h, \dots, H\}$ ;  $k \in \{1, \dots, k, \dots, K\}$  denotes the number of lines in feeder  $h$ . This definition of the data inputs can facilitate the data being captured by the model graphically. This is because the nodes in a feeder are more related to each other geographically. Their data can be aggregated by the convolutional operation more directly. Furthermore, the nearby wind and solar generation outputs can be aggregated

after several convolutional operations. In this way, the complex correlations are extracted by several convolutional operations.

#### 5.2.4 Structure of GAECN

The structure of the model is also shown in Figure 5.1. It consists of an adaptive graph convolutional layer, two 1-D convolutional neural networks (CNN) layers, and a fully connected neural layer. The adaptive graph convolutional layer can perceive and learn the correlations hidden in the input data from the beginning. This is the reason it served as the first layer. The convolutional neural network layers can deeply learn the uncertainties brought by the RES. Thus, two 1-D convolutional layers are employed here instead of the conventional neural network. A fully connected neural layer is the output layer to obtain the aiming output size. Before the CNN layer, the dimension of the input data is reshaped form  $\mathbb{R}^{B \times N}$  to  $\mathbb{R}^{B \times N \times 1}$ . And after the CNN layer, the data is flattened into two dimensions.

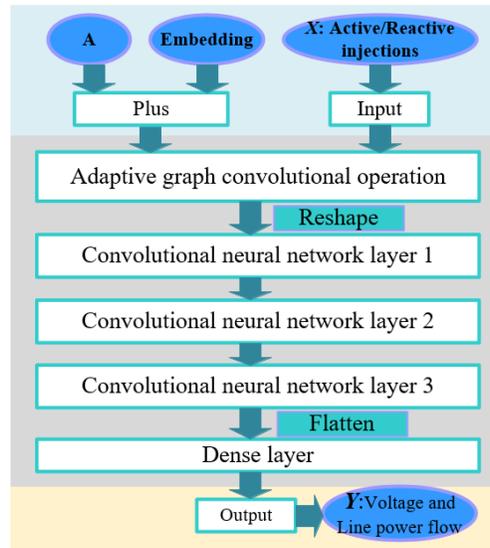


Figure 5.1. The structure of the GAECN.

The mean absolute error (MAE) is chosen to be the loss function of the proposed model, which is defined as:

$$Loss(\hat{Y}) = \frac{1}{N} \sum_{i=1}^N |\hat{Y} - Y| \quad (5.8)$$

In (5.20),  $\hat{Y}$  is the output of the proposed model,  $Y$  and denotes the actual value to guide the training process.

### 5.3 Case study

Several distribution systems, including the IEEE 33-node system, PG&E 69-node distribution system [5], 118-node system [105], and a real 76-node distribution system are employed to evaluate the performance of the proposed model. The Monte Carlo simulation is utilized to generate the power outputs of wind and solar, and load demand according to their parameters is shown in Table 5.1 and Table 5.2. The power outputs of wind and solar generators are generated by the following steps. Firstly, the wind power and solar power are generated following the Weibull and Beta density distribution functions (Parameters are in Table 5.1) and the load demand is assumed following the normal distribution with 20% of the mean value as the standard deviation. Secondly, the correlation coefficient matrix (Shown in Table 5.2) is utilized to generate the correlation injections via the  $X=L_c X_{in}$  (Eq. (5.13)). Note that the correlation coefficient matrix is added with the noise following the normal distribution  $Nor(1,0.1)$  in each injection sample. Based on these scenarios, the AC power flow is conducted by the Matpower toolbox to generate the data set. The generated data set includes the nodal net active and reactive power injections, which are the input data, and the nodal voltage magnitudes, voltage phase angles, and the line active and reactive power flow, which are the outputs of the model. After this data preparation process, the data size of the IEEE 33-node, PG&E 69-node, 118-node, and 76-node distribution systems are set to be 6.4M, 13.6M, 23.4M, and 19.4M respectively, with 90% as the training data and 10%

as the test data.

(1) System state variable error metrics:

In order to provide the indexes for evaluating the performance of the GAECN, the mean error and standard deviation error calculation is formulated.

$$\varepsilon_{\mu}^m = \left| \frac{\mu_x - \mu_{MC}}{\mu_{MC}} \right| \times 100\% \quad (5.9)$$

$$\varepsilon_{\sigma}^m = \left| \frac{\sigma_x - \sigma_{MC}}{\sigma_{MC}} \right| \times 100\% \quad (5.10)$$

In (5.9) and (5.10),  $x$  represents the PPF approaches apart from the MC methods.  $MC$  represents the MC methods in PPF, whose results are served as the true ground;  $\mu$  is the mean error of the expected value of PPF results between the method  $x$  and MC;  $\sigma$  is the mean error of the standard deviation value of the PPF results between the method  $x$  and MC;  $m$  denotes the mean or maximum (max) value of the error.

(2) Hyperparameters setup:

During the training process, the learning rates of GADNL and the employed baselines methods are set at 0.01. The size of the batch is set at 1000. The regularization coefficient is set at  $10^{-12}$ . These parameters are selected by optimization. The test system conducted by the GADNL is applied by python and performance in a computer with the windows 10 environment and Intel(R) Xeon(R) CPU E5-2650 v4 @ 2020GHz, and the graphics card of NVIDIA GeForce GTX 3090 24G.

(3) The parameters of the wind and solar power models:

The rated power of W1, W2, W3, and W4 are set at 0.06MW, and the rated power of W5 is 0.18MW. The rated powers of solar panels are 0.03MW. The wind speed and solar irradiance distribution functions' parameters are listed in Table 4.1. The correlation coefficient matrix of the RES is shown in Table 4.2. The location of the RES is listed in Table 5.1. The parameters wind-power curve are  $v_i = 4$  m/s,  $v_r = 14$  m/s and  $v_o = 25$  m/s, respectively.

Table 5. 1 The Locations of the RES units

33	Location	5	14	18	22	25	6	10	29	33
	Renewable	W1	W2	W3	W4	W5	S1	S2	S3	S4
	Rated power (MW)	0.06	0.06	0.06	0.06	0.18	0.03	0.03	0.03	0.03
69	Location	11	24	34	68	38	50	53	18	7
	Renewable	W1	W2	W3	W4	W5	S1	S2	S3	S4
	Rated power (MW)	0.06	0.06	0.06	0.06	0.18	0.03	0.03	0.03	0.03
118	Location	13	32	42	22	25	74	101	86	54
	Renewable	W1	W2	W3	W4	W5	S1	S2	S3	S4
	Rated power (MW)	0.6	0.6	0.6	0.6	1.8	0.3	0.3	0.3	0.3

### 5.3.1 Considering Independent RES Power Outputs

In order to verify the GAECN's performance in PPF calculation with independent renewable energy. Two traditional approaches, MC and PEM (Three-point) are introduced as the comparison methods. The PPF results of the MC method is served as the ground truth.

