energies

MDPI

Article

Data-Driven State Estimation Method for Distribution Networks Based on Power Flow Constraints

Jinchen Lan *, Yan Lin, Zhigeng Zhang, Shuangting Xu and Xiaoling Fang

State Grid Fujian Electric Power Co., Ltd. Electric Power Research Institute, Fuzhou 350007, China; lyepri@163.com (Y.L.); ornertito570@gmail.com (Z.Z.); sy_uaena@163.com (S.X.); belieffxl@163.com (X.F.)
* Correspondence: lanjc8304@163.com; Tel.: +86-157-6757-4862

Abstract

Addressing uneven accuracy across multiple measurement sources and the insufficient adaptability of traditional data-driven models to topological changes, a data-driven state estimation method for distribution networks based on power-flow constraints is proposed. Firstly, historical prior information is used to supplement missing data, thereby constructing a complete measurement dataset and establishing a reliability index to assess the credibility of measurements from different sources. Secondly, according to the physical topological map structure of the distribution network, an improved graph neural network algorithm is proposed to realize state estimation. The bus power balance relationship is embedded as a soft constraint in the loss function to enhance the consistency between the estimation results and physical laws. Then, an adaptive method of topology change based on transfer learning is proposed to improve the adaptability of the model to the dynamic changes of distribution network topology. Finally, the proposed method is evaluated using various metrics on the IEEE 33 and IEEE 118 systems. The analysis results show that the method can maintain high estimation accuracy under multi-source measurements and exhibits good robustness to noise disturbances.

Keywords: graph neural network; distribution network state estimation; power flow constraints; physical information learning

1. Introduction

With the increasing proportion of distributed energy resources and flexible loads integrated into the distribution network, the traditional unidirectional power transmission mode is gradually evolving into an operational form characterized by multi-source coordination and coexistence of bidirectional power flows [1]. To support operational decisions such as distributed resource consumption and voltage control, accurate and real-time system status is first required [2]. The task of distribution network state estimation is to reconstruct the state of all network buses using current measurements, given that the network topology and equipment parameters are known or identifiable [3].

In state estimation methods, the weighted least squares (WLS) method boasts a relatively mature theoretical foundation and has been widely applied in transmission grids. Compared with transmission grids, distribution grids exhibit greater complexity in terms of measurement configuration, network structure, and operational characteristics. Existing research has improved DSSE by asynchronous measurement processing, online optimization, distributed computing, uncertainty modeling, and joint topological state estimation [4–10].

Check for updates

Academic Editor: Il-Yop Chung

Received: 27 May 2026
Revised: 22 June 2026
Accepted: 27 June 2026
Published: 30 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

These methods improve the applicability of model-based DSSE, but their performance still depends heavily on measurement redundancy, model accuracy, and operating assumptions.

Data-driven approaches provide another solution for DSSE. Deep learning models can learn nonlinear mapping from measurements to states and provide fast online reasoning after training. Existing works has considered the neural networks of noise training, power-flow-informed models, Bayesian neural networks, multi-fidelity learning, and physics-guided estimation under limited measurements [11–16]. These studies have improved the robustness under incomplete or noisy data. However, many methods based on neural networks still express the measured values as vectors or tabular features with fixed length, which limits their adaptability when the feeder topology changes.

Graph neural networks (GNNs) are naturally suited to this problem because they operate directly on graph-structured data. Recent research has introduced hypergraph modeling, graph convolution, graph attention, topology adaptive learning, and physical information graph structures into DSSE [17–24]. These methods show that graph topology and physical knowledge can improve estimation accuracy, interpretability, and cross-scenario generalization. Despite these advances, actual DSSE still faces several related difficulties. Measurements from smart meters, PMUs, RTUs, and pseudo-measurement sources differ in accuracy, sampling rate, and reliability. Meanwhile, the measurement set may be sparse, the feeder topology may change, and estimated states should be consistent with the power flow physics. Existing methods usually solve these problems respectively, such as topological error detection [25–27], bad data and network attack processing [28–30], low observability estimation [31–36], or heterogeneous and asynchronous measurement fusion [37–40]. It is still difficult to establish a unified framework that can jointly deal with measurement uncertainty, topological change, and physical consistency.

Therefore, this paper proposes a graph neural network-based method for state estimation in distribution networks. This approach aims to address the challenges posed by inconsistent accuracy across multi-source measurements and the limited adaptability of traditional data-driven models to topological changes. The proposed method has been validated across multiple scenarios and comprehensively evaluated for estimation accuracy and physical consistency. Experimental results show that the proposed method achieves good estimation accuracy and improved robustness under the considered noise, gross-error, and topology-change scenarios. The specific contributions are as follows:

(1) Design a GNN structure that is oriented towards measurement heterogeneity and sparsity. The model utilizes measurement values, measurement masks, measurement variances, and pseudo-measurement labels at the input simultaneously and employs a reliability gating mechanism to distinguish the credibility of measurements from different sources.
(2) It is proposed to embed the bus power balance relationship as a soft constraint into the GNN training process. By constructing a composite objective function that combines the supervised loss with the power-flow constraint loss, the physical consistency of the results can be improved while ensuring estimation accuracy.
(3) To solve the problem of insufficient generalization ability of the distribution network topological dynamic change model, a training strategy based on transfer fine-tuning is proposed. This strategy enables the model to quickly adapt to a new topology using a small number of target topology samples, after acquiring basic generalization ability on the source topology.

Overall, Table 1 presents a comparison between the proposed IGCN method and representative existing studies, which helps clarify the technical positioning of this work. The remaining chapters of this paper are structured as follows: Section 2 elaborates on the basic theory; Section 3 introduces the graph neural network state estimation method

based on power-flow constraints; Section 4 presents simulation verification; and Section 5 provides a summary of the entire paper.

Table 1. Comparison between the proposed method and related studies.

Ref.	Topology	GNN/GCN	Attention	GRU	Physics Loss	Reliability Gate	TL
[17]	△	✓	–	–	△	△	–
[18]	△	✓	–	✓	–	–	–
[19]	✓	✓	✓	–	–	–	–
[20]	△	✓	✓	–	✓	–	✓
[21]	△	✓	–	–	✓	–	–
[22]	✓	✓	–	–	–	–	–
[23]	△	✓	–	–	✓	–	–
[24]	–	–	–	–	✓	–	–
[25]	✓	–	–	–	–	△	–
[26]	✓	–	–	–	–	△	–
[27]	✓	–	–	–	–	✓	–
[28]	–	–	–	–	–	✓	–
[29]	–	–	–	–	–	✓	–
[30]	–	–	–	–	–	✓	–
[31]	–	–	–	✓	–	–	–
[32]	–	–	–	△	–	–	–
[33]	–	–	–	–	✓	–	–
[34]	–	–	–	–	✓	–	–
[35]	–	–	–	–	–	–	–
[36]	–	–	–	✓	–	–	–
[37]	–	–	–	△	–	–	–
[38]	△	–	–	–	–	–	–
[39]	–	–	–	–	–	–	–
Ours	✓	✓	✓	✓	✓	✓	✓

Symbols: ✓ = explicitly included; △ = partially related or scenario-level relevance; – = not explicitly included.

2. Basic Theory

2.1. Foundation of Distribution Network State Estimation

Given the network topology and parameters, the state estimation of a distribution network can be formulated as a problem of reconstructing the system state vector from the measurement vector. For distribution networks containing multiple types of measurements, the relationship between the measured value and states is usually nonlinear, which can be expressed as follows:

$$z = h(x) + e \tag{1}$$

$$z = [z_{SCADA}^T, z_{PMU}^T, z_{pseudo}^T]^T \tag{2}$$

$$x = [V_1, \theta_1, V_2, \theta_2, \cdots, V_n, \theta_n]^T \tag{3}$$

where z is the measurement vector; $h(x)$ is the nonlinear measurement function determined by network topology, equipment parameters, and system state; e is the measurement error

vector; z_{SCADA} is the SCADA measurement subvector, typically including measurements such as voltage amplitude, bus injection power, and branch power; z_{PMU} is the PMU measurement subvector, typically including voltage phasors and current phasors; z_{pseudo} is the pseudo-measurement subvector generated based on historical statistics or load forecasting; the superscript T denotes transposition; x is the state vector; V_i and θ_i are the voltage amplitude and phase angle of the ith bus, respectively; and n is the number of buses.

When the measurement error follows an approximate Gaussian distribution with zero mean and known covariance, traditional state estimation typically employs the WLS model for solution:

$$\min J(x) = [z - h(x)]^T R^{-1} [z - h(x)] \tag{4}$$

where $J(x)$ is the objective function; R is the measurement error covariance matrix; and R^{-1} is its inverse matrix, which reflects the weights of different measurements.

Distribution networks often exhibit radial or weakly meshed structures, with limited high-precision measurement configurations. The absence of measurements in certain areas can easily lead to reduced observability, thereby reducing estimation accuracy. Simultaneously, frequent changes in load, distributed energy resources, and switch states lead to variations in the distribution of measurements and the network structure, necessitating an estimation model that balances accuracy and adaptability.

2.2. Basic Principles of Graph Neural Networks

For power system problems, graph models can better preserve the true connection relationships between buses, branches, and equipment parameters, making them suitable for describing distribution network state estimation problems. It can be expressed as

$$G = (S, E, X, Y) \tag{5}$$

$$m_i^l = AGG_{\{j \in N(i)\}} \phi^i (h_i^l, h_j^l, y_{ij}) \tag{6}$$

$$h_i^{l+1} = \varphi^l (h_i^l, m_i^l) \tag{7}$$

where G is the graph structure sample; S is the set of nodes; E is the set of edges; X is the node feature matrix; Y is the edge feature matrix; m_i^l is the neighborhood message aggregated by node i in the lth layer; $N(i)$ is the set of neighbors of node i; h_i^l and h_j^l are the hidden representations of node i and node j at the lth layer, respectively; y_{ij} is the feature of edge ij; $\phi^i()$ is a message function; AGG() is an aggregation operator; h_i^{l+1} is the updated representation of node i at layer $l + 1$; $\varphi^l()$ is a node update function, typically consisting of linear mapping, nonlinear activation, and normalization operations.

