IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 August 2023, accepted 13 September 2023, date of publication 15 September 2023, date of current version 4 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3316017

RESEARCH ARTICLE

State Estimation for Reconfigurable Distribution Systems Using Graph Attention Network

ZHENGXING REN, (Graduate Student Member, IEEE), XIAODONG CHU, (Member, IEEE), AND HUA YE, (Member, IEEE)

Key Laboratory of Power System Intelligent Dispatch and Control of the Ministry of Education, Shandong University, Jinan 250100, China

Corresponding author: Xiaodong Chu (chuxd@sdu.edu.cn)

This work was supported by the Science and Technology Project of the State Grid Corporation of China under Grant 520600230020.

ABSTRACT Obtaining the real-time state of the distribution system is the basis for intelligent operation of the power system. With the surge in new energy generation and the volatility in load demands, traditional state estimation methods face significant challenges. The distribution system requires frequent topology reconfigurations to maintain stable operation. However, current data-driven methods typically cater only to specific topologies. To address this issue, a complete distribution system state estimation (DSSE) framework is proposed to adapt to frequent topology reconfigurations. To ensure data quality, a measurement device configuration algorithm using node importance was designed. Then, a convolutional neural network-based topology identification model is utilized to provide real-time topology data to the DSSE, which uses measured data pre-processed by the Gramian corner field. Finally, we use graph attention network to model the DSSE as a node-level regression prediction problem on a graph abstracted from the distribution system. Simulation results on IEEE 33-bus and IEEE 118-bus distribution systems illustrate the feasibility and efficiency of the proposed framework. Further experiments show that the proposed framework has good robustness.

INDEX TERMS Distribution system, graph attention network, state estimation, topology identification.

I. INTRODUCTION

Distribution system state estimation (DSSE) is a key tool in the intelligent operation of power systems which uses available measurement data to estimate unknown state variables. DSSE plays an important role in safety monitor and optimal schedule. Traditionally, most DSSE methods are designed for scenarios with fixed topology or a small number of reconfiguration switch actions due to the stable structure of the distribution system. However, with the growth of new energy generation and the access of unconventional loads, the volatility of the distribution system has greatly increased. Distribution system require frequent topology reconfiguration through automation devices to ensure safe operation. Complex combinations of section switches and interconnection switches lead to variable topologies and make hybrid operation of radial and looped topologies increasingly common. Topology could directly affect the distribution of power flow. If reconfiguration is ignored during DSSE, the state estimation error will not be confined to the topology reconfiguration region, but will spread to the entire distribution system. Frequent changes in topology have had a significant impact on DSSE, which poses a challenge to the operation and management of the power system. Therefore, this paper attempts to design a complete DSSE framework to accommodate topology reconfiguration. This is a challenging task because the DSSE accuracy has to be guaranteed in scenarios with scarce observability and frequent topology changes. To achieve this goal, We investigated real-time state estimation technique incorporating topological information. Meanwhile, the corresponding measurement configuration and topology identification (TI) methods are designed to provide data support for the framework.

The associate editor coordinating the review of this manuscript and approving it for publication was Elizete Maria Lourenco.

A. RELATED WORK

The DSSE methods are divided into traditional model-driven methods and data-driven methods. In terms of model-driven methods, A two-stage programming model was proposed to configuration phasor measurement unit (PMU) and

realization DSSE [1]. In [2], the AC optimal power flow method based on the polar coordinate power-voltage formulation and the rectangular current-voltage formulation are used for state estimation, respectively. Reference [3] proposed a DSSE model based on equivalent current measurements and an improved iterative algorithm based on the Krawczyk operator. Traditionally, the weighted least squares (WLS) methods have been widely used in DSSE. WLS is classified into node voltage based [4] and branch current based [5] when used as a state estimator. Gradient-based multi-region algorithm was applied to distribution system divided into combination of subtree areas to solve the WLS problem [6]. Reference [7] proposed a WLS-PMU Method based on a multi-grounded distribution network model. In [8], a Gauss-Newton iterative method based on multi-region architecture is used for distributed computation of the overall WLS. Currently, due to new energy access and demand response scheduling, the iterative algorithm-based WLS has become more sensitive to the operation state of the distribution system [9]. In [10], a more computationally efficient backward-forward sweep method was proposed as an alternative to WLS. However, both methods are not applicable to looped topology. Traditional methods rely heavily on wide-area measurement data. Another significant drawback of these methods is the high computational complexity, which increase significantly with the size of the system.

With the construction of advanced measurement systems around the world, the information systems of distribution system have gradually matured. Various data-driven DSSE methods represented by machine learning have been proposed in recent years. Reference [11] uses the Taylor series of voltages in the supervisory control and data acquisition (SCADA) system to construct linear state estimation models in interval form. In response to new energy access, a model-free DSSE method based on tensor completion was proposed in [12]. An adaptive particle filter (PF) method using randomized quasi-Monte Carlo sampling was proposed in [13] for enhancing the computational efficiency of PF methods. Kalman filter is also an important tool to solve DSSE [14], [15], [16], [17]. Reference [16] combined the parameter and state space model of the distribution system and solved it using an improved adaptive unscented Kalman filter algorithm. An improved extended Kalman filter algorithm using compressive sensing technique was proposed in [17]. To correct for spurious data injection, a DSSE method combining a noisy statistics estimator and a unscented Kalman filter was proposed in [18]. To improve convergence and stability, improved unscented Kalman filter and extended Kalman filter were embedded in the interactive multiple model framework of the joint DSSE [19]. To solve the initialization sensitivity problem of Gauss-Newton method, historical and simulated data are used to train the shallow neural network and map the measurements to the real state for initialization [20]. To cope with the perturbation of the measurements, [21] used an optimal filter and a Bayesian learning process for distributed state estimation. With the development of smart distribution system, PMU with high sampling rate is gradually popularized and provides real-time data support for distribution system [22]. A state estimator based on Bayesian inference method was proposed in [23] capable of integrating smart meter, SCADA system, and PMU and ensuring the observability through pseudo-measurement. In [24], a multi-task Gaussian method was used to process inhomogeneous measurement data consisting of advanced metering infrastructure (AMI), SCADA system, and PMU, followed by a hierarchical sparse Bayesian matrix for DSSE. This type of method can effectively cope with new energy access and has good real-time performance. However, the design idea still only considers the original topology.

As an important branch of machine learning, neural networks have advantages in dealing with regression problems by mapping input values through multilayer nonlinear structures. Currently, many researchers have applied artificial neural networks to DSSE problem [25], [26]. In [25], The effect of topology change is considered, but it is limited to the power on or off. In [27], a structural deep neural network (DNN) using an alternative autoencoder that incorporates physical constraints in the latent layers was proposed. A state estimator that combines a physical model based on power flow and a DNN was proposed in [28]. In [29], the deep recurrent neural network based state prediction model was used to provide partial data support for the DNN based state estimation model. In [30], DNN was used for both DSSE and topology identification (TI). A new model needs to be retrained after topology reconfiguration. To deal with low observability, [31] combined Monte Carlo simulation data with historical data to train DNN model. To reduce the coefficients of neural network mapping parameters, a physically aware neural network based on DSSE separability to prune unnecessary neural network connections was proposed in [32]. To enhance the robustness of DSSE, [33] proposed an long short-term memory (LSTM) based Metropolis-Hastings sampling method for outlier reconstruction. In [34], convolutional neural network (CNN) and LSTM were used for multi-level data recognition and state estimation. In [9], the dense residual neural network architecture was applied to DSSE. DSSE models trained using machine learning methods are usually topologically generalized, although they are trained for a single topology. However, these models usually suffer from overfitting problems and are not applicable to distribution system with complex topology changes. Compared to general neural networks, graph neural network (GNN) is more advantageous when dealing with distribution systems with natural graph properties. DSSE was considered as a node-level regression prediction problem in [35] and solved using GNN. An unrolled spatio-temporal graph convolutional network applied to the DSSE problem by improving the graph convolutional neural (GCN) was proposed in [36]. In [37], a physics informed GNN was proposed to improve the robustness to anomalous measurements. A DSSE method combining factor graph and GNN was proposed in [38], which has discussed the measurement

configuration methodology, but did not consider the impact of topology changes on the configuration plan and results. Currently, there are relatively few studies on GNN in DSSE.

The distribution system is undergoing unprecedented change. Consumers gain greater participation rights in distribution system operation through installing solar panels and other distributed generators. Operators are actively promoting the development of new energy consumption, microgrid construction, electric vehicle charging and power demand response markets [39]. The flexibility of the distribution system topology will continue to increase in the future. New trends in distribution system have simultaneously increased the demand for and the difficulty of real-time DSSE [40]. Therefore, this paper proposes a DSSE framework using machine learning methods. To ensure DSSE is accomplished in the correct topology, we choose PMUs with high rate advantage to obtain real-time measurement data. Considering the low observability of the distribution system, the Spearman's correlation coefficient (SCC) and the extreme gradient boosting (XGBoost) method are used to evaluate the importance of nodes in TI and DSSE. This paper combines the two types of importance and applies a key nodes selection algorithm to ensure data quality under reconfiguration. To accommodate the high rate of PMUs, neural network models are used for the DSSE framework. The CNN uses measurement data enhanced by the Gramian angular field (GAF) method to provide accurate real-time topology for the framework. Finally, DSSE is treated as a regression prediction problem on nodes. To ensure the topology reconfiguration adaptability of the model, this paper extends the graph attention network (GAT) to DSSE. As suitable for graph-structured data, GAT responds to topology changes by updating the input adjacency matrix online.