Table 5. 2 The mean error of comparison of different methods without correlation of IEEE 33 system.

Method	Error (%)	$V$	$\theta$	$P$	$Q$
PEM	$\varepsilon_{\mu}^{mean}$	0.0012	0.0210	0.1274	0.0769
	$\varepsilon_{\mu}^{max}$	0.0028	0.0909	0.8111	0.4983
	$\varepsilon_{\sigma}^{mean}$	1.3873	1.0123	1.8185	1.1087
	$\varepsilon_{\sigma}^{max}$	2.8460	2.3169	4.1460	3.0690
Full-adp-GAECN	$\varepsilon_{\mu}^{mean}$	0.0012	0.0759	0.1164	0.0553
	$\varepsilon_{\mu}^{max}$	0.0029	0.1488	0.5283	0.1594
	$\varepsilon_{\sigma}^{mean}$	1.1810	1.0014	0.9875	2.5813

	$\varepsilon_{\sigma}^{\max}$	4.0839	2.1299	6.4197	4.7623
--	-------------------------------	--------	--------	--------	--------

To represent the total simulation results of the GAECN, in comparison with the MC, the relative error with MC on voltage magnitudes, phase angles, active line flow, and reactive line flow are summarized in Table 5.2. The relative errors are all less than 7%. These results further demonstrate the effectiveness of the proposed GAECN in PPF under the independent renewable energy injections.

The distribution of the voltage magnitudes and active line power flow is characterized in Figure 5.2, where the nodes 7 and 17, lines (9,10) and (17,18) are selected. Note that the voltage magnitudes and the active line power flow are assumed normal distribution, which is based on the mean values and standard deviations extracted from the MC, PEM, GPE, PCE, and GAECN. In comparison with the baseline methods MC and PEM, the GAECN shows a highly similar distribution. Taking the results of MC methods as the baseline, the results of PEM and GADLN almost coincide with that of MC. This indicates that the proposed GADLN can calculate PPF with a high accuracy under the independent renewable energy injections.

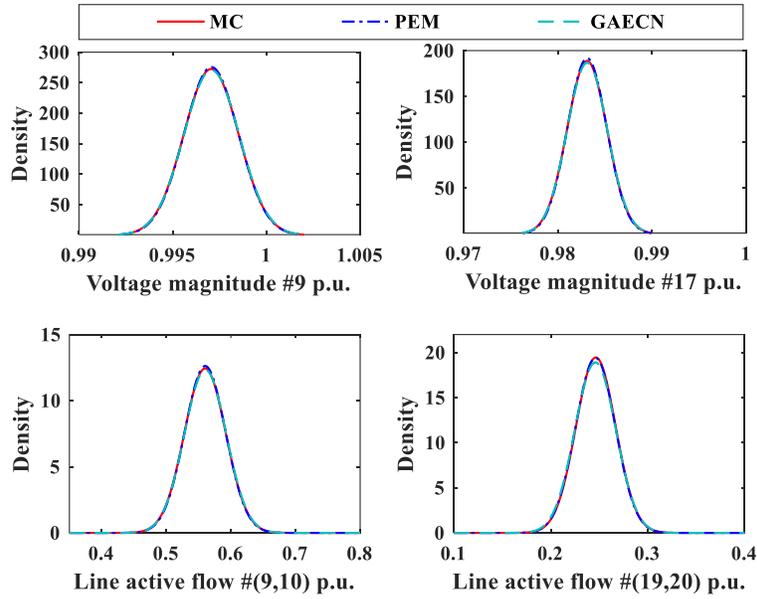


Figure 5.2. The voltage magnitudes and line active power flow without correlations.

Since the statistics information is compared for MC and GEACN without correlations, to further investigate the effectiveness of the proposed GEACN, the histograms of MC and GEACN are depicted in Figure 5.3. It shows that the distribution of the voltage magnitudes and the active line power flow are not a precise normal distribution. Besides, the histograms of GEACN are consistent with MC, which demonstrates that the GEACN effectively gives a high calculation accuracy in PPF without correlations.

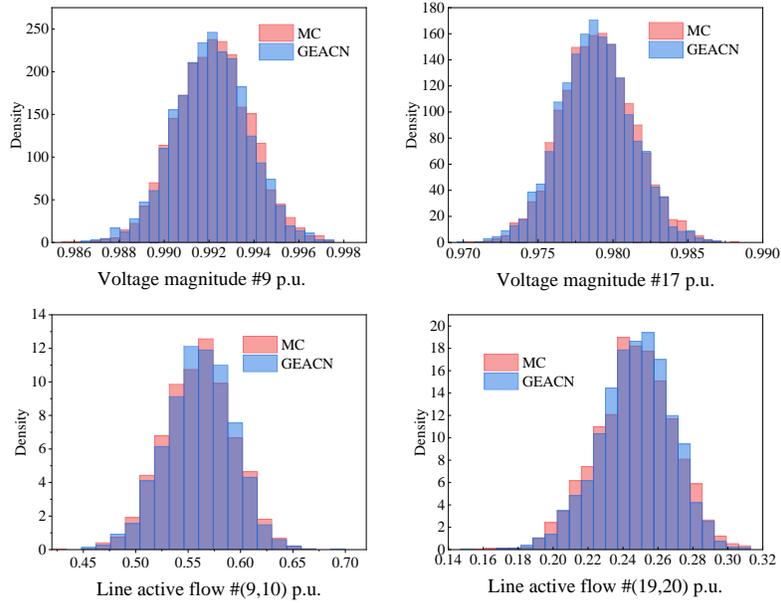


Figure 5. 3. The histograms of MC and GEACN without correlations.

### 5.3.2 Considering Correlation on RES Units

In order to verify the GADLN’s performance in PPF with renewable energy correlations, the correlation coefficient parameters are set in Table 4.2. The relative error with MC on voltage magnitudes, phase angles, active line flow, and reactive line flow is summarized in Table 5.3. It shows that the GAECN outperforms PEM by the percentages of 99.13%, 99.04% in mean value, the maximum value of the expected value of the voltage magnitudes, and 97.40%, 96.42% mean value, the maximum value of the standard deviation of the voltage magnitudes, respectively. Besides, The GAECN also outperforms PEM by the percentages of 99.99%, 99.20% in the mean value, the maximum value of the expected value of the active line power flow, and 95.27%, 95.85% in the mean value, the maximum value of the standard deviation of the active line power flow, respectively. This indicates that the proposed GAECN has a much smaller error in PPF calculation compared with that of the PEM. Therefore, the

proposed GAECN can well calculate the PPF in terms of effectiveness and robustness under the correlation of RES.

