Robust Distribution System State Estimation Considering Anomalous Real-time Measurements and Topology Change

Jiaxiang Hu, Weihao Hu, *Senior Member*, *IEEE*, Di Cao, Jianjun Chen, Sayed Abulanwar, *Senior Member*, *IEEE*, Mohammed K. Hassan, Zhe Chen, *Fellow*, *IEEE*, and Frede Blaabjerg, *Fellow*, *IEEE*

Abstract—This paper develops a physics-guided graph network to enhance the robustness of distribution system state estimation (DSSE) against anomalous real-time measurements, as well as a deep auto-encoder (DAE)-based detector and a Gaussian process-aided residual learning (GARL) to deal with challenges arising from topology changes. A global-scanning jumping knowledge network (GSJKN) is first designed to establish the regression rule between the measurement data and state variables. The structural information of distribution system (DS) and a global-scanning module are incorporated to guide the propagation of scarce measurements in the graph topology, contributing to valid estimation precision in sparsely measured DSs. To monitor the topology changes of the network, a DAE network is employed to learn an efficient representation of the measurements of the system under a certain topology, which can achieve online monitoring of the network structure by observing the variation tendency of the reconstruction error. When the topology change occurs, a Gaussian process with a composite kernel is applied to the modeling of the pre-trained GSJKN residual to adapt to the new topology. The embedding of the physical structural knowledge enables the proposed GSJKN method to restore the missing/noisy values utilizing the adjacent measurements, which enhances the robustness to typical data acquisition errors. The adopted DAE network and special GARL-based transfer method further allow the DSSE method to rapidly detect and adapt to the topology change, as well as achieve effective quantification of the estimation uncertainties. Comparative tests on balanced and unbalanced systems demonstrate the accuracy, robustness, and adaptability of the proposed DSSE method.

I. INTRODUCTION

THE widespread integration of distributed generation (DG) [1] and demand response programs [2] has transformed the distribution system operators into active market entities. Distribution system state estimation (DSSE) tasks face challenges of poor observability, complex and variable topology, and inaccurate physical parameters [3]. Additionally, distributed energy resources introduce significant uncertainties, further complicating DSSE [4]. Therefore, robust and adaptable DSSE methods are essential to deliver accurate, reliable system state data for effective DS control and management [5].

DSSE methods are typically classified into two categories: optimization-based [6] - [14] and learning-based methods. Weighted least squares (WLS)-based methods [7]-[8], initially developed for transmission systems, have been adapted for DS applications and provide accurate estimations when physical models and measurement data are reliable. However, measurement data often contain unknown noise and errors [14], and the WLS-based method lacks robustness under such conditions. Robust DSSE methods have been proposed to address these data issues [12]-[14], yet they still rely on accurate physical models. In practice, determining precise DS model parameters is challenging, and actual topology conditions frequently differ from recorded data due to ongoing changes. This discrepancy poses challenges to traditional optimization-based DSSE methods, which struggle with inaccurate system parameters and complex conditions [15].

Advancements in distribution automation (DA) systems and machine learning (ML) have introduced learning-based state estimation methods [16]-[19], which shift computation to offline training and reduce reliance on precise physical model parameters. However, standard ML methods often overlook DS-specific physical information, resulting in significant estimation errors [20] when real-time data are missing or noisy due to packet loss or communication issues. To address this, researchers are incorporating physical informa-

J. Hu, W. Hu, D. Cao (corresponding author), and J. Chen are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: jx.hu@foxmail.com; whu@uestc.edu.cn; caodi@uestc.edu.cn; chjj@std.uestc.edu.cn).

S. Abulanwar and M. K. Hassan are with the Faculty of Engineering, Horus University-Egypt, Almaadi 51718, Egypt (e-mail: abulanwar@mans. edu. eg; mkhassan@horus.edu.eg).

Z. Chen and F. Blaabjerg are with the Faculty of Engineering and Science, Aalborg University, Aalborg 9220, Denmark (e-mail: zch@energy.aau.dk; fbl@et.aau.dk).

tion into neural network (NN)-based state estimation models [20] - [27]. For instance, [20] develops a topology-specific NN for state estimation, while [21] applies a multi-layer graph convolution network (GCN) for voltage prediction.

While these methods incorporate DS topology to improve model performance and generalization, they have mostly been tested under conditions of abundant real-time measurements. In practice, DSs often lack sufficient measurements, leaving many nodes without critical data and resulting in ineffective graph aggregation due to insufficient neighbor node features. Although increasing graph aggregation layers [20], [21] can extend information gathering from more distant nodes, this method is limited by over-smoothing and fails to address sparse measurements in large-scale DSs. Furthermore, frequent topology changes in DS operations require learning-based DSSE models to be retrained to update mapping relationships, yet only limited data are typically available under new topologies, hindering the effective training process. With rising DG penetration, there is also an increasing demand for probabilistic state estimation methods to quantify estimation uncertainties. To this end, this paper develops a robust topology change-aware DSSE method by systematically integrating a structure-guided state estimator, a deep auto-encoder (DAE)-based detector, and a Gaussian process (GP)-aided residual learning (GARL). The main contributions are as follows.

1) The proposed global-scanning jumping knowledge network (GSJKN) method expands the application of physics-guided methods under scarce real-time measurements. It is realized by adaptively selecting the range of graph aggregation and designing a global-scanning module based on recurrent neural network (RNN) structures to obtain the feasible node representation, which allows it to achieve satisfying estimation precision under scarce measurements, as well as restore the missing/noisy measurement data utilizing the adjacent information during the propagation of nodal features.

2) The proposed DSSE method can achieve online detection of the topology change events according to real-time and pseudo-measurement data. The DAE takes the real and pseudo-measurements as inputs and reconstructs them via multiple layers of transformation. The encoding and decoding process enables the DAE network to learn the intrinsic structure of the measurements under a certain topology in an unsupervised manner. This allows us to identify the topology change by observing the trend of reconstruction error in an online manner.

3) The GARL-based transfer method allows the proposed model to realize fast adaption to a new topology utilizing sparse online measurement data. Instead of modeling the DSSE under a new topology as a new task, the proposed method employs the GP with a composite kernel to model the residual of the pre-trained GSJKN under the original topology. The adopted composite kernel allows the proposed method to realize inductive reasoning about the differences between the DSSE tasks under different topologies and rapidly adapt to the new topology using a limited amount of data. This differentiates from the traditional parametric transfer methods that still require a certain amount of historical data

to adapt to new topology conditions.

4) The Bayesian characteristic of the GARL enables the proposed method to effectively quantify the uncertainties of the DSSE results. This is beneficial for the operators when making uncertainty-aware decisions.

The rest of this paper is organized as follows. Section II presents the problem statement. Section III describes the proposed DSSE and fast transfer framework. Section IV presents the case study. Finally, Section V concludes this paper.

II. PROBLEM RESTATEMENT

A. Classical Optimization-based DSSE Method

Consider the system state variables as $\boldsymbol{x}=[x_1,x_2,...,x_n]^{\mathrm{T}}$ and the system measurement variables as $\boldsymbol{z}=[z_1,z_2,...,z_m]^{\mathrm{T}}$, where n and m are the numbers of nodes and system measurements, respectively. The DSSE model $h(\cdot)$ is a measurement equation based on the DS structure, line parameters, state variables $\boldsymbol{x}$, and measurement variables $\boldsymbol{z}$.

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

where $\boldsymbol{v}$ is the measurement error. State estimation solves the estimated state variable $\boldsymbol{x}^*$ so that the measured $\boldsymbol{z}$ is most likely to be observed, which can be illustrated as:

$$P(\boldsymbol{z},\boldsymbol{x}^*)=\max(P(\boldsymbol{z},\boldsymbol{x})) \tag{2}$$

where $P(\cdot)$ is the probability distribution density function. The essence of the WLS-based method is to solve the following mathematical problem:

$$\boldsymbol{x}=\arg\min(\boldsymbol{z}-h(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{W}_s(\boldsymbol{z}-h(\boldsymbol{x})) \tag{3}$$

where $\boldsymbol{W}_s=\mathrm{diag}(\sigma_y^2)\in\mathbb{R}^{m\times m}$ is the measurements weight matrix, and $y=1,2,...,m$, and σ_y is the weight factor of the yth measurement. The nonlinear optimization problem of (3) can be solved by iteration method. In this case, the accurate model parameters in $h(\cdot)$ and pre-set noise conditions $\boldsymbol{W}_s$ about measurement data are important for estimation precision. Therefore, the WLS-based method relies on the accurate physical model parameters of DS and is less robust to abnormal missing/noisy measurement values.