When the bus measurements are sparse, the model can supplement local information with neighborhood messages; when the topology changes, simply update the graph structure and edge features to continue processing within the same network framework. Therefore, this paper chooses a graph attention propagation structure with edge features to simultaneously utilize bus attributes and line parameters.

3. Graph Neural Network State Estimation Method Based on Power Flow Constraints

3.1. Construction and Feature Representation of the Distribution Network Topology Diagram

To accurately describe the structure of the distribution network and the relationships between buses, this paper maps its physical topology to a graph. Each node corresponds to a bus or equivalent bus, and each edge corresponds to a line or transformer branch. For contact switches and segment switches, they can be expressed by the switch status in the

edge existence state or edge attributes. When the switch is disconnected, the corresponding edge is removed from the graph or its weight is reset to 0, thereby keeping the graph structure consistent with the actual operating topology. For scenarios where the parameters of multiple segmented lines and branches are inconsistent, edges should be built separately based on the actual electrical connection relationship. The expression is as follows:

$$e_{ij} = [r_{ij}, q_{ij}, g_{ij}, b_{ij}, l_{ij}, s_{ij}]^T \tag{8}$$

where e_{ij} is the eigenvector of edge ij; r_{ij} and q_{ij} are the resistance and reactance of the circuit, respectively; g_{ij} and b_{ij} are the conductance and susceptance components of the series admittance, respectively; l_{ij} is the line length or equivalent length parameter; s_{ij} is in the on/off state, set to 1 when closed and 0 when disconnected.

Considering the availability and credibility of measurement information, this paper uses measurement values, measurement masks, measurement variances, and pseudo-measurement markers as node features and concatenates historical estimates or time-window statistics to enhance the representation of time-related information. The expression is as follows:

$$x_i = [V_i, P_i, Q_i, m_i^V, m_i^P, m_i^Q, \sigma_{i,V}^2, \sigma_{i,P}^2, \sigma_{i,Q}^2, \delta_i^{ps}]^T \tag{9}$$

where x_i is the input feature vector of bus i; V_i, P_i, and Q_i are the voltage amplitude, active injection, and reactive injection measurements, respectively; m_i^V, m_i^P and m_i^Q are mask variables corresponding to whether the measurement is available; $\sigma_{i,V}^2$, $\sigma_{i,P}^2$ and $\sigma_{i,Q}^2$ are the corresponding measurement variances; δ_i^{ps} is a pseudo-measurement marker, when the node mainly relies on pseudo measurement, take 1, otherwise take 0.

3.2. Data Supplementation and Reliability Gating Mechanism

Due to factors such as faults in power system measurement devices and communication link interruptions, system measurement values often exhibit phenomena such as data loss. Therefore, this paper constructs pseudo-measurements of active power, reactive power, and voltage amplitude using prior center values as the mean and relative noise as the discrete scale. The process of generating pseudo-measurements is equivalent to replacing the unknown state variables of unobservable buses with a set of random prior observations subject to large-variance constraints and then serving the learning process of graph neural networks through unified measurement vectors, covariance structures, and node feature encoding.

$$\begin{cases} P_i^{ps} = \overline{P}_i + b_i^P + \varepsilon_i^P,\ \varepsilon_i^P \sim N(0, \sigma_{P,i}^2), \sigma_{P,i} = \kappa_{ps} \cdot \max(|\overline{P}_i|, 10^{-3}), \overline{P} = \dfrac{1}{T}\sum_{t=1}^{T} P_i(t) \\ Q_i^{ps} = \overline{Q}_i + b_i^Q + \varepsilon_i^Q, \varepsilon_i^Q \sim N(0, \sigma_{Q,i}^2), \sigma_{Q,i} = \kappa_{ps} \cdot \max(|\overline{Q}_i|, 10^{-3}), \overline{Q}_i = \dfrac{1}{T}\sum_{t=1}^{T} Q_i(t) \\ V_i^{ps} = \overline{V}_i + b_i^V + \varepsilon_i^V,\ \varepsilon_i^V \sim N(0, \sigma_{V,i}^2), \sigma_{V,i} = \kappa_{ps} \cdot \max(|\overline{V}_i|, 10^{-3}), \overline{V}_i = \dfrac{1}{T}\sum_{t=1}^{T} V_i(t) \end{cases} \tag{10}$$

where i denotes the index of the bus; P_i^{ps}, Q_i^{ps}, and V_i^{ps} represent the pseudo-measurements of active power, reactive power, and voltage magnitude at bus i, respectively. These pseudo-measurements are used to supplement the input information when direct measurements are unavailable or incomplete. $\overline{P}_i$, $\overline{Q}_i$, and $\overline{V}_i$ denote the prior center values of active power, reactive power, and voltage magnitude at bus i, respectively. In this study, $\overline{P}_i$ and $\overline{Q}_i$ are obtained from the historical load means of each bus. In practical applications, they can be calculated from AMI or SCADA historical time-series data. The terms b_i^P, b_i^Q, and b_i^V are bias terms for active power, reactive power, and voltage magnitude, respectively. They are introduced to describe possible systematic deviations in pseudo-measurements or real measurements. For real measurements, the bias scale is set to (0.0, 0.01), the voltage magnitude and phase-angle biases are randomly generated within $[-0.01, 0.01]$, while

the active and reactive power biases are set within ±1% of the corresponding measured values. For pseudo-measurements, the pseudo-bias scale is set to (0.0, 0.05), which is larger than that of real measurements to reflect the higher uncertainty of prior information. The terms ε_i^P, ε_i^Q, and ε_i^V are random noise terms, which are assumed to follow zero-mean Gaussian distributions. The corresponding standard deviations $\sigma_{P,i}$, $\sigma_{Q,i}$, and $\sigma_{V,i}$ describe the uncertainty levels of the generated pseudo-measurements. For real measurements, the measurement noise is set according to the type and accuracy level of the measurement device. Specifically, the noise parameters of PMU measurements are set to (0.001, 0.0005, 0.02, 0.02), corresponding to voltage magnitude $|V|$, voltage phase angle θ, active power P, and reactive power Q, respectively. For smart meter measurements, the noise parameters are set to (0.005, None, 0.05, 0.05), where None means that the phase angle is not directly measured. For SCADA measurements, the default noise parameters are set to (0.01, None, 0.1, 0.1). The coefficient κ_{ps} is the pseudo-measurement relative noise coefficient, which controls the dispersion degree of the pseudo-measurements around their prior center values. In this study, κ_{ps} is set to 0.3, which means that the standard deviation of each pseudo-measurement is set to 30% of the corresponding mean value. In addition, the lower bound 10^{-3} is introduced to avoid an unrealistically small variance when the historical mean value is close to zero.

To improve the adaptability of the model to multi-source heterogeneous measurement data and enhance the stability of measurement information propagation in complex operating scenarios, this paper constructs a reliability gating mechanism to distinguish the trustworthiness of multi-source measurements. The weight of measurement reliability is defined and integrated into the node feature propagation process so that the model can dynamically adjust the intensity of information aggregation based on measurement quality, thereby avoiding excessive interference from low-quality pseudo-measurements on neighboring nodes during graph propagation. While retaining the dominant role of high-quality measurements in local state recovery and feature expression and improving the robustness and state perception ability of the model in scenarios such as measurement loss and high proportion of pseudo measurements. The reliability weight ω_i can be specifically expressed as

$$\omega_i = \frac{1}{1 + \sqrt{\widetilde{\sigma}_{i,P}^2 + \widetilde{\sigma}_{i,Q}^2} \cdot \eta_i / \xi_i}, \ \widetilde{\sigma}_{i,P} = \frac{\sigma_{i,P}}{S_P}, \ \widetilde{\sigma}_{i,Q} = \frac{\sigma_{i,Q}}{S_Q} \tag{11}$$

where $\widetilde{\sigma}_{i,P}^2$ and $\widetilde{\sigma}_{i,Q}^2$ are the measurement variances of active and reactive power at node i, respectively; S_P and S_Q are the base values of active and reactive power used for normalization. η_i is the penalty factor for pseudo-measurement, and the value of pseudo-measurement nodes is less than 1. In this study, η_i is set to 0.65 for pseudo-measurement nodes and 1.0 for non-pseudo-measurement nodes. ξ_i is a high reliability measurement reward factor; it is set to 1.15 for buses with phasor measurements and 1.0 for other buses.

3.3. Neural Network State Estimation Model

On the basis of the traditional GCN model [41], this paper introduces the attention mechanism, residual connection mechanism, and GRU temporal modeling unit to construct an improved graph neural network model (IGCN) that integrates spatial topological correlation and temporal dynamic features. Firstly, utilizing attention mechanism to adaptively learn the influence weights of different adjacent nodes on the target node, and combining graph convolution operation to extract spatial correlation features between buses; Secondly, the input features are fused with the output of the graph convolution through residual connections to alleviate the problems of feature smoothing and gradient attenuation caused by multi-layer graph convolution; finally, the spatial characteristics of continuous moments are incorporated into the GRU component to further elucidate the dynamic evolution of the

system state over time, thus improving the estimation accuracy of the model. The modules of the proposed IGCN algorithm are as follows:

(1) GCN

Traditional GCNs describe the connection relationships between nodes through adjacency matrices and updates node features based on neighborhood aggregation mechanisms. To avoid the loss of node information during propagation, self-loops are usually added to the adjacency matrix:

$$\widetilde{A} = A + I_N \tag{12}$$

where I_N is an N-order identity matrix. The corresponding degree matrix is

$$\widetilde{D}_{ij} = \sum_j \widetilde{A}_{ij} \tag{13}$$

The single-layer propagation process of standard GCN is as follows:

$$H_t^{l+1} = \sigma(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H_t^l W^l) \tag{14}$$

where H_t^{l+1} represents the node feature matrix of the lth layer at the tth time; $\sigma(\cdot)$ is a nonlinear activation function; W^l is the trainable weight matrix for the lth layer.