The distribution of the voltage magnitudes and active line power flow is characterized in Figure 5.4, where the nodes 7 and 17, lines (9,10), and (19,20) are selected. It shows that there is a significant bias between the distributions of state variables calculated by the PEM and the MC. It indicates that the comparison method PEM is unable to calculate the PPF with correlations considered. The GAECN can well fit the distribution of the state variables calculated by the MC method. This demonstrates that the GAECN outperforms the PEM method on the PPF calculation with correlations.

Table 5. 3. The mean error of comparison of different methods with the correlation of IEEE 33 system.

Method	Error (%)	$V$	$\theta$	$P$	$Q$
PEM	$\epsilon_{\mu}^{\text{min}}$	0.2927	2.0810	15.6586	6.1703
	$\epsilon_{\mu}^{\text{max}}$	0.5116	4.9379	237.91	44.2972
	$\epsilon_{\sigma}^{\text{min}}$	28.915	4.932	14.391	6.259
	$\epsilon_{\sigma}^{\text{max}}$	38.594	10.697	41.427	24.171
Full-adp-GAECN	$\epsilon_{\mu}^{\text{min}}$	0.0025	0.3519	0.4049	0.8598
	$\epsilon_{\mu}^{\text{max}}$	0.0049	0.7251	1.9623	4.3723
	$\epsilon_{\sigma}^{\text{min}}$	0.7541	1.1052	0.6810	1.6550
	$\epsilon_{\sigma}^{\text{max}}$	1.3821	1.9287	1.7057	5.5728

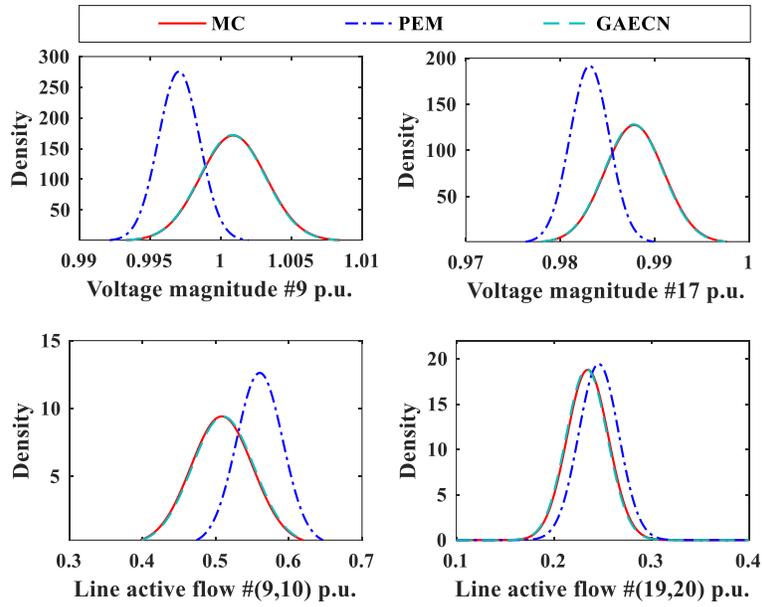


Figure 5. 4. The voltage magnitudes and line active power flow with correlations.

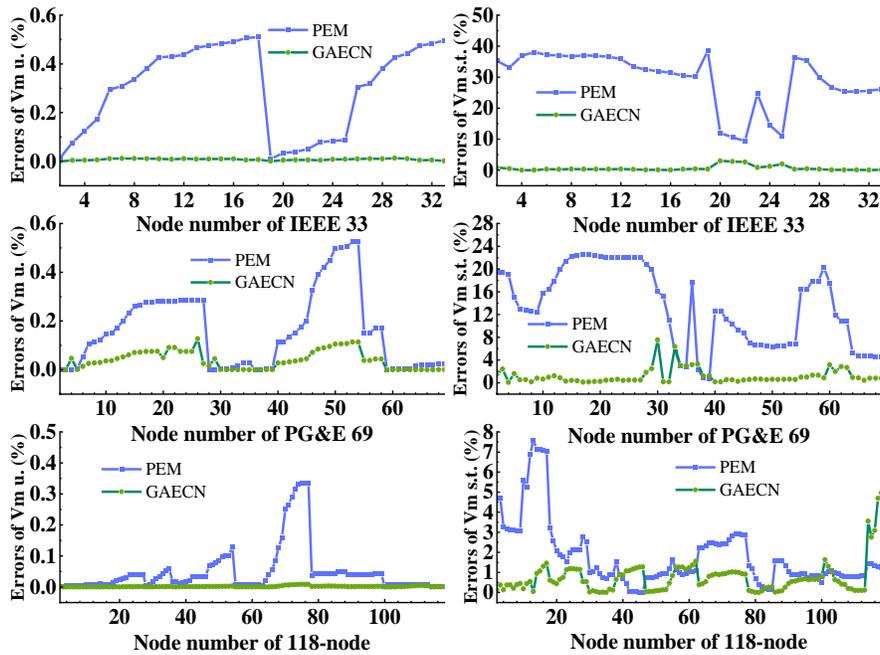


Figure 5. 5. The errors of voltage magnitudes with correlations.

To further investigate the performance of the proposed GAECN model, the errors of voltage magnitude mean value ( $V_m$  u.), and errors of voltage magnitude standard deviation ( $V_m$  s.t.) of PEM and GEACN compared with MC for IEEE 33-node system, PG&E 69, and 118-node system are depicted in Figure 5.5. The relatively large error of PEM shows that the correlations have a significant impact on conventional PPF results. The proposed GEACN can reduce the errors to below 5% for most of the voltage magnitude mean values and standard deviations.

Similarly, the statistics information for MC and GEACN with correlations is compared. To further investigate the effectiveness of the proposed GEACN, the histograms of MC and GEACN with correlations are depicted in Figure 5.6. It also shows that the histograms of GEACN are consistent with MC, which indicates that the GEACN effectively gives a high calculation accuracy in PPF with correlations.

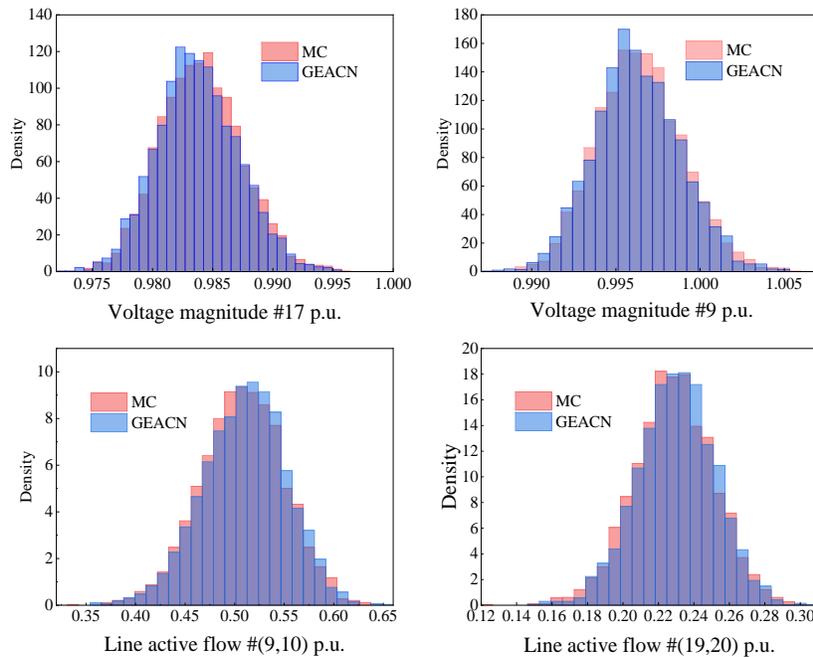


Figure 5. 6. The histograms of MC and GEACN with correlations.