B. Learning-based DSSE Method

When learning-based methods are utilized to deal with DSSE tasks, the task is generally transformed into a supervised regression learning process from the historical data. Consider a large amount of historical data collected by DA as $\{\boldsymbol{z}_i,\boldsymbol{x}_i\}_{i=1}^T$. The DSSE task can be illustrated as $\boldsymbol{x}_i^*=f(\boldsymbol{z}_i)$, where $f(\cdot)$ is the mapping function constructed by NN or ML models. However, due to the neglect of structure information in the typical learning-based methods, the abnormal data will significantly impact their estimation results. Researchers have proposed physics-guided NN to incorporate structure information, which is represented as:

$$\boldsymbol{x}^*=G_{n,Conv}(\boldsymbol{T},\boldsymbol{z}) \tag{4}$$

where $\boldsymbol{T}$ is the prior topology information; and $G_{n,Conv}$ is the graph aggregation with n layers to obtain the information from n^{th} class neighbor nodes. Therefore, the common physics-guided methods face the following challenges. ① To gather information from distant nodes, additional graph lay-

ers are required, leading to limited performance in large-scale systems with scarce real-time measurements, ② They depend on substantial historical data for DSSE model re-training, which are difficult to obtain in scenarios such as new topology change.

III. PROPOSED DSSE AND FAST TRANSFER FRAMEWORK

A. Robust DSSE Based on Proposed GSJKN Method

The proposed GSJKN-based estimator consists of two main components: ① graph jump connections through multi-layer graph aggregation to obtain the neighbor information in a certain range; and ② a subsequent adaptive global-scanning module based on RNN cells. Firstly, the graph jump connections are constructed by aggregating the node embeddings from multiple graph layers. Consider the collected historical data as $\{z_i, x_i\}_{i=1}^{\mathrm{T}}$ and the prior topology information as $\boldsymbol{T} \in \mathbb{R}^{N \times N}$, where N is the node number in DS. The graph jump connections can be represented as:

$$\boldsymbol{H}^{(l)} = \Phi_l(\boldsymbol{T}, \boldsymbol{H}^{(l-1)}) \quad l = 1, 2, \dots, L \tag{5}$$

$$\boldsymbol{H}_{JK} = concat(\boldsymbol{H}^{(1)}, \boldsymbol{H}^{(2)}, \dots, \boldsymbol{H}^{(L)}) \tag{6}$$

where $\boldsymbol{H}^{(l)}$ is the node representation after the lth graph layer; $\boldsymbol{H}_{JK}$ is the node representation after graph jump connections; $concat(\cdot)$ is the splicing operation; and Φ_l is the parameterized l^{th} graph aggregation layer. Specifically, $\boldsymbol{H}^{(l)}$ is calculated by splicing M graph embedding heads:

$$\boldsymbol{H}^{(l)} = concat(\boldsymbol{U}_1^{(l)}, \boldsymbol{U}_2^{(l)}, \dots, \boldsymbol{U}_M^{(l)}) \tag{7}$$

where $\boldsymbol{U}_m^{(l)} = \{\boldsymbol{u}_{mi}^{(l)}\}_{i=1}^N$ is the node embedding from the mth heads, and $\boldsymbol{u}_{mi}^{(l)}$ is the feature from the ith node in $\boldsymbol{U}_m^{(l)}$, which is calculated by:

$$\boldsymbol{s}_{mi}^{(l)} = \boldsymbol{W}_m^{(l)} \boldsymbol{h}_i^{(l-1)} + \boldsymbol{b}_m^{(l)} \tag{8}$$

$$\begin{cases} e_{mij}^{(l)} = Leakyrelu(\boldsymbol{\theta}_m^{(l)}(\boldsymbol{s}_{mi}^{(l)} \| \boldsymbol{s}_{mj}^{(l)})) & \boldsymbol{T}_{ij} = 1 \\ e_{mij}^{(l)} = 0 & \boldsymbol{T}_{ij} = 0 \end{cases} \tag{9}$$

$$a_{mij}^{(l)} = \frac{\exp(e_{mij}^{(l)})}{\sum\limits_{k \in N} \exp(e_{mik}^{(l)})} \tag{10}$$

$$\boldsymbol{u}_{mi}^{(l)} = \sigma\left(\sum_{j \in N} a_{mij}^{(l)} \boldsymbol{s}_{mj}^{(l)}\right) \tag{11}$$

where $\boldsymbol{h}_i^{(l-1)} \in \boldsymbol{H}^{(l-1)}$ is the feature of the ith node in $\boldsymbol{H}^{(l-1)}$; $\boldsymbol{W}_m^{(l)}$ and $\boldsymbol{b}_m^{(l)}$ are the learnable parameters at the mth graph embedding head in the lth graph layer to obtain the node embeddings $\boldsymbol{s}_{mi}^{(l)}$; $a_{mij}^{(l)}$ is the normalized attention value between nodes i and j; $\boldsymbol{T}_{ij} = 1$ is the structure connection between nodes i and j; $\|$ is the splicing operation; $e_{mij}^{(l)}$ is the feature at the edge between nodes i and j; $Leakyrelu(\cdot)$ is the leaky rectified linear unit function; and $\sigma(\cdot)$ is the selected activation function. The embeddings from nodes i and j are spliced and transformed to the edge feature $e_{mij}^{(l)}$ between them under $\boldsymbol{T}_{ij} = 1$. Then, these edge features around node i are normalized to obtain the edge attention value $\{a_{mij}^{(l)}\}_{j=1}^N$ to indicate the importance of neighbor nodes. Subsequently, the node embeddings $\{\boldsymbol{s}_{mj}^{(l)}\}_{j=1}^N$ and attention value $\{a_{mij}^{(l)}\}_{j=1}^N$ are weighted to obtain the new representation $\boldsymbol{u}_{mi}^{(l)} \in \boldsymbol{U}_{mi}^{(l)}$ for node i through $\sigma(\cdot)$.

Secondly, a global-scanning module based on the bidirectional RNN is designed to adaptively select the range of graph aggregation from the jump connections and propagate critical features in the whole topology. The new node representation $\boldsymbol{H}_G$ is calculated by:

$$\boldsymbol{H}_G = concat(RNN_F(\boldsymbol{H}_{JK}), RNN_R(\boldsymbol{H}_{JK})) \tag{12}$$

where $RNN_F(\cdot)$ and $RNN_R(\cdot)$ are the forward and reversed RNN calculations, respectively; and $\boldsymbol{H}_{JK} = [\boldsymbol{h}_{JK1}, \boldsymbol{h}_{JK2}, \dots, \boldsymbol{h}_{JKN}]$ is the aggregated node embeddings from the jump connections. The forward RNN calculation in the global-scanning module is represented as:

$$v_{Fn} = \sigma_s(\boldsymbol{W}_{Fz}(\boldsymbol{c}_{Fn-1} \boldsymbol{h}_{JKn})) \tag{13}$$

$$r_{Fn} = \sigma_s(\boldsymbol{W}_{FR}(\boldsymbol{c}_{Fn-1} \boldsymbol{h}_{JKn})) \tag{14}$$

$$\hat{\boldsymbol{c}}_{Fn} = \tanh(\boldsymbol{W}_F(r_{Fn} \boldsymbol{c}_{Fn-1} \boldsymbol{h}_{JKn})) \tag{15}$$

$$\boldsymbol{c}_{Fn} = (1 - v_{Fn}) \boldsymbol{c}_{Fn-1} + v_{Fn} \hat{\boldsymbol{c}}_{Fn} \tag{16}$$

where $\boldsymbol{h}_{JKn} \in \boldsymbol{H}_{JK}$ is the feature of the nth node in $\boldsymbol{H}_{JK}$; $\boldsymbol{W}_{Fz}$, $\boldsymbol{W}_{FR}$, and $\boldsymbol{W}_F$ are the learnable parameters at forward RNN cells; v_{Fn} and r_{Fn} are the attention values to determine the influence of $\boldsymbol{c}_{Fn-1}$ to $\boldsymbol{c}_{Fn}$; $\sigma_s(\cdot)$ is the sigmoid function; and $\hat{\boldsymbol{c}}_{Fn}$ denotes the embedding considering $\boldsymbol{c}_{Fn-1}$, and $\{\boldsymbol{c}_{Fn}\}_{n=1}^N$ is the calculated node representation from $RNN_F(\boldsymbol{H}_{JK})$. Due to the memory mechanism of forward RNN cells, the node embeddings from node 1 to node $n-1$ will be adaptively aggregated and extended to node n according to (16). Similarly, the reversed RNN cells scan the node embeddings of the system and generate the feasible node representations. Therefore, the proposed GSJKN method effectively aggregates the node embeddings of the whole DS and generate feasible node representations for estimation tasks as $\boldsymbol{x}^* = M_{LP}(\boldsymbol{H}_G)$, where M_{LP} denotes the multilayer perceptron module. This enables the proposed GSJKN method to incorporate information from more distant nodes, easing model training and enhancing estimation performance under limited real-time measurements.