(2) Attention mechanism

To improve the model's capacity to depict the roles of various neighboring nodes, the attention mechanism is introduced into the GCN, which enables the model to learn the importance of neighbor nodes adaptively according to the characteristics of nodes. Unlike traditional GCNs that use fixed adjacency matrices for feature aggregation, attention mechanisms can assign different weights to neighboring nodes [42], highlighting the impact of key nodes on the state changes of target nodes and improving the model's ability to represent complex graph structures.

For the target node i, the set of its neighboring nodes is denoted by N_i. At the tth time, the features of node i and node j are $h_{i,t}$ and $h_{j,t}$, respectively. Firstly, the node features are mapped to a high-dimensional space through a shared linear transformation:

$$\begin{cases} z_{i,t} = W_a h_{i,t} \\ z_{j,t} = W_a h_{j,t} \end{cases} \tag{15}$$

where W_a is the trainable weight matrix. Then calculate the attention relevance score of node j to node i:

$$e_{ij,t} = \text{LeakyReLU}\left(a^T[z_{i,t} \parallel z_{j,t}]\right) \tag{16}$$

where $\parallel$ represents the vector concatenation operation, and a is a trainable attention vector. To facilitate comparison between different neighboring nodes, the attention scores are normalized:

$$\alpha_{ij,t} = \frac{\exp(e_{ij,t})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik,t})} \tag{17}$$

where $\alpha_{ij,t}$ represents the attention weight of node j to node i at the tth time, which can dynamically adjust the importance of neighboring nodes based on node characteristics. Therefore, the graph convolution update formula after introducing the attention mechanism is

$$H_t^{l+1} = \sigma\left(A_t^{att} H_t^l W^l\right) \tag{18}$$

where A_t^{att} is a dynamic attention adjacency matrix composed of attention weights $\alpha_{ij,t}$.

(3) Residual Connection Mechanism

As the number of convolutional layers in the graph increases, node features tend to become more similar after multiple neighborhood aggregations, leading to over-smoothing [43]. Meanwhile, there may also be gradient attenuation issues during the training process. Therefore, this paper introduces a residual connection mechanism after the attention graph convolutional layer, enabling the model to learn higher-order neighborhood features while preserving the original node information, thereby reducing the risk of deep feature degradation and improving the model's stability and expressiveness.

If the input feature dimension is the same as the output feature dimension, the residual connection can be expressed as

$$\widetilde{H}_t = \sigma\left(H_t^{att} + \mathrm{X}_t\right) \tag{19}$$

where $\widetilde{H}_t$ represents the spatial feature after fusing residual connections at the tth time; H_t^{att} is the output of the attention graph convolutional layer.

(4) GRU time feature modeling

Attention graph convolutional layers and residual connections are primarily used to extract spatial topological features at each time step. However, in time-series prediction tasks, node states are affected not only by the current topology but also by historical states. Therefore, this paper introduces GRU units to dynamically model the spatial features of continuous time. GRU selectively preserves and forgets historical state information by updating and resetting gates, which can capture dynamic dependencies in time series with lower parameter complexity.

GRU mainly includes update gates, reset gates, and candidate hidden states [44]. When the spatial feature sequence $\left\{\widetilde{H}_{t-T+1}, \widetilde{H}_{t-T+2}, \ldots, \widetilde{H}_t\right\}$ of T consecutive time points is input into GRU, their respective states are as follows:

$$z_t = \sigma\left(W_z\widetilde{H}_t + U_z S_{t-1} + b_z\right) \tag{20}$$

$$r_t = \sigma\left(W_r\widetilde{H}_t + U_r S_{t-1} + b_r\right) \tag{21}$$

$$\widetilde{S}_t = \tanh\left(W_h\widetilde{H}_t + U_h(r_t \odot S_{t-1}) + b_h\right) \tag{22}$$

$$S_t = (1 - z_t) \odot S_{t-1} + z_t \odot \widetilde{S}_t \tag{23}$$

where z_t is the update gate, mainly used to control the degree of preservation of historical state information; r_t is a reset gate, mainly used to control the influence of the hidden state from the previous moment on the current candidate state; $\widetilde{S}_t$ is a candidate hidden state; S_t is the hidden state of GRU at the tth moment; $\odot$ represents Hadamard element wise product; W_z, W_r, W_h, U_z, U_r, U_h are trainable weight matrices; b_z, b_r, and b_h are bias terms.

(5) Output layer design

The hidden state S_t output by the GRU contains historical temporal–spatial features and temporal dynamic information, and the prediction result is obtained through a fully connected layer. To avoid the problem of regression discontinuity near $\pm\pi$, this paper does not directly output the phase angle, but instead outputs the sine and cosine components of the phase angle.

$$y_i = \left[\hat{V}_i, \sin(\hat{\theta}_i), \cos(\hat{\theta}_i)\right]^T \tag{24}$$

where y_i is the model output of node i; $\hat{V}_i$ is the estimate voltage amplitude; $\sin(\hat{\theta}_i)$ and $\cos(\hat{\theta}_i)$ are the sine and cosine representations of the estimated phase angle, respectively.

In summary, this paper has made three improvements based on the standard GCN. Firstly, regarding measurement heterogeneity, the model no longer inputs a unified numerical vector, but instead encodes the measurement values, measurement masks, measurement variances, and pseudo-measurement labels together, reflecting differences between measurement types from the input layer. Secondly, the attention mechanism is integrated into the graph convolutional layer to adaptively calculate the influence weights of neighboring nodes on the target node, based on node features. This enhancement serves to improve the model's capacity to capture intricate spatial relationships. Thirdly, the spatial features at multiple consecutive time points are fed into the GRU unit, and the historical state information is selectively remembered and forgotten through the update and reset gates, thereby capturing the dynamic evolution of node states over time.

3.4. Loss Function

To find the right balance between estimation accuracy and physical consistency, this paper introduces a combined objective function that joins supervised loss with power-flow constraint loss, as shown below:

$$L = L_{\text{sup}} + \tau L_{\text{phy}} \tag{25}$$

where L_{sup} is the supervised loss used to constrain the model output to approach the true state; τ is the physical constraint weights; L_{phy} is the power flow constraint loss used to suppress estimation results that violate power balance laws; L is a composite loss function that retains the ability of supervised learning to directly approximate the true state, while also physically correcting the output results through power flow constraints. The specific expressions of the supervised loss function and the physical loss function are as follows:

$$L_{\text{sup}} = \frac{1}{N}\sum_{i=1}^{N}\{(\hat{V}_i - V_i)^2 + \lambda_\theta[(\sin(\hat{\theta}_i) - \sin(\theta_i))^2 + (\cos(\hat{\theta}_i) - \cos(\theta_i))^2]\} \tag{26}$$

$$L_{\text{phy}} = \frac{1}{|\Omega_v|}\sum_{i\in\Omega_v}\omega_i \cdot (|\Delta P_i| + |\Delta Q_i|) + \frac{\lambda_b}{|\varepsilon|}\sum_{i,j\in\varepsilon}\left\|\Delta S_{ij}\right\|_2^2 \tag{27}$$

$$P_i = V_i\sum_{j=1}^{N} V_j[G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j)] \tag{28}$$

$$Q_i = V_i\sum_{j=1}^{N} V_j[G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j)] \tag{29}$$

$$\begin{cases} \Delta P_i = (\hat{P}_i - P_i^*) \\ \Delta Q_i = (\hat{Q}_i - Q_i^*) \end{cases} \tag{30}$$

where N is the total number of nodes; $\hat{V}_i$ and V_i are the estimated voltage amplitude and true voltage amplitude of node i, respectively; $\hat{\theta}_i$ and θ_i are the estimated phase angle and true phase angle of node i, respectively; the angle error weighting coefficient is λ_θ; Ω_v is the set of effective nodes involved in physical constraint calculations; λ_b is the residual weight of the branch; ΔS_{ij} is the residual power flow of branch ij. Considering that the node power balance term is the main factor, λ_b can be set to 0, and the branch term is retained as an extensible form. Pi and Qi are the active injection and reactive injection of node i calculated from the node voltage, respectively; G_{ij} and B_{ij} are the conductance and electronegativity components of the nodal admittance matrix elements, respectively; ΔP_i and ΔQ_i represent the imbalance between active and reactive power at node i, respectively; $\hat{P}_i$ and $\hat{Q}_i$ are the injection powers obtained by substituting the estimated voltage into the

power flow equation. P_i^* and Q_i^* are the target injection powers in the training samples, and the power-flow truth value can be prioritized for injection; when there is no truth value, it is replaced by the measured value weighted by reliability.

It should be noted that the physical loss term in Equation (27) is used only during the training stage as a physics-based regularization term. Its purpose is to constrain the estimated voltage magnitudes and phase angles to satisfy the nodal power-balance relationship, rather than to provide additional input information to the model. This constraint is especially useful for phase-angle estimation, because some measurement devices, such as SCADA, usually provide active power, reactive power, and voltage magnitude measurements, but do not directly provide voltage phase-angle measurements. Therefore, the power-balance residual can indirectly constrain the phase-angle estimation through the coupling relationship between voltage magnitude, phase angle, and power injection. Meanwhile, in the data generation process, measurement values are produced by adding device-dependent errors to the power-flow baseline values. The measurement values used by the model are not equal to the noise-free power-flow truth. For nodes without direct injection measurements, reliability-weighted pseudo-measurements are used as weak physical constraints. During the testing stage, the physical loss is not calculated, and the trained model only uses the available measurement inputs to estimate the system states. Therefore, Equation (27) does not introduce label leakage in either training or testing.

3.5. Topology Change Adaptive Strategy Based on Transfer Learning

To enhance the model's adaptability to dynamic changes in the distribution network topology, this paper proposes an adaptive strategy for topology modification based on transfer learning. When the distribution network's structure changes due to contact switch operations, fault reconstruction, or tap adjustments, traditional models often require retraining, leading to long training cycles, large sample requirements, and high computational costs. Therefore, this paper leverages transfer learning to inherit existing model parameters and feature representations, thereby enabling the model to adapt quickly to new topological scenarios.