### 5.3.3 The Performance of the GAECN

To verify the GADLN's performance in comparison with other deep learning methods, there are several cases employed as the baseline model. (1) a fully-connected neural network (FCN), with five layers parameterized by 100, 1500, 100, and 500 neurons for the first four layers; (2) A Convolutional Neural Network (CNN) model with three convolutional layers and three fully connected layers, where the sizes of its channels are 8, 8, and 8, respectively, and the kernel sizes are 5; (3) A graph convolutional network (GCN) [71] with two layers of graph convolutional layers, whose adjacent matrix is conducted based on the branch information; (4) The proposed GAECN model without adaptive operation, named Non-adp-GAECN; (5) The proposed GAECN with  $\varphi=0.1$ , named Bi-adp-GAECN; (6) The proposed GAECN with  $\varphi=0$ , named Full-adp-GAECN. The sizes of the channels of the proposed GAECN are 8, 8, and 8 respectively and the kernel sizes are 5. The structure of the methods (4), (5), and (6) are the same except for the parameters  $\varphi$ . Note that the structure of the CNN layers is chosen based on the following reasons. Because the nodes in a feeder of the distribution system are generally from 4 to 15, the kernel size is chosen to be 5 to make sure that the CNN can capture the information with 5 nodes together in the first CNN layer. By adding more than one CNN layer, the information outside 5 nodes while in a feeder can be further captured through the later CNN layers. Three CNN is enough to adapt to the distribution system and at the same time avoid the training burden. The size of channels is selected based on the GEACN performance.

The PPF calculation errors with the MC derived from different data-driven methods of the IEEE 33, PG&E 69, and 118-node distribution system are reported in Table 5.4. For IEEE 33 distribution system, compared with FCN, CNN, and GCN, the Full-adp-GAECN has a better performance on the mean error of expected value of voltage magnitudes with the percentages of 7.41%, 99.71%, and 99.11%, and the mean error of the standard deviation of the voltage magnitudes by 22.2%, 99.76%, and 99.14%, respectively. Besides, the Full-adp-GAECN also obtains much smaller maximum

errors of standard deviation in voltage phase angles by 75.78%, 72.07%, and 98.05%, respectively, compared with the baseline methods. This significant improvement of PPF by Full-adp-GAECN indicates the effectiveness of the proposed model.

Table 5. 4. The mean error of different data-driven methods with correlation.

Systems	Error (%)	IEEE 33		PG&E 69		118-node	
Methods		$V$	$\theta$	$V$	$\theta$	$V$	$\theta$
FCN	$\varepsilon_{\mu}^{mean}$	0.0027	0.3262	0.0004	0.0564	0.0018	2.7278
	$\varepsilon_{\mu}^{max}$	0.0063	0.6912	0.0013	0.1555	0.0064	204.67
	$\varepsilon_{\sigma}^{mean}$	1.164	0.8675	2.599	6.0818	6.9979	8.7629
	$\varepsilon_{\sigma}^{max}$	8.358	7.9670	70.641	80.525	61.946	65.450
CNN	$\varepsilon_{\mu}^{mean}$	0.8560	1.6663	0.2044	4.4926	0.2203	144.20
	$\varepsilon_{\mu}^{max}$	2.0174	4.7534	0.6856	10.958	1.4778	6391.7
	$\varepsilon_{\sigma}^{mean}$	6.9961	2.7658	7.0241	5.410	22.516	25.245
	$\varepsilon_{\sigma}^{max}$	17.791	6.9103	54.094	22.688	64.748	71.337
GCN	$\varepsilon_{\mu}^{mean}$	0.2819	0.5777	0.0061	0.1127	0.0067	1.8832
	$\varepsilon_{\mu}^{max}$	0.5385	1.1007	0.0254	0.4199	0.0178	128.24
	$\varepsilon_{\sigma}^{mean}$	87.560	96.523	65.318	79.328	86.724	99.111
	$\varepsilon_{\sigma}^{max}$	99.858	99.984	99.768	99.989	99.543	100.00
Non-adp-GAECN	$\varepsilon_{\mu}^{mean}$	0.0270	0.2228	0.0058	0.0710	0.0078	2.3583
	$\varepsilon_{\mu}^{max}$	0.1197	0.4158	0.0266	0.2345	0.0350	187.58
	$\varepsilon_{\sigma}^{mean}$	89.690	96.447	65.191	79.362	86.934	99.111
	$\varepsilon_{\sigma}^{max}$	99.843	100.00	99.224	100.00	99.890	100.00
Bi-adp-GAECN	$\varepsilon_{\mu}^{mean}$	0.0022	0.0400	0.0050	0.0021	0.0034	2.0103
	$\varepsilon_{\mu}^{max}$	0.0048	0.0938	0.0253	0.0113	0.0133	62.087
	$\varepsilon_{\sigma}^{mean}$	1.0773	1.5503	1.4730	0.9972	1.0903	0.8437
	$\varepsilon_{\sigma}^{max}$	6.1055	10.8101	3.5129	3.2800	5.4879	6.1453
Full-adp-GAECN	$\varepsilon_{\mu}^{mean}$	0.0025	0.3519	0.0017	0.1668	0.0048	0.8243

	$\varepsilon_{\mu}^{\max}$	0.0049	0.7251	0.0077	0.7645	0.0272	40.972
	$\varepsilon_{\sigma}^{\text{mean}}$	0.7541	1.1052	0.4281	1.3939	0.9486	0.6840
	$\varepsilon_{\sigma}^{\max}$	1.3821	1.9287	1.0427	2.8602	3.5158	4.5896

For the CNN and GCN, the accuracy reduces when the system scale increases, which demonstrates that the CNN cannot fully capture the complex graphical structure of the data, and the GCN cannot fully learn the uncertainties hidden in the data. The better performance of the Full-adp-GAECN indicates that it can fully perceive implicit correlations and the uncertainties of the data so that enhanced accuracy can be achieved.