B. Topology Change Detection Based on DAE

When new topologies emerge, the original DSSE models may encounter significant estimation errors. Topology change is critical information in the operation of DS. Researchers have investigated various methods to tackle bus-branch and node-breaker topology issues [28], [29]. Different from these methods, this paper proposes a DAE-based detector to identify topology change events, as shown in Fig. 1.
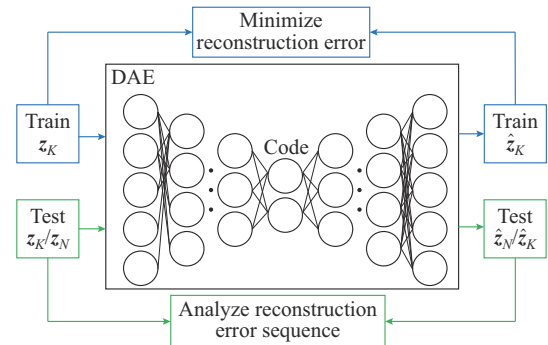

Fig. 1. Structure of DAE-based detector.

During the training process, given the measurement data z_K for the known topology, we construct the encoder mapping $\varphi(\cdot)$ to transmit z_K into a code c_K, which is formulated by:

$$c_K = \varphi(z_K) \qquad (17)$$

The code c_K retains key information from measurements z_K due to the unique bottleneck structure of DAE. We use a decoder $\rho(\cdot)$ to reconstruct $\hat{z}_K$ from c_K, which is formulated by:

$$\hat{z}_K = \rho(c_K) \qquad (18)$$

We want $\hat{z}_K$ to get close to the original measurement data z_K. The training process aims to derive the parameters that minimize the reconstruction error, which is formulated by:

$$\{W_{DAE}, l_{DAE}\} = \arg \min \sum_{i=1}^{T} |\hat{z}_{Ki} - z_{Ki}|_1 \qquad (19)$$

where W_{DAE} and l_{DAE} are the learnable parameters in DAE; and $|\hat{z}_{Ki} - z_{Ki}|_1$ is the reconstruction error e_{Ki} for the ith sample. By minimizing the reconstruction error, the encoder and decoder extract and preserve crucial feature information, particularly related to topology structure, from the input measurements z_K. However, when a topology change occurs, the power flow equation of DS undergoes modifications, resulting in measurements of new topology z_N that deviates from the previous measurement z_K. Consequently, the decoder struggles to accurately reconstruct the new measurements $\hat{z}_N$ close to z_N, leading to a significant increase in the reconstruction error e_K. To detect the topology change, we can monitor the fluctuation of the reconstruction error over time. A noticeable increase in e_K after a certain moment indicates the occurrence of topology change events for new measurements z_N. This DAE-based detector enables us to detect the topology changes in the DS, followed by the proposed GARL-based transfer method constructing the DSSE model for the new topologies.

C. Fast DSSE Model Based on GARL-based Transfer Method

Typically, obtaining historical data for a new topology is challenging, resulting in a scarcity of available data. To address this limitation and build a DSSE model for a new topology, a specially designed compound Gaussian kernel, denoted as $h_N(\cdot)$, is proposed. This kernel fully incorporates information from the known topology to overcome the lack of historical data under the new topology conditions. For the limited data of the new topology $\{z_{Ni}, x_{Ni}\}_{i=1}^{S}$, the DSSE model is expressed as:

$$x_N^* = h_N(z_N, f(T, z_N)) + f(T, z_N) \qquad (20)$$

where $f(\cdot)$ is the completed DSSE model for T. The limited data $\{z_{Ni}, x_{Ni}\}_{i=1}^{S}$ are used to train $h_N(\cdot)$, which may consist of a few dozen samples. The GARL achieves transfer learning by finding a mapping $h_N(\cdot)$, to model the residuals between $f(T, z_N)$ and x_N. This is accomplished by employing a GP with a composite kernel. $f(T, z_N)$ here is the pseudo-estimation result from the original model $f(\cdot)$, and x_N is the actual state of the new topology. The GARL-based transfer method consists of the training phase and the deployment phase [30]. In the training phase, the residuals r_i are calculated as:

$$r_i = x_{Ni} - f(T, z_{Ni}) \quad i = 1, 2, ..., S \qquad (21)$$

Let r denote the vector of all residuals and y^* denote the vector of all pseudo estimation results $f(T, z_N)$. A GP with a composite kernel is trained assuming $r \sim N(0, K_c((z_N, y^*), (z_N, y^*)) + \sigma_n^2 I)$, where I is the identity matrix, and $K_c((z_N, y^*), (z_N, y^*))$ denotes an $S \times S$ covariance matrix at all pairs of training points based on a composite kernel as:

$$K_c((z_N, y^*), (z_N, y^*)) = K_{in}(z_{Ni}, z_{Nj}) + K_{out}(y_i^*, y_j^*) \qquad (22)$$

where $i, j = 1, 2, ..., S$; K_{in} is the kernel to process z_N; and K_{out} is the kernel to process y^*. Suppose a linear kernel is used for both K_{in} and K_{out}. Then, the composite kernel can be expressed as:

$$K_c((z_N, y^*), (z_N, y^*)) = \sigma_{in}^2 (z_{Ni}^T z_{Nj}) + \sigma_{out}^2 ((y_i^*)^T y_j^*) \qquad (23)$$

The training process of GP learns the hyperparameters σ_{in}^2, σ_{out}^2, σ_n^2 by maximizing the marginal likelihood $\log_p(r|z_N, y^*)$. In the deployment phase, a test point z_{NT} is input to the DSSE model to get an output $\hat{y}^*$. The well-trained GP can establish the distribution of the residual as $\hat{r}|z_N, y^*, r, z_{NT}, \hat{y}^* \sim N(\hat{r}, var(\hat{r}))$, where $\hat{r} = k_*^T (K_c((z_N, y^*), (z_N, y^*)) + \sigma_n^2 I)^{-1} r$, and $var(\cdot)$ indicates the variance. Here, k_*^T is the vector of kernel-based covariances. The predicted residuals $\hat{r}$ will modify the output of $f(T, z_N)$ so that it can be applied to the new topology. In addition, the final estimation results x_N^* with uncertainty information is given as:

$$x_N^* \sim N(\hat{y}^* + \hat{r}, var(\hat{r})) \qquad (24)$$

$$x_N^* = h_N(z_N, \hat{y}^*) + \hat{y}^* \qquad (25)$$

GARL enables output reconstruction of the original DSSE model to adapt to new topologies while providing uncertainty estimates for state variables under these conditions. Without altering the architecture or retraining the DSSE model, GARL leverages only the model output, enabling efficient and low-cost deployment. Unlike conventional methods, GARL does not rely on explicit new topology information. Instead, it constructs the DSSE model for the new topology through residual learning, streamlining the transfer process without requiring detailed topology knowledge.

D. Implementation of Proposed GSJKN Method and GARL-based Transfer Method

The Algorithm SA1 of Supplementary Material A illustrates the training and deployment process of the proposed GSJKN method, DAE-based detector, and GARL-based transfer method. The inputs for training are the historical dataset $\{z_i, x_i\}_{i=1}^{T}$, the known topology T, and the limited dataset $\{z_{Ni}, x_{Ni}\}_{i=1}^{S}$ at the new topology. The outputs are $f(\cdot)$ for T, $\varphi(\cdot)$ and $\rho(\cdot)$, and a transferred DSSE model $h_N(\cdot) + f(\cdot)$ for the new topology. After the training process, the parameters in $f(\cdot)$ and $h_N(\cdot)$ are fixed and ready for the real-time DSSE on known and new topologies. The proposed DSSE framework is shown in Algorithm SA1.