Specifically, the model is first pre-trained on multiple representative operating topologies to learn the mapping between measurement data and system states across different operating conditions. When the system topology changes, the model is not fully retrained. Instead, a limited number of measurement samples under the target topology are employed to refine the existing model, allowing it to quickly adapt to the new network structure while retaining its original knowledge. Meanwhile, given that the sensitivity of different layer parameters to changes in graph structure varies, different parameter update methods can be selected based on actual needs during the fine-tuning process. When the amplitude of topology changes is small, only the parameters of the backend propagation layer and the output layer are adjusted, while keeping the parameters of the frontend encoding layer unchanged, thereby reducing training complexity and the size of parameter updates. When the topology changes significantly, further joint optimization of some encoding layer parameters can be performed to enhance the model's ability to express new topological features. Furthermore, to enhance the model's capacity for generalization to complex topological alterations, multiple samples exhibiting mixed topological features are used collectively during the pre-training phase, enabling the model to access a broader range of network-structure features and reducing its dependence on a single topological pattern. In this way, the model is able to maintain good feature extraction and state mapping capabilities when facing unknown or partially unseen topological structures.

3.6. Performance Indicators

To comprehensively evaluate the performance of the model, this paper sets indicators from two aspects: estimation accuracy and physical consistency, which can be specifically expressed as

$$RMSE_V = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (\hat{V}_i - V_i)^2} \tag{31}$$

$$MAE_V = \frac{1}{N} \cdot \sum_{i=1}^{N} |\hat{V}_i - V_i| \tag{32}$$

$$RMSE_\theta = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2} \tag{33}$$

$$MAE_\theta = \frac{1}{N} \cdot \sum_{i=1}^{N} |\hat{\theta}_i - \theta_i| \tag{34}$$

$$\varepsilon_{phy} = \frac{1}{N} \sum_{i=1}^{N} (|\Delta P_i| + |\Delta Q_i|) \tag{35}$$

where $RMSE_V$ is the root mean square error of voltage amplitude; MAE_V is the average absolute error of voltage amplitude; $RMSE_\theta$ is the root mean square error of the phase angle; MAE_θ is the average absolute error of phase angle; ε_{phy} is the average power imbalance, used to measure the degree of consistency between the estimated results and the power flow constraints.

4. Simulation Analysis

4.1. Experimental Setup

This article uses the IEEE 33 bus system as a test case. The rated voltage of the system is 12.66 kV, and bus 1 is a balanced node. The total load of the system is approximately 3.715 MW + j2.300 Mvar. Trend computing is implemented based on Pandapower, while the graph neural network model is constructed using the PyTorch 2.11 framework.

The measurement configuration nodes of the system is divided into three categories: the first category is PMU buses, which can provide high-precision voltage amplitude and phase angle measurements, and can selectively provide bus injection of active and reactive power; the second type is SM/SCADA buses, which only provide voltage amplitude and injection of active and reactive power measurements, without providing phase angle information; the third type is the pseudo-measurement bus, which does not rely on real-time measurement devices, but uses historical load data to generate active and reactive pseudo-measurements. Its measurement variance is large, and it also does not provide phase angles. The specific number and location of various measurement nodes are shown in Figure 1. Buses not labeled as PMUs or SMs are considered as pseudo-measurement buses. All measurement values are generated by superimposing zero-mean Gaussian white noise on the basis of the true value of the power flow. The noise standard deviations for different types of measurements are detailed in Table 2. To construct a dataset, 2000 operating sections were generated using power-flow calculations under various load-fluctuation scenarios. Each section fully recorded the voltage amplitude, phase angle, and observation values for all nodes under the above measurement configurations. Finally, the dataset was divided into training, validation, and test sets in a 7:1:2 ratio, providing a data foundation for subsequent neural network training and testing. It is worth noting that the influence of gross errors was not specifically considered in the experiment, as the measurement data

were cleaned before training, and gross errors occur less frequently in actual operation, which has a limited impact on the overall conclusion.

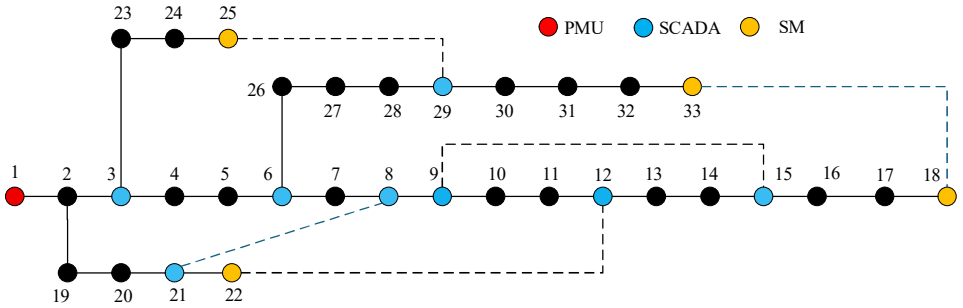


Figure 1. Schematic diagram of the 33 bus system and its measurement configuration.

Table 2. The error setting parameters for various measuring devices

Measurement Type	Voltage Amplitude Error	Active Power Error	Reactive Power Error	Phase Angle Error
PMU	0.001 p.u	0.02	0.02	0.0005 rad
SM	0.005	0.05	0.05	/
SCADA	0.01	0.1	0.1	/
Pseudo measurement	0.3	0.3	0.3	/

Finally, this experiment was programmed and implemented in MATLAB 2022b and Python 3.9, with a hardware platform configuration of Intel Core i7-1170 CPU @ 2.5 GHz, 32 GB of memory, and NVIDIA GeForce GTX 1660 graphics card.

4.2. Validation of Method Effectiveness

After completing the experimental setup, this section first validates the state estimation performance of the IGCN algorithm proposed in this paper. Four comparative methods are set up: Method 1: weighted least squares, a typical representative of mechanistic methods; Method 2: GCN without power flow constraints, which is a standard GCN trained only with supervised loss; Method 3: GCN with power flow constraints, adding the power-flow residual constraint term on the basis of supervised loss; Method 4: the IGCN algorithm proposed in this paper incorporates power-flow constraints.

All GCN models use the same graph structure for input and training hyperparameters. The setting of hyperparameters adopts Bayesian optimization and Hyperband hyperparameter optimization strategies [45] to select the optimal combination of hyperparameters. The specific training hyperparameters are as follows: the hidden layer dimension of the GNN model is set to 64, and Dropout with a rate of 0.2 is applied after each layer. The model is trained using the Adam optimizer, with an initial learning rate of 0.001, a weight decay coefficient of 0.00001, a batch size of 32, and 500 training epochs. To prevent gradient explosion, gradient norm clipping is used with a threshold set to 5. The physical information regularization term is introduced into the loss function, with a weight α of 0.1, and a linear preheating strategy is gradually increased to 0.4. Meanwhile, an early-stop strategy is adopted as a training control mechanism. During training, the validation loss is monitored after each epoch. If the validation loss does not decrease for 80 consecutive epochs, the training process is terminated to prevent overfitting and reduce unnecessary computation. Moreover, the learning rate adopts an adaptive decay strategy. When there is no improvement in the validation loss for 20 consecutive rounds, the learning rate is multiplied by a decay factor of 0.5, and the lower limit of the learning rate is set to 0.00001.

The random mask with a probability of 0.5 is applied to the input node features to enhance the robustness.

After the model training is completed, the estimation performance of each algorithm at a random time section is shown in Figure 2. It can be seen that all methods can achieve relatively accurate estimation of bus voltage amplitude. As for the phase angle of the bus voltage, due to the fact that only PMUs can directly provide phase angle measurements, the available phase angle information for method two is extremely limited, resulting in relatively large phase angle estimation errors. In contrast, the IGCN method can effectively utilize network topology and electrical coupling laws by embedding physical constraints on power flow and inferring the phase angle distribution of other nodes, thus compensating for the deficiency of insufficient phase angle measurements and achieving higher estimation accuracy. Table 3 compares the state estimation accuracy of various methods. The values in the tables were enlarged by 100 times for readability, and this was indicated by "10^{-2}" in the table titles, the numerical values in the following tables are presented in a consistent manner. According to Table 3, the WLS method has significantly higher estimation errors due to insufficient measurement redundancy; the various indicators of the IGCN method are all optimal, with the minimum estimation error. This indicates that the IGCN method can more fully explore the topological relationships between nodes in the distribution network, and effectively improve the accuracy of state estimation and model robustness under the synergistic effect of physical constraints of power flow. Figures 3 and 4 further illustrate the estimation errors and physical residuals of the IGCN algorithm at each bus. As shown in the figure, the estimation error of pseudo-measurement buses is relatively large, and the corresponding physical residuals are also large, while the opposite is true for measurement nodes. It indicates that the introduction of physical constraints has, to some extent, suppressed the interference of pseudo-measurements on estimation accuracy.

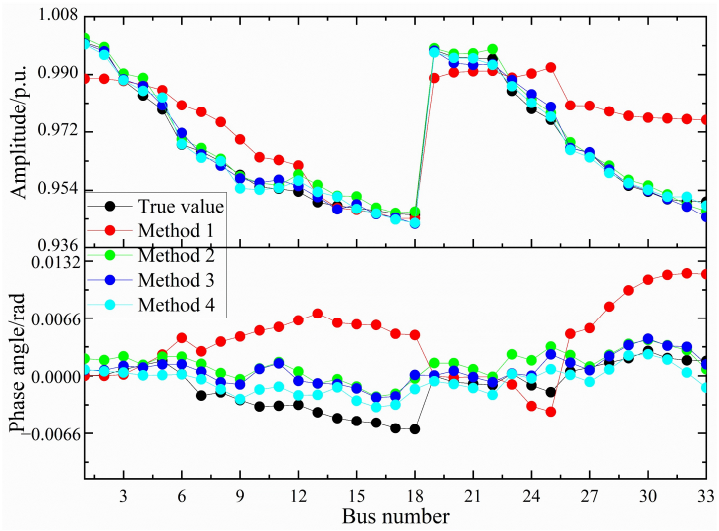


Figure 2. Comparison of the estimation effects of various algorithms.