Furthermore, compared with the Non-adp-GAECN, Bi-adp-GAECN, the Full-adp-GAECN improves the mean error of the expected value of voltage magnitudes by the percentages of 90.74%, 21.87%, and mean error of the standard deviation of the voltage magnitudes by 99.15%, 29.91%, respectively. And the Full-adp-GAECN also outperforms other PPF calculation results metrics. Besides, the Bi-adp-GAECN has a better performance compared with the Non-adp-GAECN, which demonstrates the effectiveness of the proposed full adaptive graph convolutional operation. This is because the full adaptive graph convolutional operation can capture the complex correlations among the injections extracted from the training data. The reported results indicate that the proposed Full-adp-GAECN can enhance the accuracy of the prediction results significantly either with or without correlations.

The calculation times of PPF with methods MC, PEM, GAECN for a 118-node distribution system are shown in Table 5.7. Note that the number of samples utilized in MC and GAECN is set at 5000. The average training time of each epoch for GAECN training is 3s and about 10000 epoch is needed. However, in comparison with the MC and PEM methods, the GAECN has a better calculation performance, which demonstrates the GAECN's efficiency.

Moreover, to intuitively deliver the correlations learned by the proposed GAECN, the proposed node embedding techniques based on self-adaptive graph convolutional operation parameters are depicted. Figure 5.7(a) is the heatmap of the IEEE 33 bus system's topology. Figure 5.7(b) denotes the correlation coefficient matrix. Figure

5.7(c) represents the parameters of the self-adaptive graph. In comparison with Figure 5.7(a) and Figure 5.7(b), it is obvious that the heatmap in Figure 5.7(c) is more complicated on the variety of the correlations among different nodes. This phenomenon indicates that only the linear correlation assumption and the topological structure cannot fully characterize the complex correlations hidden in the nodal power injections. On contrary, the proposed GAECN can capture hidden mapping from the injections to the power system states that involve correlations among nodal power injections and the physical topology.

Table 5.5. The calculation time of comparison of different methods.

Methods	MC	PEM	GAECN
Time (s)	18.44	2.304	0.4567

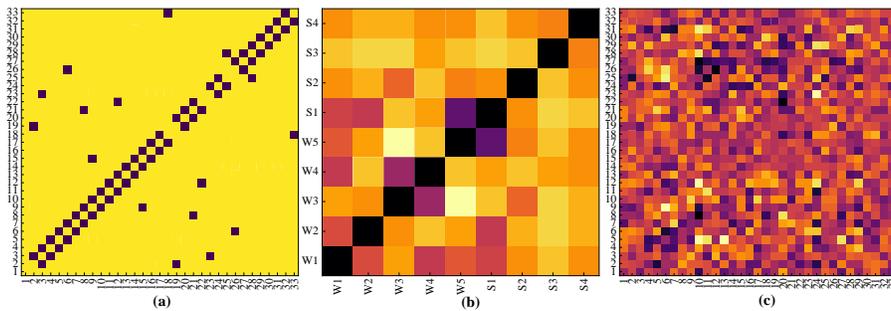


Figure 5. 7. (a) Adjacency matrix of topology. (b) Correlation coefficient matrix of RES. (c) Self-adaptive graph.

## 5.4 Summary

A novel GAECN approach is innovatively proposed to compute the PPF considering the implicit correlation of renewable energy. Compared with the conventional methods, the experiment results demonstrate that the proposed model has a significant

improvement in PPF calculation accuracy with less than 7% maximum error under correlations. Compared with the state-of-art deep learning methods, the maximum PPF accuracies are less than 10% either in test distribution systems or in the real system. This is because the full adaptive graph convolutional operation of the GAECN can deeply capture nodal power injections' correlation so that the enhanced effectiveness and efficiency of the PPF can be achieved. Thus, the GAECN model has a significant performance in PPF in terms of effectiveness, efficiency, and robustness against correlation power injections. And it shows an excellent application prospect in real power system operation.

# *Chapter 6 Conclusions and Future Works*

## **6.1 Conclusions**

The monitoring of the distribution system states faces considerable challenges due to uncertainty and intermittency brought by the increasing deployment of RES and limited measurement devices. The focus of this thesis is on overcoming these challenges through advanced system state estimation approaches. The graph embedded deep learning techniques is proposed to deal with the topology identification with very limited measurements, state estimation, and probabilistic power flow under high integrated RES. In particular, the author investigates the monitoring of distribution system states in the following four aspects.

- 1) To address the timely distribution grid topology identification problem with the limited presence of monitoring and measurement devices, a power distribution grid topological generative adversarial network (Gridtopo-GAN) model is proposed to deal with the distribution system topology identification issue threatened by the challenges of limited measurements and meshed structure. Specifically, an innovative topology preserved node embedding architecture is introduced to represent and compact the numinous topologies such that the topology identification in large-scale systems can be dealt with. The bad measurement data, as well as missing data, are not rare in practice, which inspires

the GAN with the generative capability the leveraged to ameliorate the robustness in the topology identification model. Numerical simulation results conducted on the 33/118/425-node systems demonstrate the effectiveness and time saving of the proposed model.

- 2) To timely perceive the distribution system states in a distribution system with high penetration of RES, the unrolled spatiotemporal graph convolutional network model is developed in this work for distribution system state estimation and forecasting that is exposed to complex correlations among the renewable power outputs. Specifically, three aspects of spatiotemporal correlations are captured simultaneously by the proposed unrolled spatiotemporal graph model that leverages the splicing of the spatial graphs across adjacent time steps. In this way, ameliorated forecasting accuracy and computational efficiency can be achieved. On top of this, the node embedding is leveraged to construct the dependence on the unrolled spatiotemporal graph to learn the nonlinear spatiotemporal correlations automatically instead of utilizing the linear correlation coefficient matrix that relies on full prior knowledge. Moreover, by stacking the spatiotemporal graph convolutional layers, the ahead-of-horizon state forecasting is achieved effectively. The simulation results based on the 118-node and 1746-node systems verify the accuracy and efficiency of the proposed model.
- 3) To represent the uncertain distribution system states quantificationally, a graph-aware deep learning network is leveraged to handle the probabilistic power flow that is exposed to complex dependence among the renewable power outputs. To fully capture the mapping from the fluctuated power injections and the uncertain system states, the convolutional operation is introduced to aggregate the correlations among renewables power outputs to facilitate the PPF so that the deviation pattern of the system state variables can be well learned. In this way, improved effectiveness and speed-up calculation can be achieved in the proposed model. Moreover, the numerical results conducted in the IEEE 33-node system

show the superior of the GADLN over the state-of-art with accurate and effective manners.

- 4) To calculate the PPF with the underlying complex correlation of uncertainties, graph attention enabled convolutional network is proposed to approximate PPF. Specifically, the graph attention enabled convolutional layer is proposed to aggregate the correlations of the power injections during the training process. Within this layer, the full self-adaptive graph convolutional operation is proposed to capture and learn any implicit correlation automatically so that significantly enhanced accuracy can be achieved. This layer is then followed by the convolutional neural network to capture the uncertain outputs of renewable energy to achieve the robust results of system state variable distributions. The numerical results conducted on the 33/69/118-node systems demonstrate the accuracy and efficiency of the proposed model.