B. CONTRIBUTION

In summary, this paper proposes a DSSE framework to adapt to frequent topology reconfiguration. The main contributions of this paper can be summarized as follows:

- A DSSE framework using GAT is proposed. Only the adjacency matrix inputs need to be changed after topology reconfiguration without re-training the model. The proposed method shows a substantial increase in accuracy compared to neural networks designed for Euclidean data, and exhibits good robustness under noise, missing data, and unknown topology.
- An effective method for measurement configuration is proposed. The importance of the measurement data in DSSE and TI is combined to calculate the node importance. The method is able to provide the highest possible data quality for the DSSE framework during topology reconfiguration using a limited number of measurement devices.
- A topology identification method adapted to low observable measurement is proposed by synthesizing GAF and CNN, which has a great improvement in accuracy compared to the commonly used neural network methods. The proposed method is applicable to looped topology and can provide timely topology information to the framework.

C. PAPER ORGANIZATION

The rest of this paper is organized as follows. In Section II, the preparatory work before DSSE is introduced, including the dataset construction, the measurement device configuration method, and the TI method. In Section III, the structure of GAT and the training method of DSSE model are introduced. Then, the overall framework structure of DSSE is given. Section IV discusses the results of the study case. Section V draws the main conclusions.

II. STATE ESTIMATION PREPARATION

A. DATASET CONSTRUCTION

This paper constructed the dataset by collecting different topologies and corresponding node state information from historical measurement data of the distribution system. The DSSE framework obtained based on this dataset can accomplish the state estimation task at any time point. The various machine learning methods used in this paper are based on the dataset in this section.

$$\begin{cases} D = \left\{D^1, D^2, \ldots, D^k, \ldots, D^K\right\} \\ A = \left\{A^1, A^2, \ldots, A^k, \ldots, A^K\right\} \end{cases} \tag{1}$$

$$\begin{aligned} D^k &= \left\{D^{k,1}, \ldots, D^{k,t}, \ldots, D^{k,T}\right\} \\ &= \left\{\ldots, \left\{V^{k,t} \| \theta^{k,t}\right\}, \ldots, \left\{V^{k,T} \| \theta^{k,T}\right\}\right\} \\ &= \left\{\ldots, \begin{bmatrix} V_1^{k,t} & \ldots & V_i^{k,t} & \ldots & V_N^{k,t} \\ \theta_1^{k,t} & \ldots & \theta_i^{k,t} & \ldots & \theta_N^{k,t} \end{bmatrix}, \right. \\ &\qquad \left. \ldots, \begin{bmatrix} V_1^{k,T} & \ldots & V_i^{k,T} & \ldots & V_N^{k,T} \\ \theta_1^{k,T} & \ldots & \theta_i^{k,T} & \ldots & \theta_N^{k,T} \end{bmatrix}\right\} \end{aligned} \tag{2}$$

where D^k and A^k are the node feature dataset and adjacency matrix under topology k. $D^{k,t}$ is the node feature at time t under topology k. K is the number of typical operation topologies of the distribution system. T is the number of time points in the dataset. The node features include the voltage amplitude $V^{k,t}$ and the voltage phase angle $\theta^{k,t}$. $||$ is the matrix splicing operator. $V_i^{k,t}$ and $\theta_i^{k,t}$ are the voltage amplitude and phase angle of node i. N is the number of nodes in the distribution system.

The dataset is applied in the offline learning stage. In practice, the dataset consists of various data sources such as AMI, SCADA system, and PMU. The super-resolution method proposed in [41] can recover asynchronous data into high-resolution data for neural network model training.

B. MEASUREMENT DEVICE CONFIGURATION

Distribution systems are characterized by a large number of feeders. Due to the economics and communication pressures,

it is difficult to ensure that all nodes are observable. We need to configure the measurement devices at key nodes.

On the one hand, this paper uses partial nodes to map the state information of all nodes. Key nodes need to fulfill the data requirements of DSSE. On the other hand, the topology structure affects the node state by changing the power flow distribution. Selection of key nodes needs to consider the requirement of TI.

In this paper, DSSE is treated as a regression problem. Given that the GAT is used to aggregate neighbor information, the node degree and node feature correlation have a significant impact on the DSSE results. Therefore, SCC was used to quantify the importance of the nodes on the DSSE results. The correlation coefficient of voltage amplitude is calculated as:

$$\rho_{i,j}^{V}=\frac{1}{K}\sum_{k=1}^{K}\left(\frac{\sum_{t=1}^{T}\left(V_i^{k,t}-\bar{V}_i^{k,t}\right)\left(V_j^{k,t}-\bar{V}_j^{k,t}\right)}{\sqrt{\sum_{t=1}^{T}\left(V_i^{k,t}-\bar{V}_i^{k,t}\right)^2\sum_{t=1}^{T}\left(V_j^{k,t}-\bar{V}_j^{k,t}\right)^2}}\right) \tag{3}$$

where j is a neighbor node of i. The calculation method of correlation coefficient of voltage phase angle $\rho_{i,j}^{\theta}$ is the same as (3). The state estimation importance (SEI) of node i is calculated as:

$$\rho_i^{SE}=\sum_{j\in N(i)}\left(\rho_{i,j}^{V}+\rho_{i,j}^{\theta}\right) \tag{4}$$

where $N(i)$ is the set of neighbor nodes of node i.

In this paper, TI is set up as a multi-classification problem. The candidate features are the measurement data from each node. In general, the higher the importance of the feature, the more it contributes to the TI. To measure the importance of nodes on the TI results, the XGBoost method is used. This method achieves feature selection by quantifying the importance of all candidate measurements for the multi-classification problem [42]. The prediction result of the XGBoost is:

$$\hat{y}^{k,t}=\sum_{r=1}^{R}f_r\left(X^{k,t}\right) \tag{5}$$

where R is the number of decision trees. $f_r(X^{k,t})$ is the prediction score of $X^{k,t}$ in decision tree r. $X^{k,t}$ is the train sample, which can be chosen as $V^{k,t}$ or $\theta^{k,t}$. During the iteration, the loss function of the algorithm is:

$$Obj^{Ite}=\sum_{k=1}^{K}\sum_{t=1}^{T}L\left(y^{k,t},\hat{y}^{k,t}\right)+\sum_{r=1}^{R}\Omega\left(f_r\right) \tag{6}$$

where Ite is the iteration number. $y^{k,t}$ is the target value of the sample. $L(\cdot)$ is a residual function set as the square of the difference between the predicted values and target values. $\Omega(\cdot)$ is a penal ty function that evaluates the complexity of the model based on the decision tree:

$$\Omega\left(f_r\right)=\gamma L+\frac{1}{2}\lambda\sum_{l=1}^{L}w_l^2 \tag{7}$$

where L is the number of leaf nodes. w_l is the weight of leaf node l. γ is the penalty coefficient for the leaf node number. λ is the regularization term penalty coefficient [43]. The larger the two coefficients the more conservative the tree model. Then, converting the loss function to traverse through leaf nodes:

$$Obj^{Ite}=\sum_{l=1}^{L}\left[w_l\sum_{(k,t)\in I_l}g^{k,t}+\frac{1}{2}w_l^2\left(\sum_{(k,t)\in I_l}h^{k,t}+\lambda\right)\right]+\gamma L \tag{8}$$

where $g^{k,t}$ and $h^{k,t}$ are the 1-order and 2-order derivatives of the residual function. I_l is the sample set of leaf node l. The objective is to find the minimum value of the loss function. From (8), the quadratic function is minimized when w_l is set to (9).

$$w_l=-\frac{\sum_{(k,t)\in I_l}g^{k,t}}{\sum_{(k,t)\in I_l}h^{k,t}+\lambda} \tag{9}$$

Now, the loss function is transformed into (10):

$$Obj^{*}=-\frac{1}{2}\sum_{l=1}^{L}\frac{G_l^2}{H_l+\lambda}+\gamma L \tag{10}$$

where $G_l=\sum_{(k,t)\in I_l}g^{k,t}$. $H_l=\sum_{(k,t)\in I_l}h^{k,t}$. Then, the greedy algorithm is used to determine the structure of the decision tree. This method traverses all feature data starting from the root node. The feature data with the greatest information gain is selected as the split leaf node for the division condition. The information gain is calculated as:

$$G_{ain}=\frac{1}{2}\left(\frac{G_L^2}{H_L+\lambda}+\frac{G_R^2}{H_R+\lambda}-\frac{G_L^2+G_R^2}{H_L+H_R+\lambda}\right)-\gamma \tag{11}$$

where the three items in parenthesis are the optimal solution of the left decision subtree, right decision subtree, and pre-split decision tree. The decision tree stops growing when $G_{ain}<0$. After obtaining the decision tree, the topology identification importance (TII) of node i is calculated as:

$$\rho_i^{TI}=\sum_{X=V}G_{ain}^{i,V}+\sum_{X=\theta}G_{ain}^{i,\theta} \tag{12}$$

where $G_{ain}^{i,V}$ and $G_{ain}^{i,\theta}$ are the information gain when the voltage amplitude and phase angle of node i are used as the division conditions, respectively.

The node importance in the DSSE framework of this paper is calculated through two types of importance.

$$\rho_i=\frac{\rho_i^{SE}}{\sum_{j=1}^{N}\rho_j^{SE}}+\frac{\rho_i^{TI}}{\sum_{j=1}^{N}\rho_j^{TI}} \tag{13}$$

Finally, the key nodes are selected to configure the PMUs according to the algorithm in Figure 1.