IV. CASE STUDY

A. Experiment Setting

1) Data preparation. The IEEE 33-bus test system with

photovoltaic (PV) location, modeled for DSSE tasks, is illustrated in Fig. 2. PV units, i.e., PV1, PV2, and PV3, each with a 600 kW capacity, are installed at buses 6, 13, and 31. Real-time measurements are collected from branch-injected power at branches 1-2, 2-3, 3-4, 4-5, 6-7, 7-8, 8-9, 9-10, and 10-11, while pseudo-measurements involve node load at buses 2-33. To simulate real conditions, 50% uniform noise is added to node load and 1% uniform noise is added to node injection power during state estimation. Load and PV data are sourced from a real DS over one year [31]. New topology scenarios are generated by modifying switch configurations of the original topology. The datasets include 1000 samples for training and 200 for testing on known topologies, with an additional 48 samples for training on new topologies.
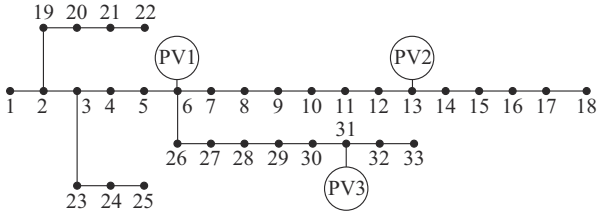

Fig. 2.   IEEE 33-bus test system with PV location.

2) The mean absolute error (MAE) is used for the performance evaluation of deterministic DSSE results.

$$MAE = \frac{1}{V} \sum_{i=1}^{V} |\hat{y}_i - y_i| \tag{26}$$

where y_i and $\hat{y}_i$ are the actual state and estimated state of the ith sample, respectively; and V is the total number of state variables in the test dataset. For interval estimation, key factors are reliability, sharpness, and calibration [32]. Pinball loss, Winkler loss, prediction interval coverage probability (PICP), and mean prediction interval width (MPIW) [33] are employed to evaluate interval DSSE performance. For the detailed definitions of these metrics, please refer to Supplementary Material A.

3) The GSJKN-based estimator for the known topology is constructed using the graph attention with 2 attention heads. The hyper-parameters of the DSSE model are presented in Table I. For the process of hyper-parameter selection, please refer to Table SBI of Supplementary Material B.

TABLE I
HYPER-PARAMETER OF DSSE MODEL

Layer type	Layer task	Layer parameter
Input layer	Accepting measurement	
Fully connected layer (FCL)	Node feature embedding	(4, 16)×2
FCL	Attention calculation	(32, 1)×2
Graph	Node information aggregation	(16, 16)×2
RNN (forward)	Global scanning	(256, 256)
RNN (reversed)	Node feature embedding	(256, 256)
MLP	Node feature embedding	(512, 2)
Output layer	Estimation result	

B. Robustness Test in IEEE 33-bus Test System

To evaluate the performance of the proposed DSSE model under missing or noisy data conditions, six cases simulating typical data acquisition errors are conducted.

1) Case 1: available real-time measurements are collected correctly (normal condition).

2) Case 2: randomly selecting 2 real-time measurements and adding 30% uniform noise.

3) Case 3: randomly selecting 5 real-time measurements and adding 30% uniform noise.

4) Case 4: missing the real-time measurements at branch 1-2 while randomly selecting 2 real-time measurements and adding 30% uniform noise.

5) Case 5: missing the real-time measurements at branches 2-3 and 7-8 while randomly selecting 2 real-time measurements and adding 30% uniform noise.

6) Case 6: missing the real-time measurements at branches 1-2, 7-8, and 10-11 while randomly selecting 2 real-time measurements and adding 30% uniform noise.

For comparison, we assess three standard learning-based methods: ① back-propagation network (BPN) methods [16] with four FCLs; ② CNN [34] with three 1D-CNN layers and two FCLs; and ③ GP [17] with an exponential kernel function. For the latest physics-guided method, we consider physics-aware neural network (PAWNN) [20] with six graph layers; PAWNN with two FCLs (PAMLP) for global information extraction; and GCN using initial residual and identity (GCNII) mapping [21] with 20 graph layers. For WLS-based methods, we investigate WLS-R, which incorporates bad data detection to eliminate erroneous data from iterations; and WLS-L, which excludes bad data detection. Missing data in WLS-L, BPN, CNN, and the proposed methods are replaced with 1×10^{-5}. The learning-based methods are trained with 1000 samples and 5000 fixed epochs with a learning rate of 3×10^{-4}.

Tables II and III summarize the voltage magnitude and angle errors in the original topology, respectively, where MAX denotes the value of the maximum error. The BPN, CNN, and GP perform well with accurate real-time measurements, whereas the WLS shows relatively high MAE for voltage angles due to noisy pseudo-measurements. Physics-guided methods, such as PAWNN and GCNII, yield even larger estimation errors, reflecting the impact of limited measurements. Adding global FCLs to PAWNN improves accuracy, showing the value of global information in addressing measurement scarcity. The proposed GSJKN method further enhances precision through jumping knowledge connections and a tailored global scanning module, showing its effectiveness in handling limited measurements.

In Cases 2-6, where measurements are anomalous, BPN and CNN show notable performance degradation, underscoring the limitations of traditional learning-based DSSE methods in addressing missing real-time data due to a lack of physical insights. Although PAMLP achieves feasible accuracy with normal noise, it exhibits sharper performance declines compared with other physics-guided methods, revealing challenges in directly integrating global information. In contrast, the proposed GSJKN method leverages structural insights and a global scanning module to effectively fill

missing values using adjacent data, sustaining accuracy even with three missing nodes in Case 6. These results demonstrate the robustness of the proposed GSJKN method in managing anomalous data.

TABLE II
VOLTAGE MAGNITUDE ERRORS IN ORIGINAL TOPOLOGY

Method	Voltage magnitude error (10^{-4} p.u.)					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
WLS-R (MAE)	3.73	8.48	12.00	8.75	8.72	9.65
WLS-R (MAX)	63.00	113.00	146.00	97.40	119.00	101.60
WLS-L (MAE)	3.73	8.48	12.00	39.00	75.50	49.20
WLS-L (MAX)	63.00	113.00	146.00	407.00	316.00	362.00
BPN [16] (MAE)	4.29	11.70	25.30	42.30	190.60	77.90
BPN [16] (MAX)	52.00	93.20	248.00	331.00	1021.00	463.00
CNN [34] (MAE)	9.29	15.60	20.90	30.30	33.60	38.20
CNN [34] (MAX)	95.40	129.00	142.00	178.00	222.00	289.00
GP [17] (MAE)	2.88	8.93	14.60	31.30	60.20	37.20
GP [17] (MAX)	67.40	178.00	187.00	183.00	258.00	232.00
PAWNN [20] (MAE)	20.00	22.00	24.20	24.90	51.00	34.50
PAWNN [20] (MAX)	413.00	407.00	365.00	407.00	407.00	453.00
PAMLP [20] (MAE)	4.00	6.61	10.60	14.60	67.90	27.20
PAMLP [20] (MAX)	57.40	139.00	221.00	172.00	222.00	184.00
GCNII [21] (MAE)	18.50	19.90	22.30	25.20	33.50	34.50
GCNII [21] (MAX)	167.00	197.00	269.00	455.00	252.00	453.00
GSJKN (MAE)	3.35	6.25	9.50	7.72	7.84	8.84
GSJKN (MAX)	70.20	82.90	98.50	96.20	93.40	88.60

For the WLS-based method, undetected missing measurements cause significant accuracy degradation in WLS-L. Although WLS-R, with missing data detection, achieves similar accuracy to the proposed WLS-based method for voltage magnitudes, its voltage angle results are notably worse in Case 6. Additionally, WLS-based method relies on precise line parameters, which are often difficult to obtain. The proposed GSJKN method alleviates this dependency by directly learning the regression relationship between measurements and state variables.

For more tests of the proposed GSJKN method under various noise conditions and fewer real-time measurements, please refer to Supplementary Material C Tables SCI-SCIII, respectively. For ablation studies of the proposed GSJKN method, please refer to Table SCIV of Supplementary Material C.