Table 3. Index values of various algorithms/10^{-2}.

Index	$RMSE_V$/p.u.	MAE_V/p.u.	$RMSE_\theta$/rad	MAE_θ/rad	ε_{phy}/p.u.
Method 1	1.26083	1.00136	0.70083	0.56328	/
Method 2	0.24046	0.19644	0.27209	0.23386	/
Method 3	0.20943	0.16334	0.23352	0.17765	14.8683
Method 4	0.17458	0.13505	0.15947	0.12079	14.8671

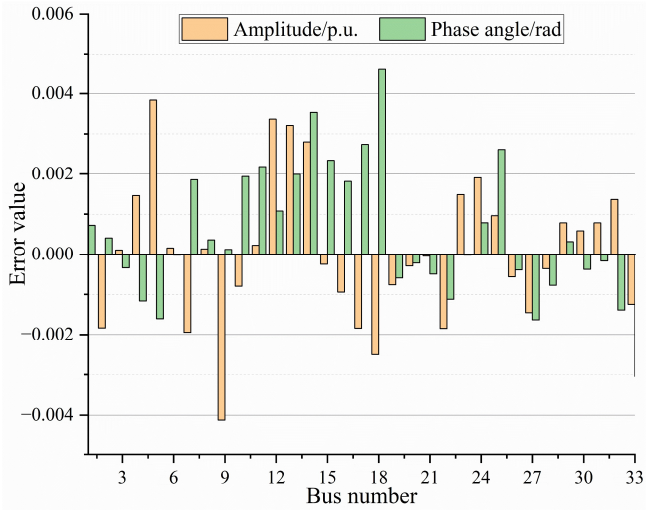


Figure 3. Amplitude and phase angle errors of each node.

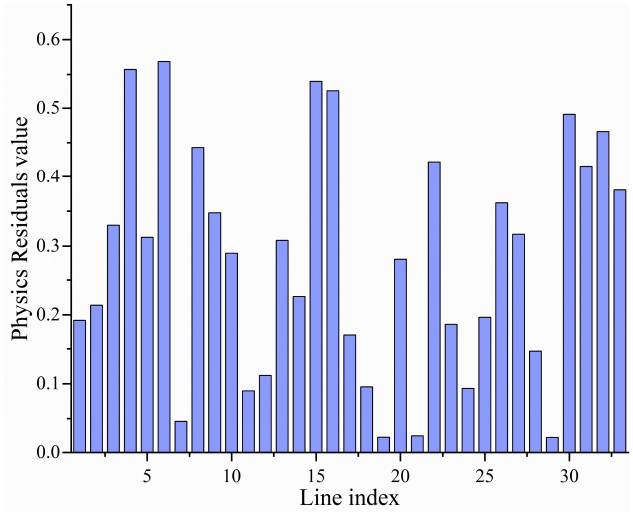


Figure 4. Physical residual value.

To analyze the contribution of each component in the proposed model, an ablation study was added. The IGCN model was used as the baseline, and four variants were constructed by removing the attention mechanism, residual connection, physical constraints, and GRU module, respectively. All variants were trained and tested under the same experimental settings to ensure a fair comparison. The results are shown in Table 4. It can be seen that the full model achieves the best performance in all evaluation metrics, indicating that the different components of the proposed IGCN framework contribute jointly to the final estimation accuracy. When the attention mechanism is removed, both voltage-related and angle-related errors increase, especially for $RMSE_\theta$ and MAE_θ. After removing the residual connection, the voltage estimation error increases significantly. This result shows that the power-flow constraints can guide the estimated states toward the physically feasible region and improve the physical consistency of the estimation results. When the physical constraints are removed from the loss function, all error indicators increase compared with the full model. The removal of the GRU module also leads to a clear increase in the estimation errors, especially in the angle-related metrics. Overall, the ablation results demonstrate that the attention mechanism, residual connection, physical constraints, and GRU module all contribute to the performance of the proposed method.

Table 4. Ablation experiment/10^{-2}.

Variant	$RMSE_V$/p.u.	MAE_V/p.u.	$RMSE_\theta$/rad	MAE_θ/rad
Full model	0.17458	0.13505	0.15947	0.12079
Without attention	0.24046	0.19644	0.27209	0.23386
Without residual	1.89598	1.75465	0.31852	0.23193
Without physical constraints	0.18833	0.14802	0.17204	0.13986
Without GRU	0.20924	0.18334	0.27624	0.21799

4.3. Robustness Analysis Under Measurement Noise and Gross Errors

To evaluate the robustness of the proposed method under abnormal measurement disturbances, the gross-error test was conducted. Three nodes were randomly selected from the measurement nodes, and gross errors with amplitudes of 20–30% were injected into their measurement values. The trained model was then directly applied to the test samples containing gross errors without retraining or parameter adjustment.

The results are shown in Table 5. Compared with the normal measurement-noise scenario, the estimation errors of all methods increase after gross errors are introduced. Nevertheless, the proposed method still maintains relatively stable estimation performance. This can be attributed to the reliability-gating mechanism and the power-flow constraints, which help reduce the influence of abnormal measurements and guide the estimated states toward a physically feasible range. These results indicate that the proposed method has a certain degree of robustness under gross-error conditions.

Table 5. Estimation accuracy of different degrees of gross errors/10^{-2}.

Gross Error Ratio	$RMSE_V$/p.u.	MAE_V/p.u.	$RMSE_\theta$/rad	MAE_θ/rad
0%	0.17458	0.13505	0.15947	0.12079
20%	0.34410	0.22255	0.16127	0.12065
25%	0.35422	0.22700	0.16131	0.12074
30%	0.35514	0.22717	0.16117	0.12075

The measurement noise level in actual distribution networks will fluctuate with equipment status and environmental changes, and measurement errors will directly affect the quality of input data, thereby affecting the state estimation results. To analyze the impact of this factor on model performance, this section sets different measurement error levels relative to the original error, compares the estimation performance of the proposed method under various noise conditions, and assesses its anti-interference ability. It is worth noting that the measurement error here is for all measurement buses.

Table 6 lists the accuracy indicators of state estimation for each algorithm under different levels of measurement noise. Methods 2 to 4 are all data-driven methods, and their estimation performance is highly dependent on the quality of input measurements. As measurement errors continue to increase, various accuracy indicators show a clear upward trend, and estimation accuracy gradually decreases. Within the 5% margin of error, this method outperforms method 2 and method 3 in $RMSE_V$ metrics by 10.76% and 8.39%, respectively. Compared to method 2 and method 3, the MAE_V index has increased by 9.16% and 5.86%, respectively; Compared to method 2 and method 3, the $RMSE_\theta$ index has increased by 40.93% and 3.18%, respectively; on the MAE_θ index, they increased by 50.42% and 9.84%, respectively. Overall, the proposed IGCN method demonstrates superior performance in most evaluation metrics and noise scenarios. In particular, the improvement

is more pronounced for angle estimation indicators, while maintaining competitive voltage estimation accuracy under different measurement noise levels.

Table 6. Estimation accuracy indicators under different measurement errors/10^{-2}.

Index	Methods	Degree of Error		
		3%	5%	10%
$RMSE_V$/p.u.	Method 2	2.17726	2.91737	3.31469
	Method 3	2.07195	2.84213	2.90457
	Method 4	1.92384	2.60356	3.00358
MAE_V/p.u.	Method 2	1.91380	2.54083	2.82536
	Method 3	1.79480	2.45183	2.39924
	Method 4	1.67828	2.30811	2.67365
$RMSE_\theta$/rad	Method 2	0.35393	0.39706	0.42616
	Method 3	0.24750	0.24225	0.23150
	Method 4	0.20622	0.2345	0.21712
MAE_θ/rad	Method 2	0.30812	0.34907	0.38273
	Method 3	0.18902	0.19196	0.19183
	Method 4	0.15342	0.17308	0.16602
ε_{phy}/p.u.	Method 2	/	/	/
	Method 3	14.88100	14.99122	15.44430
	Method 4	14.94105	15.05246	15.25623

Further comparison shows that methods 3 and 4, which introduce power-flow constraints, have better overall indicators than method 2, which does not impose physical constraints. This is because the power flow constraint can force the estimation results to remain within the physically feasible range of the power system, effectively suppressing the error amplification from measurement noise, thereby improving the accuracy and robustness of the estimation. Compared with method 3, method 4 further incorporates a temporal attention mechanism to model the dynamic evolution of system states across consecutive time intervals. By exploiting temporal dependencies in historical measurements, the model can provide additional temporal information for state estimation and achieve competitive or superior performance in most noise conditions. As the measurement noise level increases, the benefit brought by temporal correlations may be partially weakened, resulting in less pronounced improvements over method 3. Nevertheless, the proposed method generally achieves lower estimation errors in angle-related metrics and exhibits improved robustness under different measurement noise levels.

It should be noted that the robustness validation in this study is still limited. Although this setting is useful for evaluating the influence of different noise levels, it does not fully cover practical measurement problems in distribution networks.

4.4. Topology Change Adaptability Verification

In actual operation, the distribution network may undergo topology changes due to switch operations, fault isolation, and network reconstruction, which require adaptability in the state estimation model. Therefore, this section validates the estimation performance of the proposed method under structural change conditions by combining different topology scenarios, and analyzes the role of transfer learning-based adaptive strategies in new topology scenarios. This section constructs three topological change scenarios:

Scenario 1: Disconnect the communication switch between bus 28 and bus 29, and close the line between bus 25 and bus 29; Scenario 2: Disconnect the segment switch between bus 7 and bus 8, and close the connection switch between bus 8 and bus 21; Scenario 3:

Disconnect the segment switch between bus 11 and bus 12, and close the connection switch between bus 12 and bus 22.