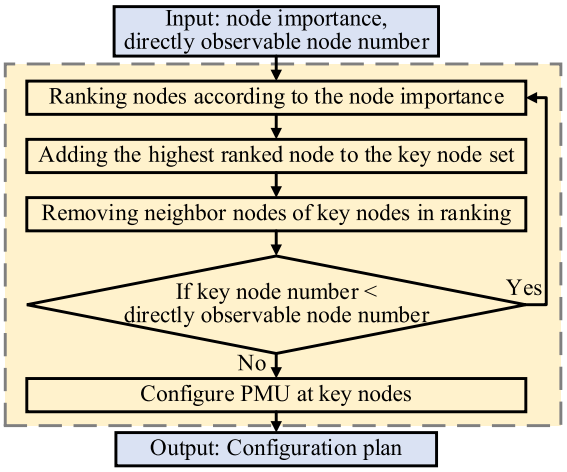


FIGURE 1. Key nodes selection algorithm.

This algorithm effectively merges the importance of DSSE and TI to accommodate the low observability of the distribution system. It also promotes a uniform distribution of PMUs to ensure the quality of the measured data after topology reconfiguration.

C. TOPOLOGY IDENTIFICATION METHOD

In this paper, a neural network model is used to establish the mapping relationship from measurement data at time points to the topology. The model uses the PMU configured in Section II-B to ensure real-time performance. The voltage amplitude and phase angle are related to the topology as:

$$\left(V^{k,t}, \theta^{k,t}\right) = F_{flow}\left(P^{k,t}, Q^{k,t}, A^{k}\right) \tag{14}$$

where $P^{k,t}$ and $Q^{k,t}$ are the active and reactive power at time t under topology k. $F_{flow}(\cdot)$ is the power flow equation.

Since topology has a significant impact on the distribution of power flow. There is a correlation between the node state information and the distribution system topology [44]. The mapping relationship can be expressed as:

$$A^{k} = F_{regu}\left(V_i^{k,t}, \theta_i^{k,t} | i \in N\,(P)\right) \tag{15}$$

where $F_{regu}(\cdot)$ is the mapping function. $N(P)$ is the set of node numbers configured with PMU. $N(P)$ is obtained from Section II-A and the total number is N_P. Therefore, this paper uses the PMU data of partial nodes to learn the mapping function.

Typically, time data is one-dimensional, but in fact another hidden dimension is time. GAF can be used to elevate the dimensionality of the time data [45]. In this paper, the nodes configured with PMU are numbered in a fixed order. GAF is extended to handle node state information at a specific time point. This method can reflect the correlation information between nodes while preserving the original node state information. In this paper, GAF is used to transform one-dimensional data into two-dimensional graph data, facilitating neural network models for more effective data mining [46]. The data pre-processing process is shown below.

Normalization of node voltage amplitude measurement data:

$$\hat{V}_i^{k,t} = \left\{ \frac{V_i^{k,t} - \min\left(V^{k,t}\right)}{\max\left(V^{k,t}\right) - \min\left(V^{k,t}\right)} \quad i \in N\,(P) \right. \tag{16}$$

where $\hat{V}_i^{k,t}$ is the normalized voltage amplitude.

Convert the normalized measurement data to the polar coordinate system:

$$\phi_i^{k,t} = \arccos V_i^{k,t} i \in N\,(P) \tag{17}$$

Trigonometric transformation of measurement data into the polar coordinate system (18), as shown at the bottom of the page, where the segment length is set to 1 during the transformation process to preserve as much topology information as possible. The voltage phase angle measurement data are pre-processed using the same method. Ultimately, three-dimensional data consisting of voltage amplitude and phase angle were obtained.

Then, CNN is used to extract the high-dimensional spatial features of the three-dimensional data. The output is the classification label corresponding to the typical operation topology of the distribution system. In this paper, the CNN classifier consists of a convolutional encoder and a multilayer perceptron, as shown in Figure 2.

The layers in the convolutional encoder correspond to 16, 32, 64 and 128 feature mapping maps, respectively. An average pooling layer is added after the last CNN layer to compress the input features, which could filter out some of the redundant information. Multiple convolution and pooling operations can effectively reduce the effect of noise in the measurement data. The multilayer perceptron consists of three fully connected layers with hidden layer sizes of 128, 256 and the number of typical operation topologies of the distribution system. The CNN classifier returns the category probability between 0-1 for each topology label. The model parameter learning process uses cross-entropy as the back-propagation loss function. The multi-classification

$$G = \begin{bmatrix} \cos\left(\phi_1^{k,t} + \phi_1^{k,t}\right) & \cos\left(\phi_1^{k,t} + \phi_2^{k,t}\right) & \cdots & \cos\left(\phi_1^{k,t} + \phi_{N_p}^{k,t}\right) \\ \cos\left(\phi_2^{k,t} + \phi_1^{k,t}\right) & \cos\left(\phi_2^{k,t} + \phi_2^{k,t}\right) & \cdots & \cos\left(\phi_2^{k,t} + \phi_{N_p}^{k,t}\right) \\ \vdots & \vdots & \vdots & \vdots \\ \cos\left(\phi_{N_p}^{k,t} + \phi_1^{k,t}\right) & \cos\left(\phi_{N_p}^{k,t} + \phi_2^{k,t}\right) & \cdots & \cos\left(\phi_{N_p}^{k,t} + \phi_{N_p}^{k,t}\right) \end{bmatrix} \tag{18}$$

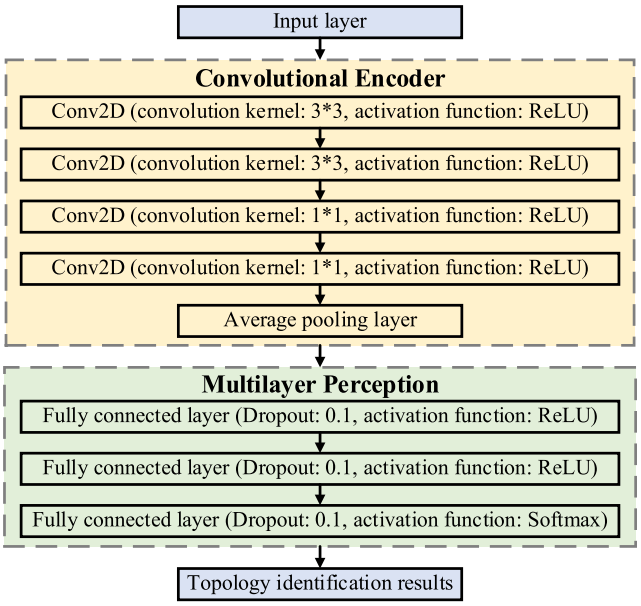


FIGURE 2. Structure of the distribution system TI model.

cross entropy function is used to calculate the gap between the predicted and true values of the model, as in:

$$loss = -\sum_{k=1}^{K} y_k \log p_k \tag{19}$$

where y is the label of topology k. p is the output of the neural network, which is the probability that the category is k.

III. DISTRIBUTION SYSTEM STATE ESTIMATION FRAMEWORK

A. DISTRIBUTION SYSTEM STATE ESTIMATION

In order to perform out tasks such as optimal scheduling and fault monitoring of the distribution system, the state information of each node needs to be available in real time through DSSE. The time series data that the distribution system can obtain online through the PMU are shown below.

$$\begin{aligned} H &= \left\{ h_1 \ h_2 \ h_3 \ h_4 \ \cdots \ h_i \ \cdots \ h_N \right\} \\ &= \begin{bmatrix} 0 & V_2 & V_3 & 0 & \cdots & V_i & \cdots & 0 \\ 0 & \theta_2 & \theta_3 & 0 & \cdots & \theta_i & \cdots & 0 \end{bmatrix} \end{aligned} \tag{20}$$

where H is the online measurement matrix of the distribution system at any time point. h_i is the node feature of node i. V_i is the voltage amplitude. θ_i is the voltage phase angle. Due to the limited number of PMU configurations, the measurement data matrix is sparse. The feature data of the nodes without PMU configuration is initialized to 0.

The objective of DSSE is to solve for the complete voltage amplitude and phase angle using real-time measurement data [47]. The voltage amplitude and phase angle of the nodes without PMU configuration are used as the state variables to be solved. This paper uses neural networks to mine the correlation between PMU measurement data and state variables. Specifically, the state variables at the current time point will be estimated by the neural network model trained from the historical measurement dataset consisting of the various time points. As shown in (21), the state variables are estimated by calculating the expectation values of the posterior conditional distribution.

$$(V, \theta) = \int_h p\,(V, \theta | h)\, p\,(h) dh \tag{21}$$

where $p(V, \theta|h)$ is a conditional probability model for the state variables under the known measurement data h. $p(h)$ is the probability distribution model. In this paper, GNN is used to construct the DSSE model.

B. THE GRAPH ATTENTION NETWORK

GNN is a machine learning method that has emerged in recent years. This method introduces graph theory into neural networks to achieve the fusion of deep learning and graph data [35]. Distribution system have natural non-Euclidean graphical structure properties [36]. The topology of the distribution system can be represented by an undirected graph $G = \{N, E, A$. N is the node set consisting of distribution system nodes. E is the edge set consisting of overhead and cable lines. A is the adjacency matrix of the topology. Unlike general neural networks, GNN is able to handle non-Euclidean data with topology information. This paper uses the GAT, a development branch of GNN, to implement DSSE. GAT uses the attention coefficients calculated through the attention mechanism as weights for information transmission between nodes. The DSSE model trained by GAT is able to achieve effective aggregation of neighbor node information. This method is suitable for state estimation of the distribution system with frequent topology reconfiguration.