TABLE III
VOLTAGE ANGLE ERRORS IN ORIGINAL TOPOLOGY

Method	Voltage angle error (10^{-3} degree)					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
WLS-R (MAE)	18.40	37.70	55.60	37.20	39.90	41.30
WLS-R (MAX)	343.00	463.00	690.00	648.00	603.00	469.00
WLS-L (MAE)	18.40	37.70	55.60	70.80	96.30	88.60
WLS-L (MAX)	343.00	463.00	690.00	951.00	773.00	930.00
BPN [16] (MAE)	9.49	58.50	66.00	62.50	121.00	87.60
BPN [16] (MAX)	190.00	469.00	391.00	950.00	1437.00	1135.00
CNN [34] (MAE)	11.20	35.00	54.40	42.10	49.50	58.10
CNN [34] (MAX)	149.00	274.00	454.00	365.00	495.00	490.00
GP [17] (MAE)	11.80	20.90	43.40	67.00	105.00	82.20
GP [17] (MAX)	235.00	529.00	1026.00	467.00	561.00	483.00
PAWNN [20] (MAE)	39.30	45.30	58.30	47.20	74.30	67.50
PAWNN [20] (MAX)	863.00	1017.00	1102.00	1017.00	1017.00	1017.00
PAMLP [20] (MAE)	9.10	21.70	43.90	36.00	66.20	45.20
PAMLP [20] (MAX)	183.00	538.00	1042.00	633.00	503.00	612.00
GCNII [21] (MAE)	18.00	26.40	40.30	32.60	39.80	50.70
GCNII [21] (MAX)	322.00	649.00	698.00	659.00	889.00	806.00
GSJKN (MAE)	9.43	19.10	37.70	22.10	22.30	24.40
GSJKN (MAX)	209.00	407.00	676.00	419.00	475.00	429.00

C. Test of Topology Change Detection Based on DAE

Additional tests are conducted to assess the topology change detection capability of the proposed DAE-based detector. Four scenarios are considered, where the network topology transitions from the original condition T1 to new configurations NT1, NT2, NT3, and NT4 as follows.

1) NT1: opening switches of branches 7-8, 28-29, and 14-15 while closing switches of branches 21-8, 25-29, and 9-15.

2) NT2: opening switches of branches 7-8 and 11-12 while closing switches of branches 21-8 and 22-12.

3) NT3: opening switches of branches 7-8, 28-29, and 11-12 while closing switches of branches 21-8, 25-29, and 12-22.

4) NT4: opening switches of branches 11-12 and 28-29 while closing switches of branches 22-12 and 25-29.

Unplanned switch actions causing these topology changes occur at the 100th hour. Figure 3 shows the reconstruction error of the DAE before and after topology changes for NT1-NT4. The reconstruction error remains low before the 100th

hour. When a topology change occurs at the 100th hour, the regression relationship between measurement data and state variables shifts, making it challenging for DAE to accurately reconstruct measurements from input features. This results in a noticeable increase in reconstruction error and enables the detection of network topology changes.
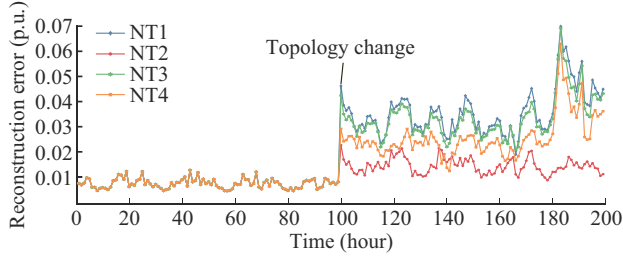

Fig. 3. Reconstruction error of DAE before and after topology changes for NT1-NT4.

To further evaluate the detection capability of the proposed DAE-based detector, four substation configurations, are tested as follows.

1) SC1: moving PV3 and PV2 to bus 31.

2) SC2: adding static var compensators at bus 29 with the capacity of 5 kvar.

3) SC3: cutting off the load at buses 28 and 29.

4) SC4: removing PV3 and cutting off the load at bus 29.

Figure 4 shows the variations in reconstruction error with the original condition T1 for SC1-SC4. When configuration changes occur, reconstruction errors deviate notably from the original curve, reflecting the configuration change of each specific substation. This allows operators to identify alterations in substation configurations, showing the effectiveness of the proposed DAE-based detector. For tests of the DAE-based detector under continuous topology changes, please refer to Fig. SC1 of Supplementary Material C.
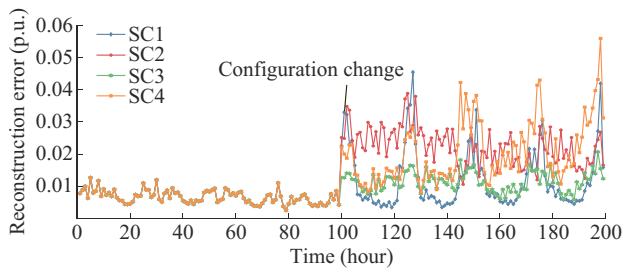

Fig. 4. Variations in reconstruction error with original condition T1 for SC1-SC4.

D. Fast Transfer of DSSE Model on New Topology

To evaluate the effectiveness of the proposed GARL-based transfer method in handling topology changes, four new topologies are considered. The base model is the DSSE model of the original topology, and the GARL-based transfer method is trained on 48 sample sets. Additionally, for the typical NN-based method, we explore BPN [16] and CNN [34] methods trained on 48 new-topology samples, and BPN-centralized training (CT) and CNN-CT methods trained on combined datasets of original and new topologies. For the ML-based methods suitable for small sample learning, we

explore GP [17] and ensembled extreme learning machine (EELM) [35] methods using only 48 sample sets and GP-CT method combining original and new topology data. For the transfer learning-based methods, we explore the BPN-Finetuned method using the original topology model as initial parameters; the Bayesian mean regression (BAR) [27] method with estimation outcomes from five topologies; the DNN+ [36] method aggregating historical data from five topologies; and the CDAR method [37], which measures conditional distribution discrepancies between original and target topologies.

For the WLS-based methods, we investigate the WLS (right) method with topology identification, using correct new topology structures; and the WLS (error) method without topology identification, using incorrect structures with one switch error. The BPN-Finetuned is trained for 200 epochs with a learning rate of 1×10^{-4}.

Table IV shows that when trained only on limited data from new topologies, the estimation errors for BPN, CNN, EELM, and GP methods are high. Traditional ML methods, including small-sample learning techniques like GP and EELM, fail to accurately estimate states for new topologies from sparse training samples. Voltage magnitude estimation is improved by aggregating data from both the original and new topologies, demonstrating that the original topology data support the training of DSSE model. However, directly combining data from different topologies can hinder model accuracy, as indicated by the MAE of CNN-CT voltage angle for NT3 and NT4. BAR and DNN+ methods outperform standard ML methods by leveraging Bayesian methods and additional information, and their accuracy remains limited due to significant differences between the new and historical topologies. Instead of augmenting training data or simultaneously learning tasks for multiple topologies, the proposed GARL-based transfer method models the estimation residual of the pre-trained GSJKN with a GP and uses a composite kernel to transfer knowledge between topologies. This enables the proposed GARL-based transfer method to outperform BAR and DNN+ with less information (using only original topology data) for training. When the correct structure of new topologies is unknown and the WLS-based method applies the original topology, estimation errors are high. Although the performance of WLS (right) improves when the correct structure is known, acquiring precise topology information frequently poses a challenge in practical scenarios, highlighting the limitations of optimization-based DSSE methods. For more tests about the time costs of various methods, please refer to Table SCV of Supplementary Material C.

E. Probabilistic DSSE Results on New Topology

The proposed GARL-based transfer method converts deterministic DSSE results into probabilistic estimates via the GP-based transfer process. Using NT3 as a test case, we assess estimation intervals with the proposed GARL-based transfer method trained on 48 samples. Additionally, for the typical non-parametric method, we explore GP [17] methods trained on 48 new-topology samples; and GP-CT method combining original and new-topology data.