Regenerate the test set for each scenario and test the estimation performance of the following two modes: (a) apply the original model directly to the new topology; (b) Transfer learning using different numbers of fine-tuning samples under the new topology.

Table 7 presents the estimation accuracy under different topology-adaptation strategies. Taking Scenario 1 as an example, when the original model is directly applied to the new topology without adaptation, the voltage-amplitude RMSE reaches 0.36169, while the value obtained by full retraining is 0.31000. This indicates that topology changes affect the mapping relationship learned by the original model and reduce the state estimation accuracy. After transfer fine-tuning with 200, 700, and 1400 samples, the $RMSE_V$ decreases to 0.35565, 0.34825, and 0.32961, respectively. These results show that transfer fine-tuning can gradually improve the adaptability of the model to the new topology. The improvement in angle-related indicators is more obvious. In Scenario 1, the $RMSE_\theta$ decreases from 0.29161 to 0.23215 after fine-tuning with 700 samples, and the MAE_θ decreases from 0.22222 to 0.17072. Compared with full retraining, transfer fine-tuning still has a certain gap in some voltage-related indicators and physical residuals. As the number of fine-tuning samples increases, the performance of the fine-tuned model gradually approaches that of full retraining. Therefore, transfer fine-tuning provides a practical adaptation strategy for topology changes, especially when rapid model updating is required and full retraining is computationally expensive.

Table 7. Evaluation index values under various topology change scenarios $/10^{-2}$.

Scenario	Index	Original Model	Finetune 200	Finetune 700	Finetune 1400	Full Retraining
1	$RMSE_V$/p.u.	0.36169	0.35565	0.34825	0.32961	0.31000
	MAE_V/p.u.	0.26428	0.26142	0.26693	0.25613	0.24334
	$RMSE_\theta$/rad	0.29161	0.28748	0.23215	0.23451	0.22978
	MAE_θ/rad	0.22222	0.21920	0.17072	0.16981	0.16209
	ε_{phy}/p.u.	2.42066	2.41181	2.28246	2.2359	2.22477
2	$RMSE_V$/p.u.	0.36368	0.35747	0.33932	0.32053	0.31887
	MAE_V/p.u.	0.27167	0.26777	0.26168	0.24605	0.24535
	$RMSE_\theta$/rad	0.60737	0.60187	0.48315	0.45105	0.43229
	MAE_θ/rad	0.47309	0.47309	0.39384	0.35383	0.32742
	ε_{phy}/p.u.	2.48651	2.47827	2.35074	2.31566	2.28258
3	$RMSE_V$/p.u.	0.35227	0.34584	0.31635	0.30792	0.28117
	MAE_V/p.u.	0.26183	0.25629	0.22849	0.22520	0.20343
	$RMSE_\theta$/rad	0.43270	0.43189	0.41852	0.40904	0.40350
	MAE_θ/rad	0.32396	0.32291	0.33749	0.31960	0.31069
	ε_{phy}/p.u.	2.52162	2.50996	2.36803	2.31791	2.29962

In addition, to reduce the randomness of individual estimates, Figures 5–7 present 100 estimates of the index for three scenarios. It is intuitive that the model trained on the original topology generally has a higher estimation error than the model fine-tuned on the target topology in the new topology. Compared with direct application, transfer learning can effectively reduce estimation errors across all three reconstruction scenarios, and the improvement in angle-based indicators is particularly significant. As the number of fine-tuning samples increased from 200 to 1400, the performance of the model further improved, indicating that introducing appropriate target topology samples can significantly enhance the model's adaptability to network reconstruction scenarios.

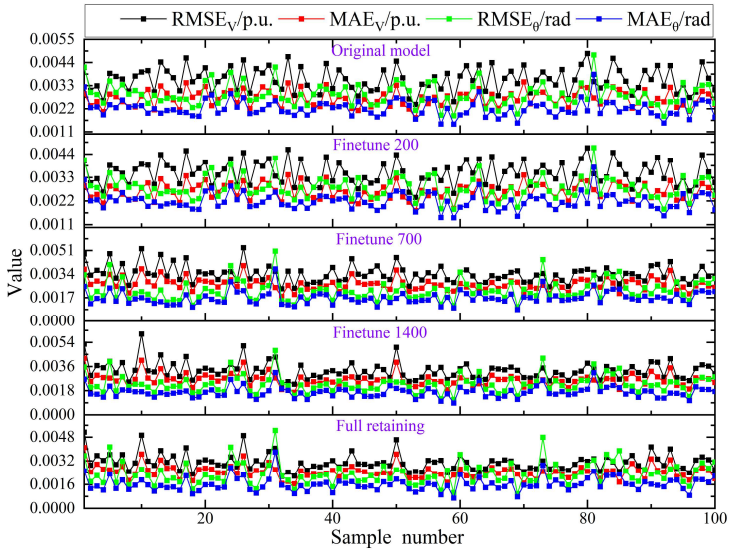


Figure 5. Estimate the index value 100 times in Scenario 1.

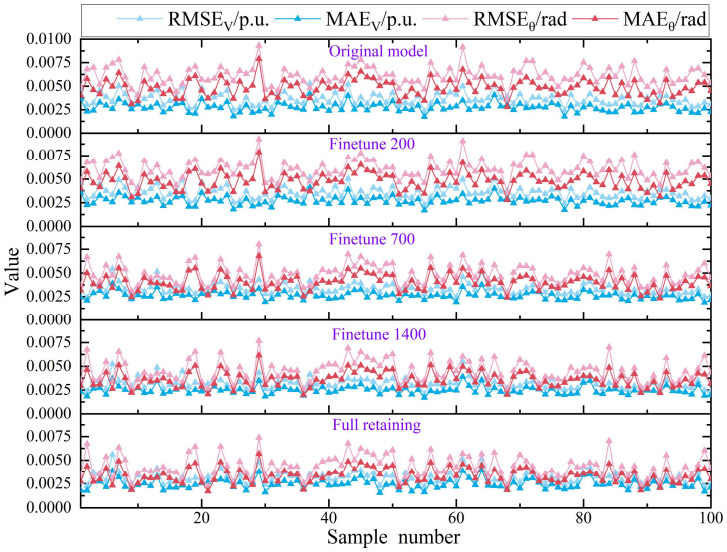


Figure 6. Estimate the index value 100 times in Scenario 2.

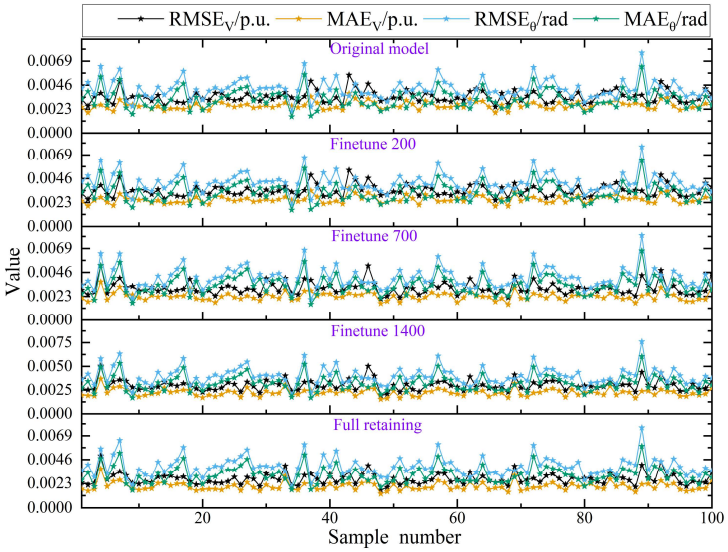


Figure 7. Estimate the index value 100 times in Scenario 3.

4.5. Method Scalability Verification

To assess the applicability of the IGCN method to large-scale systems, scalability tests were performed on the IEEE 118-bus system [46]. The measurement configuration remains based on the principle of low redundancy. The specific topology and measurement configuration are illustrated in Figure 8. The dataset construction method is consistent with the previous approach, involving the generation of samples through power flow calculations and the division of the data into training, validation, and testing sets in a ratio of 7:1:2. Acknowledging that the convergence of the model necessitates more training iterations as the system size increases, the number of training epochs was adjusted to 1000, while all other model parameters were maintained at the original settings.

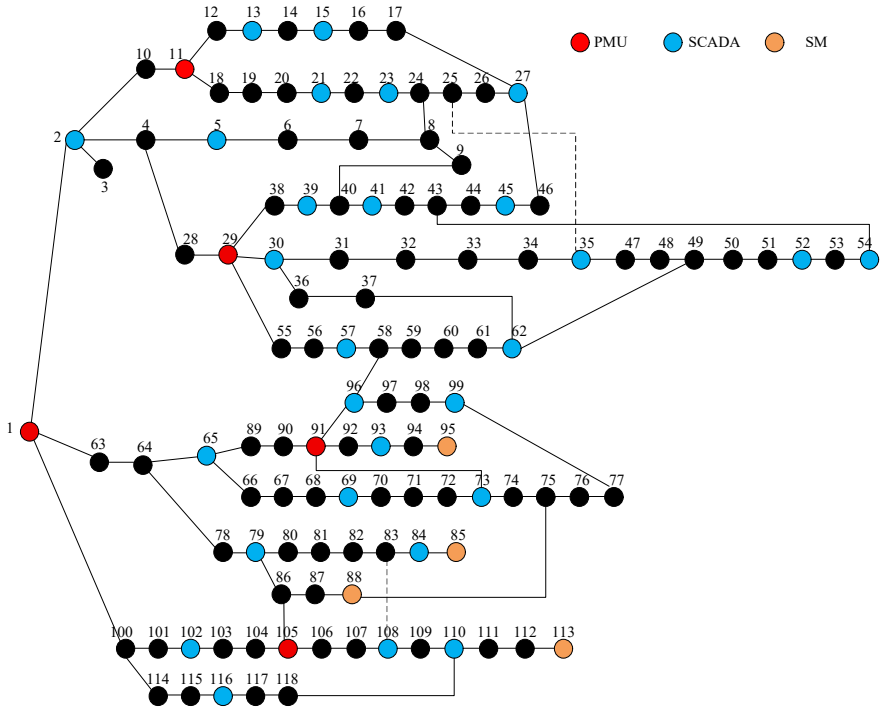


Figure 8. Schematic diagram of the 118 bus system and its measurement configuration.