## 6.2 Future Works

This thesis has proposed several advanced approaches for system state monitoring. Artificial intelligence-based methods have been developed to solve the topology identification, state estimation, and the probabilistic power flow. To extend this kind of approach to adapt to more application scenarios, the author will investigate the following problems in the future.

- 1) The distribution system is sometimes operated in meshed topology. Such structure with the integration of RES that presents a correlation nature will bring significant challenges to the topology identification problem. Therefore, the author will first endeavor to develop more deep learning-based methods to assist the distribution system topology identification problem with correlations.
- 2) Besides the impact factor of the correlations among the nodal power injections considered in chapter 3, the frequent topological changes will introduce a big

error in state estimation if it is ignored. Thus, the author will investigate an effective state estimation model to hedge against topology changes.

- 3) The probabilistic power flow represents the uncertain pattern on the system states. More efficient data-driven methods will be investigated to obtain higher moment information of the system states in PPF so that more accurate probabilistic distribution information can be leveraged to support the system operator decision-making.

# References

- [1] H. E. Murdock, D. Gibb, T. Andre *et al.*, “Renewables 2021-Global status report,” 2021.
- [2] H. Wu, P. Dong, and M. Liu, “Random fuzzy power flow of distribution network with uncertain wind turbine, PV generation, and load based on random fuzzy theory,” *IET Renewable Power Generation*, vol. 12, no. 10, pp. 1180-1188, 2018.
- [3] P. Shah and X. Zhao, “Network Identification using micro-PMU and Smart Meter Measurements,” *IEEE Transactions on Industrial Informatics*, 2022.
- [4] K. Dehghanpour, Z. Y. Wang, J. H. Wang *et al.*, “A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems,” *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312-2322, Mar 2019.
- [5] H. Wu, P. Dong, and M. Liu, “Distribution Network Reconfiguration for Loss Reduction and Voltage Stability With Random Fuzzy Uncertainties of Renewable Energy Generation and Load,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5655-5666, 2020.
- [6] S. Chai, Z. Xu, Y. Jia *et al.*, “A Robust Spatiotemporal Forecasting Framework for Photovoltaic Generation,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5370-5382, 2020.
- [7] E. Caro, A. J. Conejo, and R. Minguez, “Power System State Estimation Considering Measurement Dependencies,” *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1875-1885, 2009.
- [8] J. Jithendranath, D. Das, and J. M. Guerrero, “Probabilistic optimal power flow in islanded microgrids with load, wind and solar uncertainties including intermittent generation spatial correlation,” *Energy*, vol. 222, p. 119847, 2021.
- [9] S. Bolognani, N. Bof, D. Michelotti *et al.*, “Identification of power distribution network topology via voltage correlation analysis,” in *2013 IEEE 52nd Annual Conference on Decision and Control (Cdc)*, Firenze, Italy, 2013, pp. 1659-1664.
- [10] Y. Weng, Y. Z. Liao, and R. Rajagopal, “Distributed Energy Resources Topology Identification via Graphical Modeling,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682-2694, Jul 2017.

- [11] D. Deka, S. Backhaus, and M. Chertkov, "Estimating distribution grid topologies: A graphical learning based approach," in *2016 Power Systems Computation Conference (Psc)*, 2016, pp. 1-7.
- [12] C. Muscas, M. Pau, and P. A. Pegoraro, etc. , "Impact of input data correlation on Distribution System State Estimation," in *2013 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, pp. 114-119.
- [13] A. Ranković, B. M. Maksimović, and A. T. Sarić, et al., "ANN-based correlation of measurements in micro-grid state estimation," *International Transactions on Electrical Energy Systems*, vol. 25, no. 10, pp. 2181-2202, 2015.
- [14] M. Shafiei, G. Nourbakhsh, and A. Arefi, et al., "Single iteration conditional based DSSE considering spatial and temporal correlation," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 644-655, 2019.
- [15] J. Zhao, G. Zhang, and Z. Y. Dong, et al., "Robust Forecasting Aided Power System State Estimation Considering State Correlations," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2658-2666, 2018.
- [16] M. Aien, M. Fotuhi-Firuzabad, M. Rashidinejad *et al.*, "Probabilistic power flow of correlated hybrid wind-photovoltaic power systems," *IET Renewable Power Generation*, vol. 8, no. 6, pp. 649-658, 2014.
- [17] Y. Xu, L. Mili, and J. Zhao, "Probabilistic Power Flow Calculation and Variance Analysis Based on Hierarchical Adaptive Polynomial Chaos-ANOVA Method," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3316-3325, 2019.
- [18] X. Fang, S. Misra, G. L. Xue *et al.*, "Smart Grid - The New and Improved Power Grid: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 944-980, 2012.
- [19] D. Deka, S. Backhaus, and M. Chertkov, "Learning topology of the power distribution grid with and without missing data," in *2016 European Control Conference (Ecc)*, 2016, pp. 313-320.
- [20] R. E. Brown, "Impact of Smart Grid on distribution system design," in *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1-4.
- [21] Y. Zhang, J. Wang, and M. E. Khodayar, "Graph-Based Faulted Line Identification Using Micro-PMU Data in Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 3982-3992, 2020.

- [22] J. Zhao, C. Huang, L. Mili *et al.*, “Robust Medium-Voltage Distribution System State Estimation using Multi-Source Data,” presented at the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2020.
- [23] A. E. Saldaña-González, A. Sumper, M. Aragüés-Peñalba *et al.*, “Advanced Distribution Measurement Technologies and Data Applications for Smart Grids: A Review,” *Energies*, vol. 13, no. 14, p. 3730, 2020.
- [24] D. Deka, S. Talukdar, M. Chertkov *et al.*, “Graphical Models in Meshed Distribution Grids: Topology Estimation, Change Detection & Limitations,” *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4299-4310, 2020.
- [25] M. Mao, J. Xu, Z. Wu *et al.*, “A Multi-area State Estimation for Distribution Networks under Mixed Measurement Environment,” *IEEE Transactions on Industrial Informatics*, 2021.
- [26] Y. Liu, J. Sun, Q. Chen *et al.*, “Distribution Network Topology Error Identification Method Based on D-PMU and Branch State Function,” in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, Chengdu, China, 21-24 May, 2019.
- [27] G. Cavararo, V. Kekatos, and S. Veeramachaneni, “Voltage Analytics for Power Distribution Network Topology Verification,” *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 1058-1067, Jan 2019.
- [28] Y. Z. Liao, Y. Weng, and R. Rajagopal, “Urban Distribution Grid Topology Reconstruction via Lasso,” *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1- 5, 2016.
- [29] X. Li, S. Li, W. Li *et al.*, “Distribution Network Topology Reconstruction Method Based on Lasso and Its Supplementary Criteria,” *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, Cham., pp. 839 - 851, 2017.
- [30] J. Zhao, L. Li, Z. Xu *et al.*, “Full-Scale Distribution System Topology Identification Using Markov Random Field,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4714 - 4726, Nov 2020.
- [31] S. J. Pappu, N. Bhatt, R. Pasumarthy *et al.*, “Identifying Topology of Low Voltage Distribution Networks Based on Smart Meter Data,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113-5122, Sep 2018.