TABLE IV
TRANSFER RESULTS IN NT1-NT4

Method	Voltage magnitude error (10^{-4} p.u.)				Voltage angle error (10^{-3} degree)			
	NT1	NT2	NT3	NT4	NT1	NT2	NT3	NT4
BPN [16]	37.10	35.90	37.40	32.00	33.60	48.30	55.80	54.00
CNN [34]	41.10	31.00	30.30	31.50	27.10	33.70	26.80	32.20
GP [17]	20.20	21.10	20.30	19.10	61.00	66.00	62.80	57.70
BPN-CT	24.30	22.60	27.90	24.30	30.00	37.00	36.30	37.10
CNN-CT	27.80	23.80	25.20	23.50	25.20	29.40	31.20	39.30
GP-CT	14.80	13.70	15.70	14.00	43.00	41.50	48.10	43.70
WLS (right)	4.96	6.66	6.92	6.05	24.70	34.60	35.40	31.00
WLS (error)	34.00	11.00	40.00	34.00	110.00	55.30	93.60	81.40
EELM [35]	27.80	33.50	27.10	26.40	32.70	37.40	36.80	41.90
BPN-Finetuned [34]	20.90	17.70	20.70	18.90	28.10	30.90	34.30	32.10
CDAR [37]	24.90	20.40	25.50	19.90	21.80	26.80	31.20	31.50
DNN+ [36]	13.70	17.60	14.90	13.10	18.00	34.90	29.40	29.50
BAR [27]	11.30	9.95	11.90	8.99	16.80	33.70	30.50	22.00
Proposed	4.22	4.31	4.84	4.23	9.94	13.80	12.70	12.10

For the NN-based method, we investigate the quantile regression neural network (QRNN) -CT method, using the same data as GP-CT. For the transfer learning-based methods, we explore BAR [27] method and the QRNN-Finetuned [34] method which is finetuned from the model of original topology.

Tables V presents the probabilistic results on voltage magnitude in NT1 under voltage magnitude error of 1×10^{-4} p.u.. Table VI presents the probabilistic results on voltage angle in NT1 under voltage angle error of 1×10^{-3} degree. When trained solely on NT3 data, the GP method exhibits high Pinball and Winkler losses. The magnitude estimation performance of GP is improved by adding historical topology data. While the BAR method, which leverages historical topology information, performs better than NN-based methods like QRNN-CT and QRNN-Finetuned, which cannot meet the standard of the GARL-based transfer method. The proposed GARL-based transfer method, using a GP method with a composite kernel for knowledge transfer, adapts effectively to significant topology changes and captures uncertainties from sparse online measurements. For example, the proposed GARL-based transfer method reduces Pinball loss by 61.7% and 71.8% for voltage magnitude compared with BAR and QRNN-Finetuned, respectively, showing a substantial advantage in Winkler loss. When $\alpha = 0.1$, the interval score of the proposed GARL-based transfer method outperforms more than threefold compared with BAR in voltage magnitude estimation, demonstrating its superior ability to generate effective estimation intervals and quantify uncertainties.

Tables VII and VIII present the interval results of voltage magnitude and voltage angle in NT1, respectively, where better performance is indicated by a smaller MPIW and a larger PICP. The QRNN-Finetuned method is unreliable, with PICP values below 0.8 for both magnitude and angle tasks when $\alpha = 0.2$.

TABLE V
PROBABILISTIC RESULTS ON VOLTAGE MAGNITUDE IN NT1

Method	Pinball loss	Winkler loss			
		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$
Proposed	1.89	32.7	25.6	21.6	18.8
BAR [27]	4.94	116.0	76.7	59.9	49.8
GP [17]	8.22	154.0	117.0	98.0	84.4
GP-CT	6.65	162.0	105.0	81.5	67.4
QRNN-CT	7.83	157.0	125.0	101.0	85.5
QRNN-Finetuned	6.70	196.0	121.0	100.0	81.0

TABLE VI
PROBABILISTIC RESULTS ON VOLTAGE ANGLE IN NT1

Method	Pinball loss	Winkler loss			
		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$
Proposed	5.16	96.7	74.0	61.7	53.0
BAR [27]	13.30	379.0	225.0	166.0	134.0
GP [17]	26.20	499.0	380.0	316.0	271.0
GP-CT	20.30	498.0	323.0	250.0	206.0
QRNN-CT	21.40	413.0	314.0	258.0	220.0
QRNN-Finetuned	16.40	446.0	2463.0	194.0	164.0

Although the GP achieves better coverage with limited samples, its intervals for magnitude and angle estimates remain wide due to insufficient training data. For example, MPIW of GP is 276.9% larger than that of the proposed GARL-based transfer method for the voltage magnitude task when $\alpha = 0.1$. Adding training data and training tasks simultaneously reduces MPIW for GP-CT, but naive aggregation across topologies hinders the performance, as can be observed in PICP values of GP-CT. While the BAR achieves narrower intervals, its PICP does not satisfy DSSE requirements. In contrast, the proposed GARL-based transfer meth-

od adapts to new topologies with limited data, while its Bayesian characteristics enable uncertainty quantification, achieving superior coverage with narrower intervals. For more displays about the probabilistic results of various methods, please refer to Fig. SC2 of Supplementary Material C.

TABLE VII
INTERVAL RESULTS ON VOLTAGE MAGNITUDE IN NT1

Method	PICP (%)		MPIW (10^{-4} p.u.)	
	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
Proposed	0.9045	0.8448	22.73	17.72
BAR [27]	0.7034	0.6386	29.03	22.61
GP [17]	0.8712	0.8171	85.66	66.74
GP-CT	0.7066	0.6384	36.87	28.73
QRNN-CT	0.7827	0.6428	79.71	60.24
QRNN-Finetuned	0.6324	0.5221	52.55	34.27

TABLE VIII
INTERVAL RESULTS ON VOLTAGE ANGLE IN NT1

Method	PICP (%)		MPIW (10^{-3} degree)	
	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.2$
Proposed	0.8757	0.8168	58.67	45.71
BAR [27]	0.5974	0.5324	46.83	36.49
GP [17]	0.8703	0.8106	280.20	218.30
GP-CT	0.7395	0.9375	120.30	93.75
QRNN-CT	0.7248	0.6134	140.00	103.20
QRNN-Finetuned	0.6028	0.4890	77.47	59.15

F. Scalability Test in IEEE 119-bus Test System

To evaluate the adaptability of the proposed GARL-based transfer metthod, more tests are carried out on the IEEE 119-bus test system [38]. The PV units are placed at buses 22, 50, 74, 80, 96, and 110, each with a capacity of 400 kW. The real-time measurements include the injection power of branches 10-11, 18-19, 19-20, 20-21, 21-22, 22-23, 23-24, 28-29, 29-30, 30-31, 31-32, 32-33, 33-34, 34-35, 63-64, 64-65, 65-66, 66-67, 67-68, 68-69, 69-70, 101-102, 102-103, 103-104, 104-105, 105-106, 106-107, and 107-108, and added with 1% uniform noise. The load data are taken as pseudo measurement data and added with 50% uniform noise.

Moreover, different topologies are performed as follows, where HT1 is selected as the original topology for GSJKN method. HNT1 and HNT2 are selected as the new topologies.

1) HT1: original topology.

2) HNT1: opening switches of branches 34-35, 72-73, and 107-108 while closing switches of branches 25-35, 91-73, and 83-108.

3) HNT2: opening switches of branches 23-24, 34-35, 72-73, and 107-108 while closing switches of branches 8-24, 25-35, 91-73, and 83-108.

Six cases are considered to simulate data acquisition errors as follows.

1) Case 1: normal condition.

2) Case 2: randomly selecting 20% real-time measurements and adding 30% uniform noise.

3) Case 3: randomly selecting 50% real-time measurements and adding 30% uniform noise.

4) Case 4: missing the real-time measurements at branches 18-19, 29-30, 64-65, and 102-103 while randomly selecting 20% real-time measurements and adding 30% uniform noise.

5) Case 5: missing the real-time measurements at branches 18-19, 22-23, 29-30, 33-34, 64-65, 68-69, 102-103, and 106-107 while randomly selecting 20% real-time measurements and adding 30% uniform noise.

6) Case 6: missing the real-time measurements at branches 18-19, 21-22, 23-24, 29-30, 32-33, 34-35, 64-65, 67-68, 69-70, 102-103, 105-106, and 107-108 while randomly selecting 20% real-time measurements and adding 30% uniform noise.

Table IX shows the estimation errors for voltage magnitude and angle in HT1. The classical BPN method struggles with anomalous measurements, leading to significant estimation deviations, underscoring the limitations of standard learning-based methods. While GCNII and PAWNN provide some resilience to abnormal measurements, their precision remains limited due to sparse real-time data. In contrast, the proposed GSJKN method consistently performs well under both normal and anomalous conditions, aligning with results from the IEEE 33-bus test system. These findings highlight the high precision and reliability of the proposed GSJKN method, even with noisy or incomplete measurements.