In aggregating node information, GAT uses the attention coefficient as a weight for information dissemination [48]. First, the correlation coefficient e is calculated using the node features.

$$e_{i,j} = a\left(\left[Wh_i || Wh_j\right]\right) \ j \in N\ (i) \tag{22}$$

where $e_{i,j}$ is the correlation coefficient between node i and neighbor node j. W is a learnable shared weight parameter used as a linear transformation. W enhances the expressive capability of the model by increasing the dimensionality of the node feature h. During model training, W is updated at each iteration through backward propagation process and gradient descent optimization algorithm. $||$ is used to splice the node features. $a(\cdot)$ is a feature mapping function consisting of a single-layer feedforward neural network.

Then, the attention coefficient matrix a is obtained by normalization using the SoftMax function. GAT considers its own node information while calculating the attention score.

$$a_{i,j} = \frac{exp\left(LeakyReLU\left(e_{i,j}\right)\right)}{\sum_{n \in N(i) \cup i} exp\left(LeakyReLU\left(e_{i,n}\right)\right)} \tag{23}$$

where $a_{i,j}$ is the attention coefficient between node i and neighbor node j. $LeakyReLU(\cdot)$ is the activation function.

Finally, new node features are formed by aggregating information using the attention coefficient matrix. To enhance the

performance of the attention layer, this paper introduces a multi-head attention mechanism to obtain aggregated results.

$$h_i'(N_h) = \frac{1}{N_h}\sum_{n_h=1}^{N_h}\sigma\left(a_{i,i}^{n_h}W^{n_h}h_i + \sum_{j\in N(i)}a_{i,j}^{n_h}W^{n_h}h_j\right) \quad (24)$$

where σ is an activation function. N_h is the number of attention heads in GAT. The introduction of multi-head attention will lead to multiple aggregated results. In this paper, all results are averaged to obtain the final output state variable values. The information transfer process of GAT is shown in Figure 3.

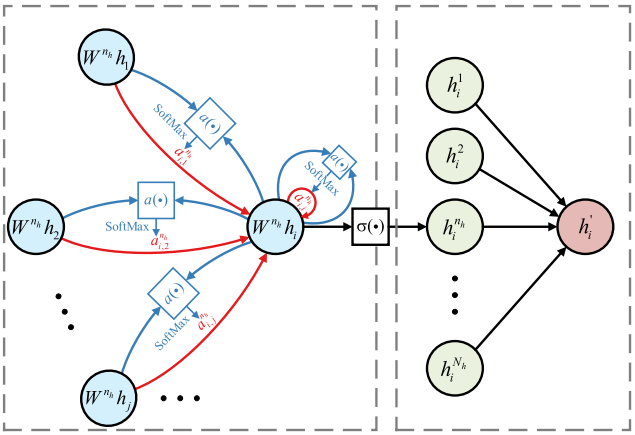


FIGURE 3. Information transfer process of GAT.

C. STATE ESTIMATION MODEL STRUCTURE

In this paper, GAT is applied to the problem of regression prediction on nodes. As shown in (25), when training the parameters of the DSSE model, the inputs are the adjacency matrix A^* corresponding to topology, the label Y, and the node feature matrix X.

$$\begin{cases} A^* = A^k \\ Y = \left\{D^{k,1},\ldots,D^{k,t},\ldots,D^{k,T}\right\} \\ X = \left\{D^{k,1}\odot C,\ldots,D^{k,t}\odot C,\ldots,D^{k,T}\odot C\right\} \end{cases} \quad (25)$$

where $\odot$ is the multiplication operation of the elements in the corresponding positions. C is a measurement data mask of the same length as D^k.

The value of the element in the measurement data mask is determined according to (26).

$$c_i = \begin{cases} 1, & if\ i \in N(P) \\ 0, & else \end{cases} \quad (26)$$

Typically, the adjacency matrix A^* is very sparse. Each GAT layer in the DSSE model can aggregate the 1-order neighbor node information. When constructing the DSSE model, a reasonable number of GAT layers must be determined. A too small number of GAT layers results in ineffective information transfer between distant nodes. Conversely, too many GAT layers can lead to overfitting and training difficulties. The structure of the DSSE model in this paper is shown in Table 1.

During each iteration, the DSSE model provides state estimation results consisting of node voltage amplitude and phase angle of the same size as the training set. The mean square error is used as the loss function in the training process.

$$\begin{cases} L_{mse}^{V} = \dfrac{1}{M_{Grain}}\displaystyle\sum_{m=1}^{M_{Gran}}\sum_{i=1}^{N}\left(V_{m,i}-\hat{V}_{m,i}\right)^2 \\ L_{mse}^{\theta} = \dfrac{1}{M_{Gtrain}}\displaystyle\sum_{m=1}^{M_{Gratan}}\sum_{i=1}^{N}\left(\theta_{m,i}-\hat{\theta}_{m,i}\right)^2 \end{cases} \quad (27)$$

where M_{Gtrain} is the number of samples in the state estimation training set. $\hat{V}_{m,i}$ and $\hat{\theta}_{m,i}$ are the estimated node voltage amplitude and phase angle.

When evaluating the state estimation results, the mean absolute percentage error (MAPE) is used for the voltage amplitude. The node mean absolute percentage error (N-MAPE) of the voltage amplitude for any topology is:

$$\gamma_i^V = \frac{1}{M_{Gtest}}\sum_{m=1}^{M_{Gist}}\frac{\left|V_{m,i}-\hat{V}_{m,i}\right|}{V_{m,i}} i \in N\left(P^*\right) \quad (28)$$

where M_{Gtest} is the number of samples in the state estimation test set. $N(P^*)$ is the set of node numbers configured without PMU. N_{P^*} is the total number of nodes in $N(P^*)$. The MAPE for any topology is calculated by N-MAPE.

$$\gamma^{\theta} = \frac{1}{N_{P^*}}\sum_{i=1}^{N_{p^*}}\gamma_i^{\theta} i \in N\left(P^*\right) \quad (29)$$

TABLE 1. Structure of the DSSE model.

Number of layers	Voltage amplitude				Voltage phase angle			
	Structure	Activation function	Input	Output	Structure	Activation function	Input	Output
Layer 1	GAT	ReLU	N	N*8	GAT	ReLU	N	N*8
Layer 2	GAT	ReLU	N*8	N*16	GAT	ReLU	N*8	N*16
Layer 3	GAT	ReLU	N*16	N*32	GAT	ReLU	N*16	N*32
Layer 4	GAT	ReLU	N*32	N*64	GAT	ReLU	N*32	N*64
Layer 5	Batch normalization	-	N*64	N*64	GAT	ReLU	N*64	N*128
Layer 6	Dense	Sigmoid	N*64	N*32	GAT	ReLU	N*128	N*256
Layer 7	Dense	Sigmoid	N*32	N	Dense	Tanh	N*256	N*128
Layer 8	-	-	-	-	Dense	Tanh	N*128	N

In addition, the node mean absolute error (N-MAE) of the voltage phase angle for any topology is:

$$\gamma_i^{\theta} = \frac{1}{M_{Gtest}} \sum_{m=1}^{M_{Gimes}} \left| \theta_{m,i} - \hat{\theta}_{m,i} \right| i \in N\left(P^*\right) \tag{30}$$

The mean absolute error (MAE) is used as the evaluation index of the voltage phase angle state estimation results.

$$\gamma^{\theta} = \frac{1}{N_{P^*}} \sum_{i=1}^{N_{p^*}} \gamma_i^{\theta} i \in N\left(P^*\right) \tag{31}$$

D. FRAMEWORK STRUCTURE

As shown in Figure 4, a DSSE framework is proposed to meet the demand for real-time awareness of distribution system operation state under frequent topology reconfiguration.

The offline learning stage initiates with the construction of the historical measurement dataset. To handle the challenges posed by low observability and topology reconfiguration, we compute the node importance by considering both DSSE and TI requirements for measurement data. PMUs are configured at key nodes selected based on their importance. Then, the dataset is pre-processed using GAF and the TI model is trained using CNN. Finally, the dataset and topology are used as input to the GAT for training the DSSE model.

In the online stage, the data center collects measurement data from PMUs. The TI model is invoked to identify the distribution system topology. Then, the DSSE model estimates the voltage amplitude and phase angle of each node based on the latest TI results. If a new topology is detected, the topology generalization capability of the existing DSSE model is used to perform state estimation. This framework is able to obtain the operation state online at any given time point. These data can serve as the foundation for distribution system state analysis and optimal dispatching.

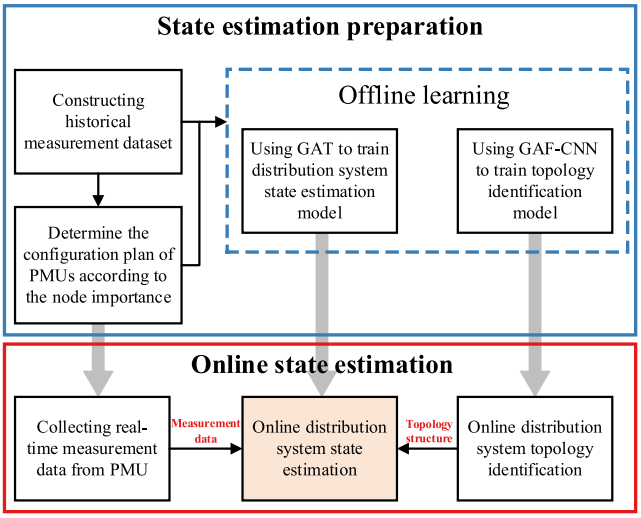


FIGURE 4. DSSE framework.