After the model training is completed, randomly select a time section from the IEEE 118 bus test set, and the estimation performance of each algorithm is shown in Figure 9. It is evident that the estimation curve of the IGCN algorithm proposed in this article closely aligns with the true value curve. This suggests that the IGCN approach is more effective at capturing the distribution patterns of electrical states in complex topologies, and the introduction of power-flow constraints effectively suppresses estimation deviations caused by sparse measurements in local areas.

Table 8 compares the state estimation accuracy of various methods on the IEEE 118 bus test set. Compared with the 33 bus systems, as the number of buses and network topology complexity increase, the error metrics of each method increase to varying degrees. However, the method proposed in this paper still maintains a low error across various evaluation indicators, indicating that it can effectively extract inter-node correlation information in complex power grid topologies. Therefore, the IGCN method still achieves good estimation accuracy. Figures 10 and 11 further illustrate the estimation errors and physical residuals of the IGCN algorithm at each bus. It can be seen that the estimation errors of most buses are concentrated within 0.08 p.u., indicating that the overall estimation results of the algorithm are relatively stable. Meanwhile, the error reduction of the IGCN method is more pronounced at the end buses, indicating that the IGCN can enhance long-distance

information propagation and effectively alleviate the information insufficiency caused by the remote topology of the end buses. In addition, the physical residuals on these buses remain low, indicating that the estimation results better comply with the physical constraints of the power system, further verifying the engineering applicability of our method in large-scale testing systems.

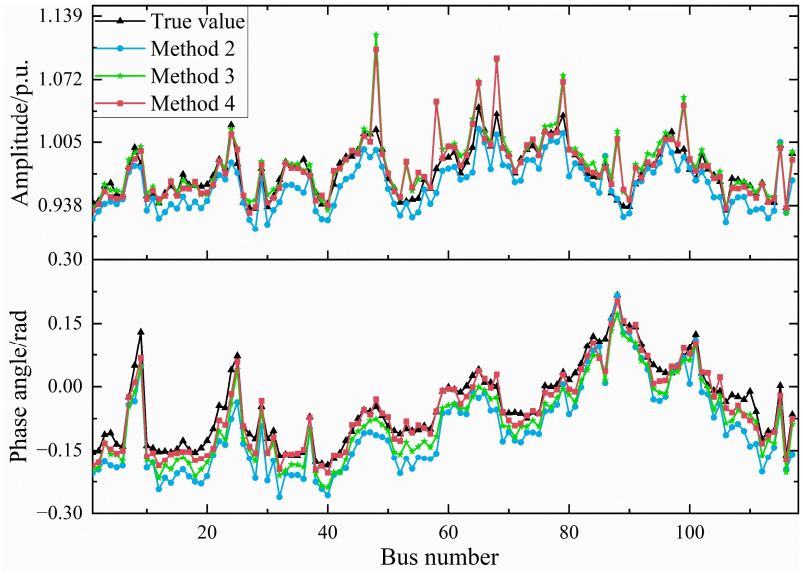


Figure 9. Comparison of the estimation effects of each bus in the 118 bus system.

Table 8. Evaluation index values under multiple scenarios/10^{-2}.

Index	Method 2	Method 3	Method 4
$RMSE_V$/p.u.	1.7004	1.8037	1.6613
MAE_V/p.u.	1.5143	0.9849	0.9116
$RMSE_\theta$/rad	6.2754	4.1266	2.3869
MAE_θ/rad	5.7472	3.7840	1.8720
ε_{phy}/p.u.	/	136.13779	136.80586

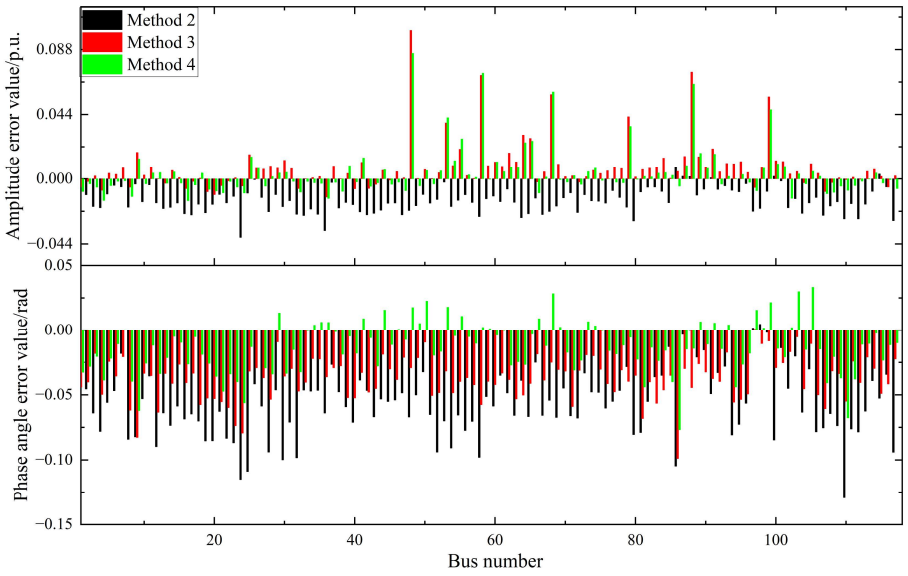


Figure 10. Amplitude phase angle error of each bus in the 118 bus system.

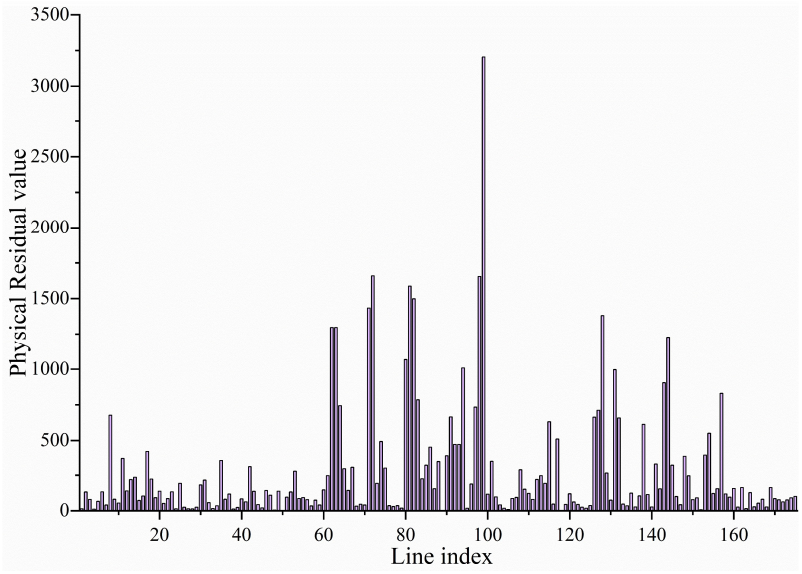


Figure 11. Physical residual value of IGCN in the 118 bus system.

5. Conclusions

This paper proposes an IGCN-based state estimation method for distribution networks with heterogeneous measurements and topology changes. The proposed method constructs an attributed graph representation of the distribution network by jointly considering measurement values, availability masks, measurement variances, and pseudo-measurement labels. By introducing edge feature encoding, graph attention, reliability gating, and power flow constraints, the model improves the use of heterogeneous measurement information and enhances the physical consistency of the estimation results. The experimental results on the IEEE 33-bus and IEEE 118-bus systems show that the proposed method achieves better overall estimation performance than the compared methods under different measurement noise levels and topology reconfiguration scenarios.

However, the current study still has some practical limitations. First, the experiments are mainly based on simulated data generated from standard IEEE test systems, and the performance of the method has not yet been verified using real distribution network operation data. Second, although different noise levels and topology changes are considered, more complex practical conditions, such as missing measurements, asynchronous data, communication delays, inaccurate device parameters, and delayed topology updates, are not fully covered. These factors may affect the performance and reliability of the model in industrial applications.

In future work, we will further evaluate the proposed method using real-world distribution network data and more complex measurement error scenarios. In addition, we will explore ways to embed physical constraints more tightly into the model architecture to improve the robustness, interpretability, and engineering applicability of the proposed state estimation method.

Author Contributions: Writing—original draft preparation, J.L.; funding acquisition, J.L.; writing—review and editing, Y.L.; visualization, Z.Z.; software, Z.Z.; supervision, S.X.; data curation, S.X.; project administration, X.F.; Formal analysis, X.F. All authors have read and agreed to the published version of the manuscript.

Funding: State Grid Fujian Electric Power Co., Ltd.—Research on power quality impact analysis and accurate evaluation technology under multi-mode operation of rail transit (521304250006).