TABLE IX
ESTIMATION ERRORS FOR VOLTAGE MAGNITUDE AND ANGLE IN HT1

Method	Voltage magnitude error (10^{-4} p.u.)						Voltage angle error (10^{-3} degrees)					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
WLS-L	3.10	4.79	6.10	26.90	30.50	34.0	8.10	15.6	21.2	40.4	43.8	51.3
BPN [16]	5.60	7.13	9.22	29.60	34.90	38.3	5.23	12.6	23.2	53.3	60.2	61.3
GCNII [21]	22.20	24.70	25.20	25.70	26.60	27.5	19.50	21.1	22.4	30.3	33.4	35.2
PAWNN [20]	15.80	16.70	17.80	25.60	27.20	30.0	12.10	15.9	21.2	39.9	42.7	51.0
GSJKN	3.80	4.52	5.97	8.13	9.02	10.1	7.51	11.3	16.9	15.5	16.4	17.9

Transfer tests in HNT1 and HNT2 are conducted to evaluate the effectiveness of the proposed GARL-based transfer method. Table X shows performance results across methods

in HNT1 and HNT2. The WLS experiences substantial estimation errors following topology changes, highlighting the need for prior topology information in optimization-based

methods. Learning-based methods, such as BPN, struggle to build a reliable DSSE mapping with limited new-topology samples, resulting in large estimation deviations. Although adding information from other topologies improves the accuracy of BPN-CT, BPN-Finetuned, and BAR, their performance remains below practical standards, with BAR showing large estimation errors due to significant deviations in transferred knowledge. In contrast, the proposed GARL-based transfer method significantly outperforms other methods, showing the superiority of residual learning.

TABLE X
PERFORMANCE RESULTS ACROSS METHODS IN HNT1 AND HNT2

Method	Voltage magnitude error (10^{-4} p.u.)		Voltage angle error (10^{-3} degree)	
	HNT1	HNT2	HNT1	HNT2
BPN [16]	73.40	92.10	34.00	29.70
BPN-CT	40.90	56.00	26.10	24.70
BPN-Finetuned [34]	39.60	34.70	27.40	24.20
BAR [27]	22.30	19.90	29.40	28.20
WLS (right)	3.76	3.08	14.60	12.20
WLS (error)	28.60	17.00	83.70	51.00
Proposed	9.13	8.52	23.50	21.40

For tests of the proposed GARL-based transfer method at the post-transfer stage, please refer to Table SCVI of Supplementary Material C.

G. Scalability Test on IEEE 342-node Test System

To evaluate the adaptability of the proposed GSJKN method and GARL-based transfer method, tests are conducted on the IEEE 342-node test system, which represenets low-voltage, three-phase unbalanced networks widely used in North America [39]. The system includes 48 PV units (300 kW each), with real-time current measurements from 56 branches collected with 1% uniform noise. One year of recorded load and PV generation data [31] is used, and 50% uniform noise is added to the load data to simulate measurements. Moreover, different topologies are performed as follows.

GT1 is selected as the original topology for the proposed GSJKN method, and GNT1 and GNT2 are selected as the new topologies.

1) GT1: original topology.
2) GNT1: opening switches of branches P134-135 and S148-S27.
3) GNT2: opening switches of branches S148-S27.

Six cases are simulated to replicate typical data acquisition errors as follows.

1) Case 1: normal condition.
2) Case 2: randomly selecting 20% real-time measurements and adding 30% uniform noise.
3) Case 3: randomly selecting 50% real-time measurements and adding 30% uniform noise.
4) Case 4: missing the real-time measurements at branches P82 and P83.
5) Case 5: missing the real-time measurements at branches P122, P123, S52, and S53.
6) Case 6: missing the real-time measurements at branches P122, P123, S82, S83, S202, and S203.

Table XI presents estimation errors for voltage magnitude and angle across methods in GT1. The classical BPN method struggles to capture the complex mapping between measurement data and state variables in unbalanced systems, leading to high MAE values for both tasks. While CNN shows improved precision over BPN, its voltage magnitude error remains relatively high. Due to limited real-time measurements and the large scale of the system, the GCNII and PAWNN methods also fail to achieve accurate angle estimates. In contrast, the proposed GSJKN method demonstrates superior performance on both tasks under normal conditions. Under noisy/missing measurements, the accuracy of CNN declines significantly, as can be observed in voltage angle errors in Case 3 and voltage magnitude errors in Case 6, showing the limitations of classical learning-based methods. By embedding physical structure, the proposed GSJKN method maintains accuracy even with outliers. Consistent with results from the IEEE 33-bus test system, these results demonstrate the robustness of the proposed GSJKN method.

TABLE XI
ESTIMATION ERRORS FOR VOLTAGE MAGNITUDE AND ANGLE ACROSS METHODS IN GT1

Method	Voltage magnitude error (10^{-4} p.u.)						Voltage angle error (10^{-3} degrees)					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
BPN [16]	37.70	37.90	41.00	39.00	36.80	40.50	96.10	100.00	104.00	110.00	99.90	172.00
CNN [34]	20.80	23.70	27.20	223.00	180.00	193.00	48.50	76.50	99.90	84.70	67.20	154.00
GCNII [21]	34.10	34.60	35.10	36.30	39.20	44.80	139.00	141.00	142.00	144.00	151.00	159.00
PAWNN [20]	30.60	31.20	32.00	38.40	39.30	41.60	151.00	154.00	162.00	182.00	231.00	236.00
GSJKN	8.61	9.63	10.80	12.40	12.70	16.60	48.10	51.40	57.50	66.80	58.50	78.70

In Table XII, the transfer results of new topologies GNT1 and GNT2 are presented. With limited samples, the BPN and CNN methods struggle to map measurement data to system state variables, resulting in high MAEs for both voltage magnitude and angle. Aggregating data and training simultaneously improve the model precision; however, the BAR

method still shows significant voltage angle errors due to task complexity. By contrast, the proposed GARL-based transfer method first applies a GSJKN model to learn the initial regression rule, followed by residual learning for fast adaptation to topology changes. This contributes to superior performance in estimation of voltage magnitude and angle,

underscoring the effectiveness of the proposed GARL-based transfer method.

TABLE XII
TRANSFER RESULTS OF NEW TOPOLOGIES IN IEEE 342-NODE TEST SYSTEM

Method	Voltage magnitude error (10^{-4} p.u.)		Voltage angle error (10^{-3} degrees)	
	GNT1	GNT2	GNT1	GNT2
BPN [16]	1871.0	1664.0	155.0	128.0
CNN [34]	1363.0	821.0	110.0	120.0
BPN-CT	45.1	64.2	116.0	124.0
CNN-CT	27.1	32.8	73.6	72.6
BPN-Finetuned [34]	34.2	29.8	115.0	104.0
BAR [27]	16.9	16.2	133.0	199.0
Proposed	14.4	13.4	65.9	61.1

For probabilistic results of various methods at the IEEE 342-node test system, please refer to Table SCVII and Table SCVIII of Supplementary Material C.

V. CONCLUSION

We introduce a robust DSSE method based on a physics-guided GSJKN method, a DAE-based detector and a GARL-based transfer method, aiming at tackling anomalous real-time measurements and potential topology changes, respectively. Specifically, the proposed GSJKN method establishes a complex mapping between measurement data and system state variables, followed by a DAE-based detector to detect topology changes and a GARL-based transfer method to capture residuals after topology changes occur. Comparative tests with benchmark methods show that: embedding physical structural information within the GSJKN improves robustness against missing/noisy measurements; the DAE-based detector monitors topology changes online by tracking reconstruction errors; the GARL-based transfer method enables rapid adaptation to new topologies with minimal online data and effectively quantifies estimation uncertainty, producing probabilistic DSSE results with higher reliability, sharpness, and resolution than other methods.

As system scale and measurement diversity increase, fusing multi-rate multi-sensor data becomes critical for enhancing state estimation precision and efficiency. Future research will focus on developing a multi-source information fusion module to effectively integrate diverse measurement data into the estimation process. Additionally, more complex node-breaker substation models will be incorporated into the topology change detector to enhance the monitoring of substation configurations. Advanced semi-supervised and meta-learning frameworks will also be explored to reduce model dependency on extensive training data, broadening the applicability of learning-based DSSE methods.

REFERENCES

[1] Z. Liu, P. Li, C. Wang et al., "Robust state estimation of active distribution networks with multi-source measurements," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 5, pp. 1540-1552, Sept.

2023.

[2] V. Gundu, S. P. Simon, V. Kasi et al., "Priority-based residential demand response for alleviating crowding in distribution systems," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 2, pp. 502-510, Mar. 2023.