The proposed framework offers the following benefits:

- Economical aspects: The increased frequency of topology reconfiguration requires more measurement devices. This framework only requires measurement data at a few key nodes to perform DSSE.
- Real-time aspect: The neural network model used in this paper runs with short time. The accuracy of TI and DSSE does not change significantly with increasing system size.
- Robustness aspect: GAF-CNN can efficiently mine the relationship between node measurement data and topologies. The information transfer process of GAT ensures robustness under topology reconfiguration.

IV. CASE STUDIES

A. STATE ESTIMATION RESULTS

This paper uses an IEEE 33-bus distribution system as a test case. As shown in Figure 6, this system has 32 branches and 4 feeders. There are interconnection switches between nodes 8-21, 9-15, 12-22, 18-33 and 25-29. In this paper, the dataset collects PMUs measurement data at 1min intervals for each time point. Except for the parent node, all the nodes have load and distributed PV access. Daily load curves and solar power curves are used to simulate the variation of load and PV generation. The MATPOWER toolbox is used to generate 24h steady-state flow data for 200 topologies based on different combinations of section switches and interconnection switches. The topology library contains 128 radial and 72 looped distribution system typical operation topologies. Therefore, the dataset contains 288,000 sets of samples. Each set of samples contains the topology number and the voltage amplitude, voltage phase angle, injected active power, and injected reactive power of all nodes. Pytorch v.1.13.0 was used in Python v.3.7 to implement the machine learning algorithms in this paper. All simulations were performed on a computer with 32-GB RAM, Intel Core 11700 CPU at 2.50 GHz and Nvidia T600 4-GB GPU.

First, the node importance is calculated using the SCC and XGBoost. In XGBoost, γ is set to 0 and λ is set to 0.5. After that, the key nodes are selected. The node importance results are shown in Figure 5.

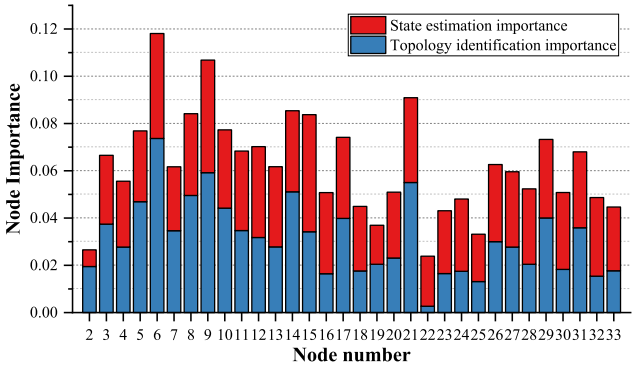


FIGURE 5. Node importance results.

Now, there is insufficient measurement capacity in the distribution system. This situation will be difficult to improve in short term due to the large scale. The number of nodes configured with PMUs in the distribution system is typically

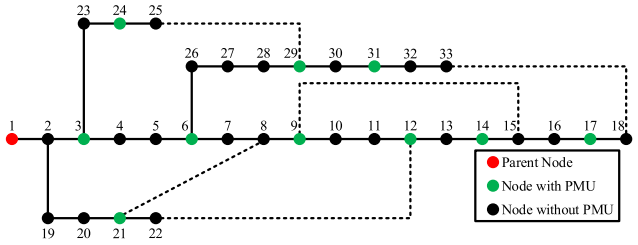


FIGURE 6. Configuration plan of the measurement devices.

around one-third or even less. Therefore, this paper selects the top ten nodes in terms of node importance to configure the PMU, as shown in Figure 6.

In addition, the parent node is also configured with advanced measurement devices. We found that important nodes tend to be on or around high degree nodes. According to the measurement device configuration plan, this paper arranges the node voltage amplitude and phase angle in the dataset in the order of node numbers 3, 6, 9, 12, 14, 17, 21, 24, 29, and 31 as inputs to train the TI model. The parameters of the CNN are trained using the Adam optimizer and the initial learning rate is set to 0.001. The 5-fold cross-validation method is used to divide the dataset. This method enables hierarchical random folding of multi-label dataset and guarantees the percentage of labels in each fold. During training, the dataset is randomly divided into 5 equal parts. Each sub-dataset is used as a validation set in turn. The batch size and the number of iterations are set to 32 and 200, respectively.

To verify the efficiency of GAF-CNN, the commonly used machine learning methods SVM, DNN, and CNN are selected as controls. The TI results of the different methods are shown in Table 2.

TABLE 2. TI results of different methods.

Method	Accuracy
SVM	88.48%
DNN	90.23%
CNN	95.67%
GAF-CNN	99.82%

The TI model constructed by GAF-CNN in this paper has more accurate identification results. GAF-CNN improves accuracy by 4.15% compared to CNN. The main reason is that the multi-dimensional data obtained by GAF can effectively represent the characteristics between different topologies, as shown in Figure 7.

After obtaining the TI model, the distribution system topology is used as a known quantity to train the DSSE model. The parameters of the GAT are trained using the Adam optimizer and the initial learning rate is set to 0.001. The dataset is divided into train set, validation set and test set according to the ratio of 8:1:1. The number of attention heads is set to 5. The number of iterations is set to 50.

To test the performance of the DSSE method, six distribution system topologies are selected to compare the DSSE models trained by GAT with DNN, CNN, and GCN.

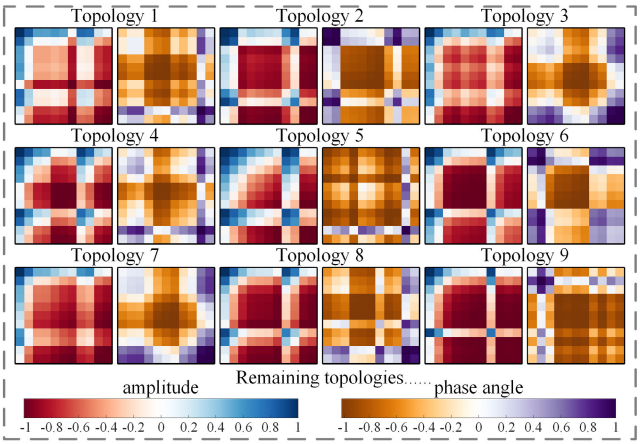


FIGURE 7. Visualization of TI data pre-processed by GAF.

TABLE 3. DSSE results of different topologies.

Method	MAPE of voltage amplitude (%)					
	Topology 1	Topology 2	Topology 3	Topology 4	Topology 5	Topology 6
DNN	0.5518	0.5861	0.6015	0.5741	0.5841	0.6318
CNN	0.5494	0.6153	0.5746	0.5945	0.5256	0.6082
GCN	0.3757	0.3662	0.4297	0.3525	0.3631	0.4061
GAT	0.1122	0.1107	0.1023	0.1248	0.1138	0.1279
Method	MAE of voltage phase angle (10^{-3}rad)					
	Topology 1	Topology 2	Topology 3	Topology 4	Topology 5	Topology 6
DNN	5.5642	5.6818	5.4492	5.3932	5.6607	5.7685
CNN	5.4624	5.1450	5.3556	5.3912	5.6236	5.5171
GCN	3.3627	3.4034	3.5643	3.8144	3.3981	3.4352
GAT	1.0343	0.9862	1.1552	1.0949	1.0166	1.1252