- [32] Y. Z. Liao, Y. Weng, M. Wu *et al.*, “Distribution Grid Topology Reconstruction: An Information Theoretic Approach,” presented at the 2015 North American Power Symposium (Naps), Charlotte, NC, Oct., 2015.
- [33] Z. S. Hosseini, M. Mahoor, and A. Khodaei, “AMI-Enabled Distribution Network Line Outage Identification via Multi-Label SVM,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5470-5472, 2018.
- [34] X. Z. Li, B. Han, G. J. Li *et al.*, “Dynamic Topology Awareness in Active Distribution Networks Using Blockchain-Based State Estimations,” *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5185-5197, Nov 2021.
- [35] M. Babakmehr, M. G. Simões, M. B. Wakin *et al.*, “Compressive Sensing-Based Topology Identification for Smart Grids,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 532-543, 2016.
- [36] L. Ding, W. Li, S. Nie *et al.*, “Topology Identification for Power Systems via Compressive Sensing Based on Transient Dynamics,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3202-3206, 2020.
- [37] Y. Sharon, A. M. Annaswamy, A. L. Motto *et al.*, “Topology identification in distribution network with limited measurements,” presented at the 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), Washington, DC, USA, Jan., 2012.
- [38] G. Cavraro, A. Bernstein, V. Kekatos *et al.*, “Real-Time Identifiability of Power Distribution Network Topologies With Limited Monitoring,” *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 325-330, 2020.
- [39] D. Deka, S. Backhaus, and M. Chertkov, “Learning Topology of Distribution Grids using only Terminal Node Measurements,” in *2016 IEEE International Conference on Smart Grid Communications (Smartgridcomm)*, Sydney, NSW, Australia, 2016.
- [40] D. Deka, S. Backhaus, and M. Chertkov, “Structure Learning in Power Distribution Networks,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061-1074, Sep 2018.
- [41] H. S. Karimi and B. Natarajan, “Joint Topology Identification and State Estimation in Unobservable Distribution Grids,” *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5299-5309, 2021.
- [42] H. Wu, Z. Xu, J. Zhao *et al.*, “Gridtopo-GAN for Distribution System Topology Identification,” *IEEE Transactions on Industrial Informatics*, pp. 1-1, 2022.

- [43] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: online learning of social representations," in *International Conference on Knowledge Discovery and Data Mining*, ed, 2014, pp. 701-710.
- [44] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*, 2017, pp. 2642-2651: PMLR.
- [45] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems 27 (Nips 2014)*, vol. 27, 2014.
- [46] J. Gui, et al., "A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications," *arXiv preprint arXiv:2001.06937*, 2020.
- [47] X. L. Lv, G. A. Bi, and C. R. Wan, "The Group Lasso for Stable Recovery of Block-Sparse Signal Representations," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1371-1382, Apr 2011.
- [48] R. A. Jabr, R. Singh, and B. C. Pal, "Minimum Loss Network Reconfiguration Using Mixed-Integer Convex Programming," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1106-1115, 2012.
- [49] J. T. Townsend, "Theoretical Analysis of an Alphabetic Confusion Matrix," *Perception & Psychophysics*, vol. 9, no. 1, pp. 40-50, 1971.
- [50] Y. Liao, Y. Weng, G. Liu *et al.*, "Urban MV and LV Distribution Grid Topology Estimation via Group Lasso," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 12-27, 2019.
- [51] J. Zheng, D. W. Gao, and L. Lin, "Smart Meters in Smart Grid: An Overview," in *2013 IEEE Green Technologies Conference (GreenTech)*, Denver, CO, USA, Denver, CO, USA, 2013, pp. 57-64.
- [52] A. Primadianto and C.-N. Lu, "A Review on Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875-3883, 2017.
- [53] S. Xin, Q. Guo, and H. Sun, et al., "Information-Energy Flow Computation and Cyber-Physical Sensitivity Analysis for Power Systems," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 2, pp. 329-341, 2017.
- [54] J. Zhao, G. Zhang, and M. L. Scala, et al., "Enhanced Robustness of State Estimator to Bad Data Processing Through Multi-innovation Analysis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1610-1619, 2017.

- [55] C. Muscas, M. Pau, and P. A. Pegoraro, et al., “Effects of Measurements and Pseudomeasurements Correlation in Distribution System State Estimation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2813-2823, 2014.
- [56] G. Valverde, A. T. Saric, and V. Terzija, “Stochastic Monitoring of Distribution Networks Including Correlated Input Variables,” *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 246-255, 2013.
- [57] S. Deshmukh, B. Natarajan, and A. Pahwa, “State Estimation and Voltage/VAR Control in Distribution Network With Intermittent Measurements,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 200-209, 2014.
- [58] M. Shafiei, G. Ledwich, and G. Nourbakhsh, et al., “One-Step Multilayer Unbalanced Three-Phase DSE for LV Distribution Networks,” *IEEE Systems Journal*, vol. 15, no. 3, pp. 4403-4412, 2021.
- [59] M. Hassanzadeh and e. a. Evrenosoglu, “A Short-Term Nodal Voltage Phasor Forecasting Method Using Temporal and Spatial Correlation,” *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3881-3890, 2016.
- [60] L. Wang, Q. Zhou, and S. Jin, “Physics-guided Deep Learning for Power System State Estimation,” *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 607-615, 2020.
- [61] L. Zhang, G. Wang, and G. B. Giannakis, “Real-Time Power System State Estimation and Forecasting via Deep Unrolled Neural Networks,” *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069-4077, 2019.
- [62] A. S. Zamzam and N. D. Sidiropoulos, “Physics-Aware Neural Networks for Distribution System State Estimation,” *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347-4356, 2020.
- [63] N. Bhusal, R. M. Shukla, and M. Gautam, et al., “Deep ensemble learning-based approach to real-time power system state estimation,” *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106806, 2021.
- [64] L. Zhao, Y. Song, C. Zhang *et al.*, “T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848-3858, 2020.
- [65] A. Mohamed, K. Qian, M. Elhoseiny *et al.*, “Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14424-14432.
- [66] P. Ghosh, Y. Yao, L. Davis *et al.*, “Stacked spatio-temporal graph convolutional networks for action segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 576-585.
- [67] C. Song, Y. Lin, S. Guo *et al.*, “Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 914-921, 2020.
- [68] P. A. Pegoraro, A. Angioni, M. Pau *et al.*, “Bayesian Approach for Distribution System State Estimation With Non-Gaussian Uncertainty Models,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 2957-2966, 2017.
- [69] B. Yu, H. Yin, and Z. Zhu, “Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting,” in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 3634-3640.
- [70] L. Ruiz, F. Gama, and A. Ribeiro, “Gated Graph Recurrent Neural Networks,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 6303-6318, 2020.
- [71] W. M. Kipf T N, “Semi-Supervised Classification with Graph Convolutional Networks,” *International Conference on Learning Representations (ICLR)*, 2017.
- [72] S. Zhang, H. Tong, and J. Xu, et al., “Graph convolutional networks: a comprehensive review,” *Computational Social Networks*, vol. 6, no. 1, pp. 1-23, 2019.
- [73] Y. L. a. R. Y. a. C. S. a. Y. Liu, “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting,” *International Conference on Learning Representations*, 2018.
- [74] Z. Wu, S. Pan, G. Long *et al.*, “Graph WaveNet for Deep Spatial-Temporal Graph Modeling,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 1907-1913.
- [75] P. J. Huber, “Robust Estimation of a Location Parameter,” Springer New York, 1992, pp. 492-518.
- [76] I. Pena, C. B. Martinez-Anido, and B.-M. Hodge, “An Extended IEEE 118-Bus Test System With High Renewable Penetration,” *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 281-289, 2018.
- [77] H. Ahmadi and J. R. Marti, “Distribution System Optimization Based on a Linear Power-Flow Formulation,” *IEEE Transactions on Power Delivery*, vol. 30, no. 1, pp. 25-33, 2015.