[3] W. Wang and N. Yu, "Estimate three-phase distribution line parameters with physics-informed graphical learning method," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3577-3591, Sept. 2022.

[4] J. Hu, W. Hu, D. Cao et al., "Robust multiarea distribution system state estimation based on structure-informed graphic network and multitask Gaussian process," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10599-10612, Aug. 2024.

[5] D. Cao, J. Zhao, J. Hu et al., "Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 233-246, Jan. 2024.

[6] A. Primadianto and C. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875-3883, Sept. 2017.

[7] X. Zhou, Z. Liu, Y. Guo et al., "Gradient-based multi-area distribution system state estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5325-5338, Nov. 2020.

[8] G. Cheng, Y. Lin, Y. Chen et al., "Adaptive state estimation for power systems measured by PMUs with unknown and time-varying error statistics," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4482-4491, Sept. 2021.

[9] B. Rout, S. Dahale, and B. Natarajan, "Dynamic matrix completion based state estimation in distribution grids," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7504-7511, Nov. 2022.

[10] M. Mao, J. Xu, Z. Wu et al., "A multiarea state estimation for distribution networks under mixed measurement environment," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3620-3629, Jun. 2022.

[11] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Enhanced tensor completion based approaches for state estimation in distribution systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5938-5947, Sept. 2021.

[12] P. A. Pegoraro and S. Sulis, "Robustness-oriented meter placement for distribution system state estimation in presence of network parameter uncertainty," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 954-962, May 2013.

[13] T. Chen, L. Sun, K. Ling et al., "Robust power system state estimation using t-distribution noise model," *IEEE Systems Journal*, vol. 14, no. 1, pp. 771-781, Mar. 2020.

[14] T. Wu, W. Xue, H. Wang et al., "Extreme learning machine-based state reconstruction for automatic attack filtering in cyber physical power system," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1892-1904, Mar. 2021.

[15] M. Netto and L. Mili, "A robust data-driven Koopman Kalman filter for power systems dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7228-7237, Nov. 2018.

[16] B. Zargar, A. Angioni, F. Ponci et al., "Multiarea parallel data-driven three-phase distribution system state estimation using synchrophasor measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6186-6202, Sept. 2020.

[17] D. Cao, J. Zhao, W. Hu et al., "Topology change aware data-driven probabilistic distribution state estimation based on Gaussian process," *IEEE Transactions on Smart Grid*, vol. 14, no. 2, pp. 1317-1320, Mar. 2023.

[18] G. Tian, Y. Gu, D. Shi et al., "Neural-network-based power system state estimation with extended observability," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1043-1053, Sept. 2021.

[19] Y. Chen, H. Chen, Y. Jiao et al., "Data-driven robust state estimation through off-line learning and on-line matching," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 897-909, Jul. 2021.

[20] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347-4356, Nov. 2020.

[21] T. Su, J. Zhao, Y. Pei et al., "Probabilistic physics-informed graph convolutional network for active distribution system voltage prediction," *IEEE Transactions on Power Systems*, vol. 38, no. 6, pp. 5969-5972, Nov. 2023.

[22] J. Hu, W. Hu, D. Cao et al., "Feature graph-enabled graphical learning for robust DSSE with inaccurate topology information," *IEEE Transactions on Power Systems*, vol. 39, no. 4, pp. 6091-6094, Jul. 2024.

[23] H. Wu, Z. Xu, and M. Wang, "Unrolled spatiotemporal graph convolu-

tional network for distribution system state estimation and forecasting," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 297-308, Jan. 2023.

[24] B. Huang and J. Wang, "Applications of physics-informed neural networks in power systems-a review," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 572-588, Jan. 2023.

[25] L. Wang, Q. Zhou, and S. Jin, "Physics-guided deep learning for power system state estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 607-615, Jul. 2020.

[26] J. Hu, W. Hu, J. Chen *et al.*, "Fault location and classification for distribution systems based on deep graph learning methods," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 35-51, Jan. 2023.

[27] D. Cao, J. Zhao, W. Hu *et al.*, "Physics-informed graphical learning and Bayesian averaging for robust distribution state estimation," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 2879-2892, Mar. 2024.

[28] A. Abur, H. Kim, and M. K. Celik, "Identifying the unknown circuit breaker statuses in power networks," *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 2029-2037, Nov. 1995.

[29] A. de la V. Jaen and A. Gomez-Exposito, "Implicitly constrained substation model for state estimation," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 850-856, Aug. 2002.

[30] X. Qiu, E. Meyerson, and R. Miikkulainen, "Quantifying point-prediction uncertainty in neural networks via residual estimation with an I/O kernel," in *Proceeding of The International Conference on Learning Representations (ICLR)*, online, Jun. 2020, pp. 1-35.

[31] K. Dehghanpour and Z. Wang. (2019, Jan.). A Real 240-Bus Distribution System with One-Year Smart Meter Data. [Online]. Available: http://wzy.ece.iastate.edu/Testsystem.html.

[32] J. Hu, W. Hu, D. Cao *et al.*, "Probabilistic net load forecasting based on transformer network and Gaussian process-enabled residual modeling learning method," *Renewable Energy*, vol. 225, pp. 120253, Apr. 2024.

[33] J. Hu, W. Hu, D. Cao *et al.*, "Bayesian averaging-enabled transfer learning method for probabilistic wind power forecasting of newly built wind farms," *Applied Energy*, vol. 355, p. 122185, Aug. 2024.

[34] H. Zang, Y. Guo, M. Huang *et al.*, "State estimation for power systems with time-varying topology based on deep transfer learning," *Automation of Electric Power Systems*, vol. 45, no. 24, pp. 49-56, Jan. 2021.

[35] Y. Xu, Z. Dong, J. Zhao *et al.*, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1253-1263, Aug. 2012.

[36] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83-95, Jan. 2016.

[37] X. Liu, Y. Li, Qi. Meng *et al.*, "Deep transfer learning for conditional shift in regression," *Knowledge-Based Systems*, vol. 227, p. 107216, Jan. 2021.

[38] H. Wang, W. Zhang, and Y. Liu, "A robust measurement placement method for active distribution system state estimation considering network reconfiguration," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2108-2117, May 2018.

[39] J. A. D. Massignan, J. B. A. London, M. Bessani *et al.*, "Bayesian inference approach for information fusion in distribution system state estimation," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 526-540, Jan. 2022.

Jiaxiang Hu received the B.Eng. and M.E. degrees from the University of Electronics Science and Technology of China (UESTC), Chengdu, China, in 2021 and 2024, respectively. He is currently working toward the Ph.D. degree in electrical engineering at KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include distribution network state estimation, wind power prediction, and machine learning.

Weihao Hu received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in electrical engineering, and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 2012. He is currently a Full Professor and the Director of Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China, Chengdu, China. His research interests include artificial intelligence in modern power system and renewable power generation.

Di Cao received the Ph.D. degree in control science and engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2021. He is currently an Associate Professor at UESTC. His research interests include optimization of distribution network and application of machine learning in power system.

Jianjun Chen received the Ph.D. degree in control science and engineering from the University of Electronics Science and Technology of China (UESTC), Chengdu, China, in 2024. His research interests include mechanical fault diagnosis and application of machine learning in health assessment of power system equipment.

Sayed Abulanwar received the B.S. and M.S. degrees in electrical engineering from Mansoura University, Mansoura, Egypt, in 2005 and 2010, respectively, and the Ph.D. degree from the Energy Technology Department, Aalborg University, Aalborg, Denmark, in 2016. He is currently an Associate Professor of electrical engineering at Mansoura University. His research interests include hybrid AC/DC microgrid, protection of AC and DC system, wind energy conversion system, transients in power system, and grid-connected converter.

Mohammed K. Hassan received the M.Sc. and Ph.D. degrees in computer engineering and systems from Mansoura University, Mansoura, Egypt. He is currently an Assistant Professor with the Mechatronics Department, Faculty of Engineering, Horus University, Damietta, Egypt. His current research interests include artificial intelligence, machine learning, deep learning, cloud computing, Internet of Things, and big data.

Zhe Chen received the B.Eng. and M.Sc. degrees from the Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph.D. degree from the University of Durham, Durham, UK. He is a Full Professor with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. His research interests include power electronics and electric machine, wind energy, and modern power system.

Frede Blaabjerg received the Ph.D. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1995. He was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. His current research interests include power electronics and its applications such as in wind turbine, photovoltaic system, reliability, harmonics and adjustable speed drive.