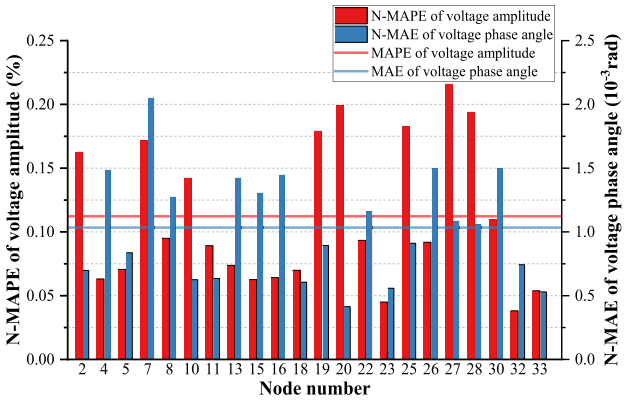
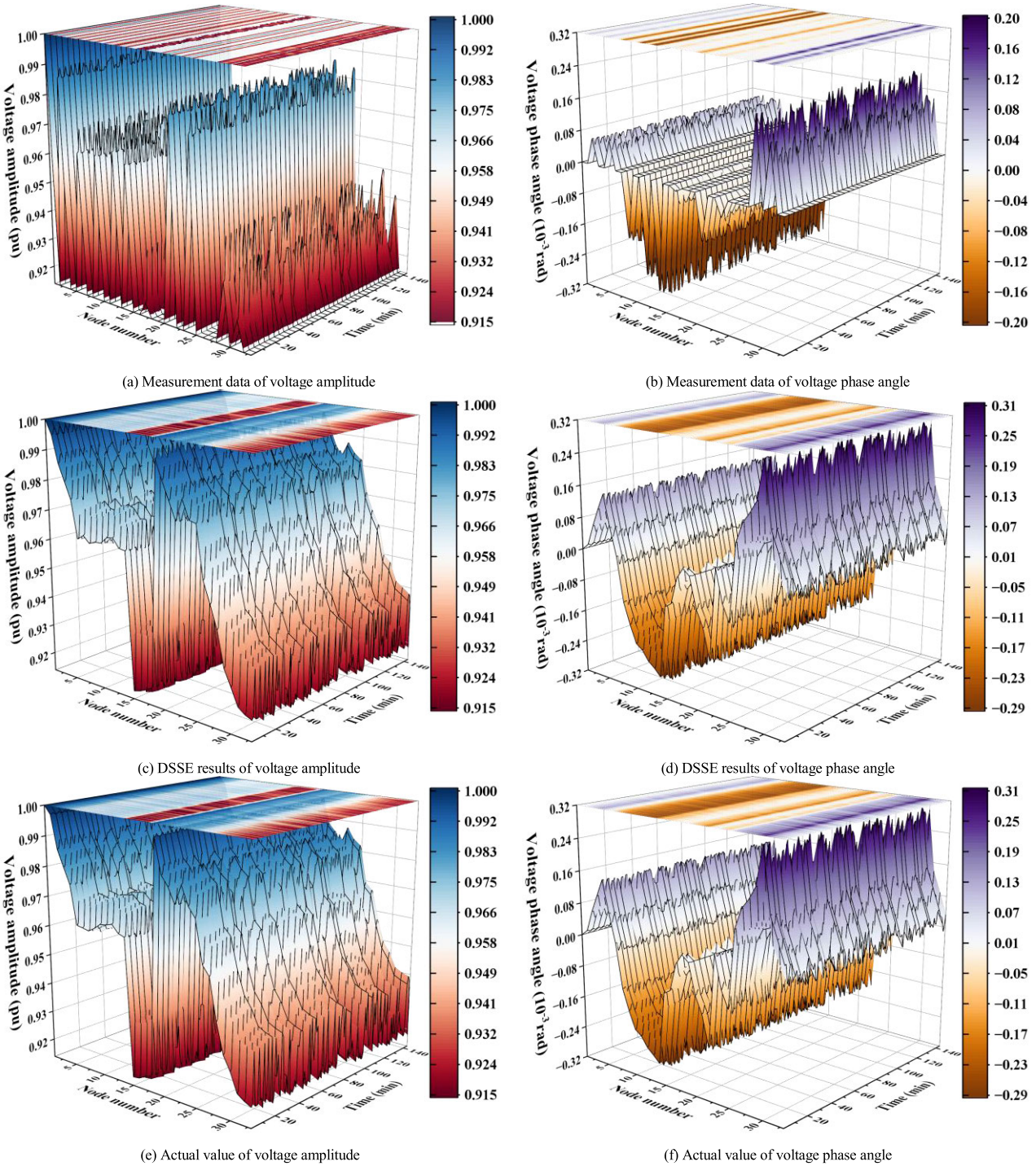


FIGURE 8. Evaluation indexes of DSSE results in topology 1.

DNN and CNN only learn the mapping relationship between the measurement data and the state variables without considering the influence of the topology. Compared to the method in this paper, the error of the DSSE model trained with DNN and CNN is larger. GAT enables the aggregation of information across spatial domains through the attention mechanism. Compared to GCN, GAT solves the problem that only same weights can be assigned to same order neighbor nodes. Meanwhile, GAT is able to cope with topology changes by adaptively adjusting the attention coefficient between nodes. Therefore, the DSSE model trained using GAT is able to recover the state variables of each node more accurately.



(a) Measurement data of voltage amplitude

(b) Measurement data of voltage phase angle

(c) DSSE results of voltage amplitude

(d) DSSE results of voltage phase angle

(e) Actual value of voltage amplitude

(f) Actual value of voltage phase angle

FIGURE 9. DSSE results of the voltage amplitude and phase angle of all nodes.

Topology 1 in Table 3, which includes both radial and looped networks, is selected for detailed analysis of the DSSE method. In topology 1, section switch 14-15 is open and interconnection switches 18-33, 12-22 are closed. The state estimation error of the DSSE is shown in Figure 8.

The worst N-AMPE of voltage amplitude and N-MAE of voltage phase angle in the DSSE results are only 0.2161% and 2.0525×10^{-3} rad, which are within twice of the average values. The state estimation errors of each node are small and similar, proving the effectiveness of the key node selection algorithm in this paper. We selected some of the time points in topology 1 as the validation set. The DSSE results of all nodes are shown in Figure 9.

Figures (a) and (b) show the real-time measurement data obtained by the PMU. Due to the limitation of measurement devices, the input of the DSSE model is very sparse. However,

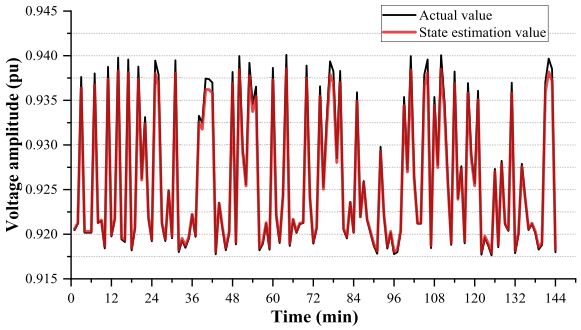


FIGURE 10. DSSE results of voltage amplitude of node 33.

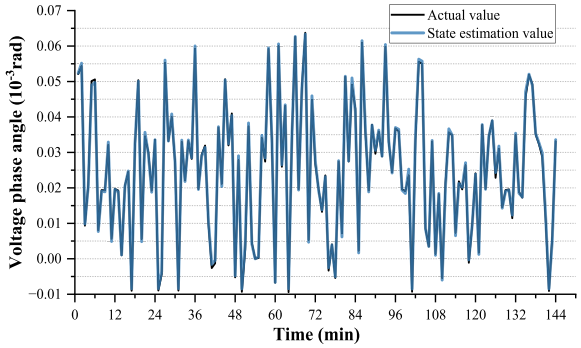


FIGURE 11. DSSE results of voltage phase angle of node 33.

GAT is able to accurately estimate the details of the state variables. The DSSE results in Figures (c) and (d) are generally consistent with the actual values in Figures (e) and (f). The framework in this paper has the advantage of high precision. By replacing the adjacency matrix inputs, we obtain the same excellent results on other topologies in the topology library. The method is of high practical value. Figures 10 and 11 show the DSSE results for 144 time points of node 33 in topology 1.

The DSSE results of the GAT method follow the same trend as the actual values of voltage amplitude and phase angle on the time series. The following section verifies the applicability of the method proposed in this paper in practical scenarios. Section IV-B is tested on topology 1. For a more intuitive representation of the performance of the DSSE model, Sections IV-C-F use the average of the evaluation indexes of the topologies in the dataset.

B. INFLUENCE OF MEASUREMENT DEVICE

This Section demonstrates the superiority of our method in addressing the low observable DSSE problem by configuring varying quantities of measurement devices. According to key nodes selection algorithm, we design five PMU configuration plans. The proportions of directly observable nodes in the distribution system are 18.75%, 25.00%, 31.25%, 37.50%, and 43.75% respectively. The dataset, constructed using the same method and size as in Section IV-A, is numbered from 1 to 5. The test results are shown in Table 4.

As the observability decreases, there is no serious decline in the evaluation indexes of the DSSE model. In Table 4, dataset 6 is constructed from 10 randomly selected PMU nodes. The objective is to analyze the effect of the configuration location of the measurement device on the state estimation accuracy. The state estimation error of dataset 6 is shown in Figure 12.

TABLE 4. Results of different measurement device configuration plans.

Dataset number	Location	MAPE (%)	MAE (10^{-3}rad)	TI (%)
1	6,9,14,17,21,29	0.3494	3.0101	95.46%
2	6,9,12,14,17,21,29,31	0.1950	1.8984	98.37%
3	3,6,9,12,14,17,21,24,29,31	0.1122	1.0343	99.82%
4	3,6,9,12,14,17,21,24,27,29,31,33	0.1067	1.0259	99.86%
5	3,6,8,9,12,14,17,19,21,24,27,29,31,33	0.0979	0.9843	99.91%
6	3,5,8,11,15,21,26,28,29,31	0.1809	1.5756	99.78%

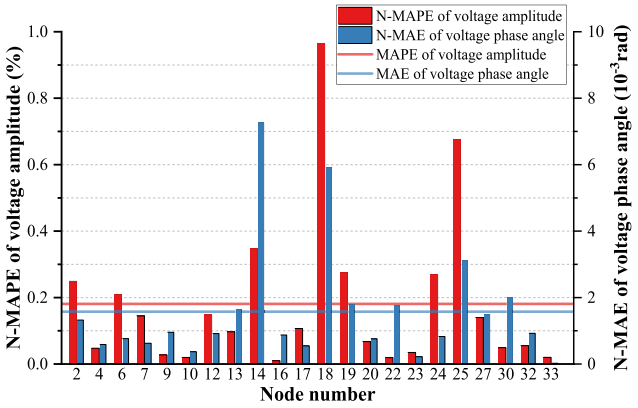


FIGURE 12. Evaluation indexes of DSSE results of dataset 6 in topology 1.

The state estimation errors of nodes 14, 18, and 25 are significantly higher than the average level. This occurs because these nodes only have PMUs configured on 3-order neighbors. Such nodes contribute the vast majority of the error. The measurement device configuration algorithm in this paper effectively circumvents the impact on state estimation accuracy due to improper location selection.