- [78] B. C. Meriem M, Abdelaziz B, et al., "Study of State Estimation Using Weighted-LeastSquares Method (WLS)," *2016 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, Oct. 2016.
- [79] M. Gol and A. Abur, "LAV Based Robust State Estimation for Systems Measured by PMUs," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1808-1814, 2014.
- [80] P. Chen, Z. Chen, and B. Bak-Jensen, "Probabilistic load flow: A review," in *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, Nanjing, China, 2008.
- [81] J. Zhang, G. Lin, and C. Y. Chung, "Instantaneous sensitivity identification in power systems - challenges and technique roadmap," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, USA, 2016.
- [82] M. Abdelaziz, "GPU-OpenCL accelerated probabilistic power flow analysis using Monte-Carlo simulation," *Electric Power Systems Research*, vol. 147, pp. 70-72, 2017.
- [83] G. E. Constante-Flores and M. S. Illindala, "Data-Driven Probabilistic Power Flow Analysis for a Distribution System With Renewable Energy Sources Using Monte Carlo Simulation," *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 174-181, 2019.
- [84] X. Xiaoyuan and Y. Zheng, "Probabilistic load flow evaluation with hybrid Latin Hypercube Sampling and multiple linear regression," in *2015 IEEE Power & Energy Society General Meeting*, Denver, CO, 2015.
- [85] M. Hajian, W. D. Rosehart, and H. Zareipour, "Probabilistic Power Flow by Monte Carlo Simulation With Latin Supercube Sampling," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1550-1559, 2013.
- [86] M. Tourandaz Kenari, M. S. Sepasian, M. Setayesh Nazar *et al.*, "Combined cumulants and Laplace transform method for probabilistic load flow analysis," *IET Generation, Transmission & Distribution*, vol. 11, no. 14, pp. 3548-3556, 2017.
- [87] W. Wu, K. Wang, and G. Li, et al., "Probabilistic load flow calculation using cumulants and multiple integrals," *IET Generation, Transmission & Distribution*, vol. 10, no. 7, pp. 1703-1709, 2016.
- [88] H. Wu, Y. Zhou, S. Dong *et al.*, "Probabilistic Load Flow Based on Generalized Polynomial Chaos," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 820-821, 2017.

- [89] C. S. Saunders, "Point Estimate Method Addressing Correlated Wind Power for Probabilistic Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1045-1054, May 2014.
- [90] X. Ai, J. Wen, T. Wu *et al.*, "A Discrete Point Estimate Method for Probabilistic Load Flow Based on the Measured Data of Wind Power," *IEEE Transactions on Industry Applications*, vol. 49, no. 5, pp. 2244-2252, 2013.
- [91] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish–Fisher expansion," *Electric Power Systems Research*, vol. 89, pp. 129-138, 2012.
- [92] Y. Chen, J. Wen, and S. Cheng, "Probabilistic Load Flow Method Based on Nataf Transformation and Latin Hypercube Sampling," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 294-301, 2013.
- [93] J.-Z. Zhu and Y. Zhang, "Probabilistic load flow with correlated wind power sources using a frequency and duration method," *IET Generation, Transmission & Distribution*, vol. 13, no. 18, pp. 4158-4170, 2019.
- [94] B. R. Prusty and D. Jena, "A Sensitivity Matrix-Based Temperature-Augmented Probabilistic Load Flow Study," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 2506-2516, 2017.
- [95] Y. Yang, Z. Yang, J. Yu *et al.*, "Fast Calculation of Probabilistic Power Flow: A Model-Based Deep Learning Approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2235-2244, 2020.
- [96] D. Wang, K. Zheng, Q. Chen *et al.*, "A data-driven probabilistic power flow method based on convolutional neural networks," *International Transactions on Electrical Energy Systems*, vol. 30, no. 7, 2020.
- [97] C. Chen, S. Duan, T. Cai *et al.*, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856-2870, 2011.
- [98] "The website for wind energy data in Switzerland" <https://wind-data.ch/tools/weibull.php>.
- [99] H. Wu, W. Minghao, Z. Xu *et al.*, "Probabilistic Power Flow of Distribution System Based on a Graph-Aware Deep Learning Network," in *2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, Chengdu, China, 2021.

- [100] H. Wu, M. Wang, Z. Xu *et al.*, “Graph Attention Enabled Convolutional Network for Distribution System Probabilistic Power Flow,” *IEEE Transactions on Industry Applications*, 2022.
- [101] Z. Zhang, P. Cui, and W. Zhu, “Deep Learning on Graphs: A Survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 249-270, 2022.
- [102] “Graph WaveNet for Deep Spatial-Temporal Graph Modeling,” *arXiv pre-print server*, 2019.
- [103] S.-C. Wang, “Artificial Neural Network,” in *Interdisciplinary Computing in Java Programming*, S.-C. Wang, Ed. Boston, MA: Springer US, 2003, pp. 81-100.
- [104] T. Guo, J. Dong, H. Li *et al.*, “Simple convolutional neural network on image classification,” in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Beijing, China, 2017.
- [105] A. A. E. Fergany and A. Y. Abdelaziz, “Reactive power compensation in distribution networks using cuckoo search algorithm,” *International Journal of Bio-Inspired Computation*, vol. 6, no. 4, pp. 230-238, 2014.