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: All Authors were employed by State Grid Fujian Electric Power Co., Ltd. Electric Power Research Institute. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

WLS	Weighted least squares
PMU	Phasor measurement unit
SM	Smart meter
GNN	Graph neural network
GCN	Graph convolutional network
IGCN	Improved graph neural network

References

1. Zhou, L.; Shao, Z.; Cheng, G.; Lin, J.; Chen, F. Interval harmonic state estimation of three-phase asymmetric distribution network by integrating multi-source data. *Prot. Control Mod. Power Syst.* **2026**, *11*, 62–76.
2. Sundaray, P.; Weng, Y. Alternative auto-encoder for state estimation in distribution systems with unobservability. *IEEE Trans. Smart Grid* **2023**, *14*, 2262–2274.
3. Hua, D.; Zhang, L.; Tang, W.; Wu, J.; Qian, T. A spatiotemporal diffusion framework for state estimation in low-observability distribution systems. *Int. J. Electr. Power Energy Syst.* **2026**, *177*, 111839. [CrossRef]
4. Cavraro, G.; Comden, J.; Dall'Anese, E.; Bernstein, A. Real-time distribution system state estimation with asynchronous measurements. *IEEE Trans. Smart Grid* **2022**, *13*, 3813–3822. [CrossRef]
5. Huang, J.; Zhou, X.; Cui, B. Online distribution system state estimation via stochastic gradient algorithm. *Electr. Power Syst. Res.* **2022**, *213*, 108305. [CrossRef]
6. Li, K.; Han, X. A distributed Gauss-Newton method for distribution system state estimation. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107694.
7. Xu, J.; Liu, J.; Wu, Z.; Jiang, W.; Wu, Q. Multi-area state estimation for active distribution networks under multiple uncertainties: An affine approach. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109632.
8. Soltani, Z.; Khorsand, M. Real-time topology detection and state estimation in distribution systems using micro-PMU and smart meter data. *IEEE Syst. J.* **2022**, *16*, 3554–3565.
9. Mokaribolhassan, A.; Nourbakhsh, G.; Ledwich, G.; Arefi, A.; Shafiei, M. Distribution system state estimation using PV separation strategy in LV feeders with high levels of unmonitored PV generation. *IEEE Syst. J.* **2023**, *17*, 684–695.
10. Azimian, B.; Biswas, R.S.; Moshtagh, S.; Pal, A.; Tong, L.; Dasarathy, G. State and topology estimation for unobservable distribution systems using deep neural networks. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 9003514. [CrossRef] [PubMed]
11. Liu, Y.; Ma, J. Noise-trained deep learning-based distribution system state estimation considering the penetration of distributed energy resources. *IEEE Trans. Smart Grid* **2025**, *16*, 2292–2303.
12. Liang, D.; Li, G.; Liu, X.; Zeng, L.; Chiang, H.-D.; Wang, S. Bayesian state estimation for partially observable distribution networks via power flow-informed neural networks. *Int. J. Electr. Power Energy Syst.* **2025**, *170*, 110886. [CrossRef]
13. Zhang, J.; Zhao, J.-B.; Cheng, G.; Rouhani, A.; Chen, X. Explainable multi-fidelity Bayesian neural network for distribution system state estimation. *Appl. Energy* **2025**, *392*, 125972. [CrossRef]
14. Raghuvamsi, Y.; Teeparthi, K.; Vinod Kumar, D.M.; Abdul, I.; Parri, S. Distribution system state estimation using physics-guided deep learning approach. *Electr. Power Syst. Res.* **2024**, *236*, 110922. [CrossRef]
15. Gomez de la Varga, J.; Pineda, S.; Morales, J.M.; Porras, A. Learning-based state estimation in distribution systems with limited real-time measurements. *Electr. Power Syst. Res.* **2025**, *241*, 111268.
16. Shafiei, M.; Ledwich, G.; Nourbakhsh, G.; Arefi, A.; Pezeshki, H. One-step multilayer unbalanced three-phase DSE for LV distribution networks. *IEEE Syst. J.* **2021**, *15*, 4403–4412.
17. Habib, B.; Isufi, E.; van Breda, W.; Jongepier, A.; Cremer, J.L. Deep statistical solver for distribution system state estimation. *IEEE Trans. Power Syst.* **2024**, *39*, 4039–4050.
18. Wu, H.; Xu, Z.; Wang, M. Unrolled spatiotemporal graph convolutional network for distribution system state estimation and forecasting. *IEEE Trans. Sustain. Energy* **2023**, *14*, 297–308.
19. Ren, Z.; Chu, X.; Ye, H. State estimation for reconfigurable distribution systems using graph attention network. *IEEE Access* **2023**, *11*, 107237–107250. [CrossRef]

20. Wu, H.; Xu, Z.; Wang, M.; Lyu, X. Meta-learning enhanced physics-informed graph attention convolutional network for distribution power system state estimation. *IEEE Trans. Netw. Sci. Eng.* **2025**, *12*, 1186–1197.
21. Lin, H.; Sun, Y. EleGNN: Electrical-model-guided graph neural networks for power distribution system state estimation. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 5292–5298.
22. Wu, H.; Jia, Y.; Xu, Z. Adaptive graph convolutional network-based distribution system state estimation. In Proceedings of the 2022 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 17–21 July 2022; pp. 1–5.
23. Ngo, Q.-H.; Nguyen, B.L.H.; Vu, T.V.; Zhang, J.; Ngo, T. Physics-informed graphical neural network for power system state estimation. *Appl. Energy* **2024**, *358*, 122602. [CrossRef]
24. Tran, M.-Q.; Zamzam, A.S.; Nguyen, P.H. Enhancement of distribution system state estimation using pruned physics-aware neural networks. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 28 June–2 July 2021; pp. 1–5.
25. Crawford, K.; Baran, M.E. A signature-based analysis for topology error detection in distribution system state estimation. *IEEE Trans. Power Syst.* **2026**, *41*, 2571–2578. [CrossRef]
26. Parlaktuna, K.; Dursun, E.; Göl, M. WLAV state estimation based topology error detection and identification in distribution networks with limited number of measurements. *Electr. Power Syst. Res.* **2024**, *232*, 110406. [CrossRef]
27. Raghuvamsi, Y.; Batchu, S.; Teeparthi, K. Topology and FDIA identification in distribution system state estimation using a data-driven approach. *Measurement* **2025**, *253*, 117741. [CrossRef]
28. Rajasekaran, J.R.K.; Natarajan, B.; Jiang, J. Online bad data detection in compressed sensing-based distribution system state estimation. *IEEE Trans. Smart Grid* **2025**, *16*, 3449–3452. [CrossRef]
29. Raghuvamsi, Y.; Teeparthi, K. Detection and reconstruction of measurements against false data injection and DoS attacks in distribution system state estimation: A deep learning approach. *Measurement* **2023**, *210*, 112565. [CrossRef]
30. Hebling, G.M.; Massignan, J.A.D.; London, J.B.A.; de Oliveira, R. Sparse and orthogonal method for fast bad data processing in distribution system state estimation. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 28 June–2 July 2021; pp. 1–6.
31. Zhang, X.; Ge, S.; Zhou, Y.; Liu, H. Deep learning framework for low-observable distribution system state estimation with multitimescale measurements. *IEEE Trans. Ind. Inform.* **2024**, *20*, 13273–13283. [CrossRef]
32. de Jongh, S.; Mueller, F.; Canizares, C.A.; Leibfried, T.; Bhattacharya, K. Distribution grid state estimation with limited actual and pseudo measurements. *IEEE Trans. Smart Grid* **2025**, *16*, 3638–3652. [CrossRef]
33. Zeraati, M.; Sheibani, M.R.; Jabari, F.; Heydarian-Forushani, E. A novel state estimation method for distribution networks with low observability based on linear AC optimal power flow model. *Electr. Power Syst. Res.* **2024**, *228*, 110085.
34. Kamal, M.; Li, W.; Deka, D.; Mohsenian-Rad, H. Physics-conditioned generative adversarial networks for state estimation in active power distribution systems with low observability. In Proceedings of the 2022 International Conference on Smart Grid Synchronized Measurements and Analytics, Split, Croatia, 24–26 May 2022; pp. 1–6.
35. Chen, Z.; Ziras, C.; Bindner, H.W. A mixed formulation of matrix completion in low voltage distribution network state estimation. In Proceedings of the 2023 IEEE Belgrade PowerTech, Belgrade, Serbia, 25–29 June 2023; pp. 1–6.
36. Liu, Y.; Wang, Y.; Yang, Q. Spatio-temporal generative adversarial network based power distribution network state estimation with multiple time-scale measurements. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9790–9797.
37. Li, J.; Gao, M.; Liu, B.; Cai, Y. Forecasting aided distribution network state estimation using mixed micro-PMU-RTU measurements. *IEEE Syst. J.* **2022**, *16*, 6524–6534.
38. Bajagain, S.; Dubey, A. Iterative distribution system state estimation for integrated primary and split-phase secondary feeder monitoring. *IEEE Trans. Power Deliv.* **2023**, *38*, 1674–1687.
39. Zhou, Y.; Centeno, V. Adaptive linear state estimation for unbalanced distribution system. *IEEE Access* **2025**, *13*, 63417–63427. [CrossRef]
40. Niu, S.; Liu, W.; Liu, C.; Lyu, J.; Yang, C.; Chau, K.T. Wireless power transfer for IoT bolts in railways: Overcoming coupling challenges due to conductive and ferromagnetic rails. *IEEE Trans. Ind. Electron.* **2026**, *73*, 5409–5419.
41. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858.
42. Zhou, L.; Shao, Z.; Chen, F.; Cheng, G.; Lin, C.; Chen, Z.; Zheng, Z. A physics-informed topology-adaptive graph convolutional network for harmonic source location in three-phase distribution networks with renewable energy integration. *IEEE Trans. Power Deliv* 2026, *Epub ahead of printing*.
43. Zhang, M.; Yu, Z.; Xu, Z. Short-term load forecasting using recurrent neural networks with input attention mechanism and hidden connection mechanism. *IEEE Access* **2020**, *8*, 186514–186529. [CrossRef]
44. Baloch, M.; Honnurvali, M.S.; Ahmed, T.; Sheikh, A.M.; Chaudhary, S.T. Federated learning-based distributed solar forecasting for smart buildings in Muscat, Oman using GRU networks. *Energies* **2026**, *19*, 2496. [CrossRef]

45. Falkner, S.; Klein, A.; Hutter, F. BOHB: Robust and efficient hyperparameter optimization at scale. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 1437–1446.
46. Haddadi, A.; Rezaei-Zare, A.; Gerin-Lajoie, L.; Hassani, R.; Mahseredjian, J. A modified IEEE 118-bus test case for geomagnetic disturbance studies-Part I: Model data. *IEEE Trans. Electromagn. Compat.* **2020**, *62*, 955–965.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.