C. INFLUENCE OF NOISE

In the real operating environment, measurement error is inevitable. Factors such as sensor accuracy, measurement algorithms, and hardware aging influence the measurement error of the PMU. This Section evaluates the adaptability of the proposed method to noise by introducing different levels of noise into the measurement data. As shown in Table 5, five datasets with different Gaussian noise levels were constructed.

TABLE 5. Datasets with different noise levels.

Dataset number	Error range of voltage amplitude	Error range of voltage phase angle	Sample size
1	±0.025%	±0.005°	288,000
2	±0.050%	±0.010°	288,000
3	±0.100%	±0.020°	288,000
4	±0.200%	±0.040°	288,000
5	±0.400%	±0.080°	288,000

The dataset number, as constructed in Section IV-A, is set to 0. As shown in Figure 13, both the accuracy of TI and the DSSE worsen as the noise level increases.

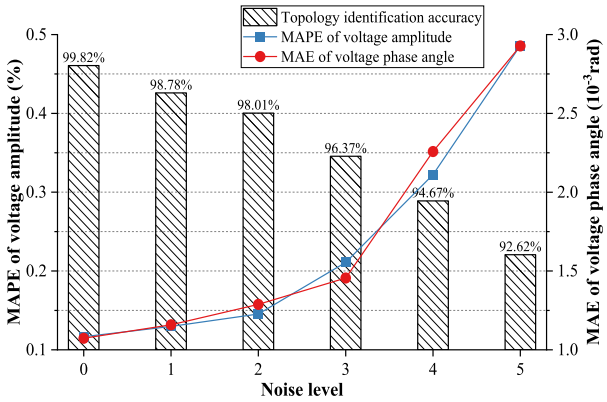


FIGURE 13. DSSE results at different noise levels.

Despite a ±0.2% voltage amplitude error and ±0.04° voltage phase angle error, the TI model still maintains an accuracy over 94%. The error of the DSSE model remains low level at noise levels 1-3. Currently, the measurement accuracy of the PMU is able to keep the error of voltage amplitude and phase angle within ±0.05% and ±0.01° [49]. Therefore, the proposed method can meet the requirements of noise immunity in practical state estimation.

D. INFLUENCE OF DATA MISSING

Due to factors such as communication interference and hardware failure, the PMU measurement data has a certain degree of data missing problem. PMU data missing directly impairs real-time observability, reducing the reliability of DSSE results. Moreover, TI models trained using GAF-CNN might fail to provide high accuracy topology data for state estimation due to missing graph data. In order to evaluate the adaptability of the proposed method to data missing, this paper constructs five datasets with varying percentages of data missing. The percentages are 4%, 8%, 12%, 16%, and 20%, respectively. The dataset construction method and size are identical to V.A. The test results are shown in Figure 14.

The GAF-CNN method shows strong robustness in the data missing test. Even with the highest 20% of data missing, the TI accuracy still reaches 96.59%. Prior to initiating the DSSE for the dataset with data missing, the data missing points are filled using the measurement data from the previous time point. Only a simple preprocessing of the dataset is sufficient to ensure the accuracy of the GAT method. With the highest percentage of missing data, the evaluation indexes of voltage amplitude and phase angle only increased by 0.0739% and 0.8620×10^{-3} rad compared to the original state.

E. ANOTHER DISTRIBUTION SYSTEM TEST

To verify the impact of system size on the accuracy and efficiency of the proposed data-driven method, this section uses an IEEE 118-bus distribution system as a test case. The same method as in Section IV-A is used to construct a dataset

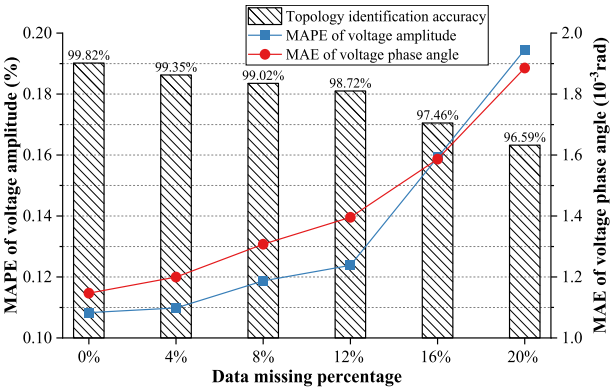


FIGURE 14. DSSE results with different percentage of data missing.

of 390 topologies. First, PMUs are configured based on the node importance and the number is 35. Then, the average of MAPE and MAE for all topologies in the IEEE 118-bus topology library are used to evaluate the DSSE model. The test results are shown in Table 6.

TABLE 6. Comparison of TI and DSSE results of two distribution systems.

Test results	IEEE 33-bus	IEEE 118-bus
Average online TI time (10^{-3} s)	36.4796	38.8701
Average online DSSE time (10^{-3} s)	1.9889	2.0269
Topology identification accuracy (%)	99.82	99.57
MAPE of voltage amplitude (%)	0.1147	0.1320
MAE of voltage phase angle (10^{-3}rad)	1.0832	1.2378

As the distribution system size increases, the accuracy of the TI model slightly diminishes. Since GAT focuses on the electrical characteristics of neighbor nodes, the performance of the DSSE model is less affected by the size of the distribution system within a certain range. The DSSE framework in this paper only requires the state information of partial nodes. Therefore, neither the training difficulty nor the online running time increased significantly in the IEEE 118-bus distribution system. The method proposed in this paper is suited to online analysis scenarios of distribution systems. However, node degree does not change significantly with increasing distribution system size. Foreseeably, large and complex distribution systems have higher sparsity. During the aggregation process, the accuracy of some nodes decreases due to the large distance from the observable nodes. TI will face the pressure of too large topology library and the difficulty of updating them. Therefore, large systems require decoupling and partitioning before further expansion. The framework is then applied in parallel to each sub-region.

F. TOPOLOGY GENERALIZATION TEST

When a new topology appears, the TI model is able to be updated with less new data. Compared to the TI model, the regression prediction problem handled by the DSSE model is more demanding on the dataset. During the actual operation of the distribution system, the dataset often lacks sufficient historical samples of new topologies to drive model training.

Therefore, this paper uses the existing IEEE 118-bus DSSE model for generalization testing of six new reconfigured topologies. The test results are shown in Figure 15.

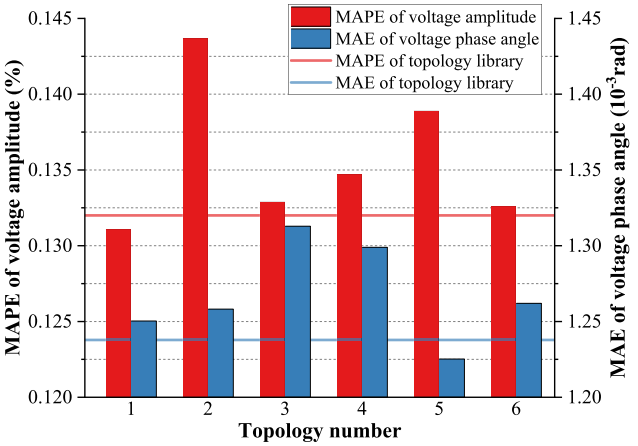


FIGURE 15. Topology generalization test results of IEEE 118-bus system.

Even without learning new topologies, the DSSE model trained with GAT still maintains high accuracy. The proposed method can integrate the spatial correlation of topology and node state information to fully learn the mapping relationship between the measurement data and state variables. Therefore, the model possesses good generalization capability and is suitable for frequent topology reconfiguration.

V. CONCLUSION

In this paper, we propose a DSSE framework that can adapt to frequent topology changes. To meet the data requirements, the configuration plan of the measurement device is designed based on the node importance results calculated by SCC and XGBoost. To capture the real-time topology of the distribution system, a GAF-CNN method using graph data is proposed for TI. Finally, GAT is trained to perform online DSSE. Experimental results show that this framework has good robustness in case of data quality degradation. Even with decreasing observability, the framework sustains high performance. Furthermore, the study tests the scalability of the framework by deploying it on an IEEE 118-bus distribution system. In addition, topology generalization experiments demonstrate that the DSSE model can effectively handle complex operation state.

While the proposed framework enables accurate DSSE under frequent reconfiguration, we need further research in several important directions. First, we are researching how to introduce physical constraints into the GAT training process. As a possible extension, power flow constraints can be added to the loss function to make the DSSE results more in line with the electrical laws and help the model jump out of the local optimal solution. Second, distributed generation is monitored on different time scales. These measurements can add redundancy. We could incorporate the measured and predicted values into the model training to further improve the accuracy. Finally, scaling to large systems will be an important research direction. We need to study parallel DSSE methods including decoupling and partitioning based on the proposed framework.

REFERENCES

[1] A. Abdolahi and N. T. Kalantari, "State estimation of asymmetrical distribution networks by μ-PMU allocation: A novel stochastic two-stage programming," *Electric Power Syst. Res.*, vol. 213, Dec. 2022, Art. no. 108738.

[2] Z. Soltani and M. Khorsand, "Real-time topology detection and state estimation in distribution systems using micro-PMU and smart meter data," *IEEE Syst. J.*, vol. 16, no. 3, pp. 3554–3565, Sep. 2022.

[3] H. Long, Z. Wu, C. Fang, W. Gu, X. Wei, and H. Zhan, "Cyber-attack detection strategy based on distribution system state estimation," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 4, pp. 669–678, Jul. 2020.

[4] I. Džafic, R. A. Jabr, and T. Hrnjic, "Hybrid state estimation in complex variables," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5288–5296, Sep. 2018.

[5] Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "A joint distribution system state estimation framework via deep actor-critic learning method," *IEEE Trans. Power Syst.*, vol. 38, no. 1, pp. 796–806, Jan. 2023.

[6] X. Zhou, Z. Liu, Y. Guo, C. Zhao, J. Huang, and L. Chen, "Gradient-based multi-area distribution system state estimation," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5325–5338, Nov. 2020.

[7] P. M. De Oliveira-De Jesus, N. A. Rodriguez, D. F. Celeita, and G. A. Ramos, "PMU-based system state estimation for multigrounded distribution systems," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1071–1081, Mar. 2021.

[8] K. Li and X. Han, "A distributed Gauss–Newton method for distribution system state estimation," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107694.

[9] N. Bhusal, R. M. Shukla, M. Gautam, M. Benidris, and S. Sengupta, "Deep ensemble learning-based approach to real-time power system state estimation," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106806.

[10] M. Pau and Z. Tamim, "A backward–forward sweep algorithm for distribution system state estimation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[11] Y. Zhang and J. Wang, "Towards highly efficient state estimation with nonlinear measurements in distribution systems," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2471–2474, May 2020.

[12] Y. Liu, A. S. Zamzam, and A. Bernstein, "Multiarea distribution system state estimation via distributed tensor completion," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4887–4898, Nov. 2022.

[13] T. Zhang, W. Zhang, Q. Zhao, J. Chen, and Y. Zhang, "Adaptive particle filter with randomized quasi-Monte Carlo sampling for unbalanced distribution system state estimation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[14] A. Louis, G. Ledwich, G. Walker, and Y. Mishra, "Measurement sensitivity and estimation error in distribution system state estimation using augmented complex Kalman filter," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 4, pp. 657–668, Jul. 2020.

[15] J. Li, M. Gao, B. Liu, and Y. Cai, "Forecasting aided distribution network state estimation using mixed μPMU-RTU measurements," *IEEE Syst. J.*, vol. 16, no. 4, pp. 6524–6534, Dec. 2022.

[16] Y. Wang, M. Xia, Q. Yang, Y. Song, Q. Chen, and Y. Chen, "Augmented state estimation of line parameters in active power distribution systems with phasor measurement units," *IEEE Trans. Power Del.*, vol. 37, no. 5, pp. 3835–3845, Oct. 2022.

[17] R. Mohammadrezaee, J. Ghaisari, G. Yousefi, and M. Kamali, "Dynamic state estimation of smart distribution grids using compressed measurements," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4535–4542, Sep. 2021.

[18] J. Xu, Z. Wu, T. Zhang, Q. Hu, and Q. Wu, "A secure forecasting-aided state estimation framework for power distribution systems against false data injection attacks," *Appl. Energy*, vol. 328, Dec. 2022, Art. no. 120107.

[19] X. Kong, X. Zhang, X. Zhang, C. Wang, H.-D. Chiang, and P. Li, "Adaptive dynamic state estimation of distribution network based on interacting multiple model," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 643–652, Apr. 2022.

[20] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4796–4805, Nov. 2019.

[21] M. M. Rana, R. Bo, and A. Abdelhadi, "Distributed grid state estimation under cyber attacks using optimal filter and Bayesian approach," *IEEE Syst. J.*, vol. 15, no. 2, pp. 1970–1978, Jun. 2021.
[22] J. Kim, H.-T. Kim, and S. Choi, "Performance criterion of phasor measurement units for distribution system state estimation," *IEEE Access*, vol. 7, pp. 106372–106384, 2019.
[23] J. A. D. Massignan, J. B. A. London, M. Bessani, C. D. Maciel, R. Z. Fannucchi, and V. Miranda, "Bayesian inference approach for information fusion in distribution system state estimation," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 526–540, Jan. 2022.
[24] S. Dahale and B. Natarajan, "Bayesian framework for multi-timescale state estimation in low-observable distribution systems," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4340–4351, Nov. 2022.
[25] F. Ahmad, M. Tariq, and A. Farooq, "A novel ANN-based distribution network state estimator," *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 200–212, May 2019.
[26] B. Zargar, A. Angioni, F. Ponci, and A. Monti, "Multiarea parallel data-driven three-phase distribution system state estimation using synchrophasor measurements," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6186–6202, Sep. 2020.
[27] P. Sundaray and Y. Weng, "Alternative auto-encoder for state estimation in distribution systems with unobservability," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2262–2274, May 2023.
[28] L. Wang, Q. Zhou, and S. Jin, "Physics-guided deep learning for power system state estimation," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 4, pp. 607–615, Jul. 2020.
[29] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep unrolled neural networks," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 4069–4077, Aug. 2019.
[30] B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarathy, "State and topology estimation for unobservable distribution systems using deep neural networks," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 9003514.
[31] G. Tian, Y. Gu, D. Shi, J. Fu, Z. Yu, and Q. Zhou, "Neural-network-based power system state estimation with extended observability," *J. Modern Power Syst. Clean Energy*, vol. 9, no. 5, pp. 1043–1053, Sep. 2021.
[32] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4347–4356, Nov. 2020.
[33] M. Xia, J. Sun, and Q. Chen, "Outlier reconstruction based distribution system state estimation using equivalent model of long short-term memory and metropolis-hastings sampling," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 6, pp. 1625–1636, Nov. 2022.
[34] Z. Gao, S. Hu, H. Sun, J. Liu, Y. Zhi, and J. Li, "Dynamic state estimation of new energy power systems considering multi-level false data identification based on LSTM-CNN," *IEEE Access*, vol. 9, pp. 142411–142424, 2021.
[35] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Graph neural network-based distribution system state estimators," *IEEE Trans. Ind. Informat.*, early access, Feb. 23, 2023, doi: 10.1109/TII.2023.3248082.
[36] H. Wu, Z. Xu, and M. Wang, "Unrolled spatiotemporal graph convolutional network for distribution system state estimation and forecasting," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 297–308, Jan. 2023.
[37] D. Cao, J. Zhao, W. Hu, N. Yu, J. Hu, and Z. Chen, "Physics-informed graphical learning and Bayesian averaging for robust distribution state estimation," *IEEE Trans. Power Syst.*, early access, Jun. 2, 2023, doi: 10.1109/TPWRS.2023.3282413.
[38] O. Kundacina, M. Cosovic, D. Miskovic, and D. Vukobratovic, "Graph neural networks on factor graphs for robust, fast, and scalable linear state estimation with PMUs," *Sustain. Energy Grids Netw*, vol. 34, Jun. 2023, Art. no. 101056.
[39] A. Primadianto and C.-N. Lu, "A review on distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3875–3883, Sep. 2017.
[40] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.
[41] Z. Wang, Y. Chen, S. Huang, X. Zhang, and X. Liu, "Temporal graph super resolution on power distribution network measurements," *IEEE Access*, vol. 9, pp. 70628–70638, 2021.
[42] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Intrusion detection in SCADA based power grids: Recursive feature elimination model with majority vote ensemble algorithm," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2559–2574, Jul. 2021.
[43] N. Li, B. Li, and L. Gao, "Transient stability assessment of power system based on XGBoost and factorization machine," *IEEE Access*, vol. 8, pp. 28403–28414, 2020.
[44] B. Patnaik, M. Mishra, R. C. Bansal, and R. K. Jena, "MODWT-XGBoost based smart energy solution for fault detection and classification in a smart microgrid," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116457.
[45] Y.-Y. Hong, J. J. F. Martinez, and A. C. Fajardo, "Day-ahead solar irradiation forecasting utilizing Gramian angular field and convolutional long short-term memory," *IEEE Access*, vol. 8, pp. 18741–18753, 2020.
[46] Y. Wu, B. Wang, R. Yuan, and J. Watada, "A Gramian angular field-based data-driven approach for multiregion and multisource renewable scenario generation," *Inf. Sci.*, vol. 619, pp. 578–602, Jan. 2023.
[47] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit: Algorithms and Cramér-rao bound," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1649–1658, Mar. 2018.
[48] P. Veličkovič, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
[49] M. Dey, S. P. Rana, C. V. Simmons, and S. Dudley, "Solar farm voltage anomaly detection using high-resolution μPMU data-driven unsupervised machine learning," *Appl. Energy*, vol. 303, Dec. 2021, Art. no. 117656.

ZHENGXING REN (Graduate Student Member, IEEE) received the B.Eng. degree in electrical engineering from Jilin University, Changchun, China, in 2021. He is currently pursuing the master's degree with the School of Electrical Engineering, Shandong University. His research interests include cyber-physical system modeling, machine learning, and state estimation in smart grids.

XIAODONG CHU (Member, IEEE) received the B.Eng. degree in electrical engineering from the Shandong University of Technology, Jinan, China, in 2000, and the Ph.D. degree in electrical engineering from Shandong University, Jinan, in 2006.

She was with Shandong University, as an Assistant Professor, in 2006. She is currently an Associate Professor with the Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University. Her research interests include power system stability and control and AI applications in power systems.

HUA YE (Member, IEEE) received the B.Eng. and Ph.D. degrees in electrical engineering from Shandong University, Jinan, China, in 2003 and 2009, respectively.

He is currently a Professor with the Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University. His research interests include power system dynamic stability analysis and control and cyber-physical systems.