



Online Power System Dynamic Security Assessment: A GNN-FNO Approach Learning from Multi-Source Spatial-Temporal Data

Genghong Lu, Siqi Bu *Senior Member, IEEE*

Abstract—Data-driven online dynamic security assessment offers system operators a computationally efficient approach for monitoring system dynamics. However, the challenges of processing multi-source spatial-temporal data from different measurement systems remain unsolved, thus resulting in potentially biased results. In addition, most existing data-driven dynamic security assessment methods that focus on state estimation/prediction overlook the fault location identification, which is important to real-time decision-making. To address the above limitations, an advanced online dynamic security assessment, which learns system dynamics and fault characteristics from multi-source spatial-temporal data, is developed. Considering the challenge posed by different sampling rates and sensor numbers, global and local spatial-temporal data from various measurement systems are modeled as graphs with different numbers of nodes and edges. Then, two different sets of graph neural networks are customized to learn global and local spatial-temporal features, respectively. With the learned multi-source spatial-temporal features, a Fourier Neural Operator-based dynamics trajectory predictor and a multilayer perceptron-based fault location identifier are developed for the advanced online dynamic security assessment. Case studies on IEEE 39 Bus System and IEEE 118 Bus System validate the effectiveness and efficiency of the developed online dynamic security assessment.

Index Terms—Dynamics trajectory prediction, fault location identification, multi-source spatial-temporal data, online dynamic security assessment

I. INTRODUCTION

Given the increasing penetration of renewable energy resources (RES), which introduces significant uncertainties and complicates the dynamics of the power system, online dynamic security assessment (DSA) [1] is pivotal for ensuring the secure operation of the power system. To alarm operators in the event of an unsafe condition, system dynamics are monitored.

Manuscript received 29 Oct. 2024; revised 5 May 2024; accepted 25 May 2024. This work was supported by the Hong Kong Research Grant Council for the Research Project under Grant 15205424. (Corresponding author: Siqi Bu).

Genghong Lu is with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong. (e-mail: genghong.lu@polyu.edu.hk).

Siqi Bu is with Department of Electrical and Electronic Engineering, Shenzhen Research Institute, Research Centre for Grid Modernisation, Research Institute for Smart Energy, Policy Research Centre for Innovation and Technology, International Centre of Urban Energy Nexus, and Centre for Advances in Reliability and Safety, The Hong Kong Polytechnic University, Kowloon, Hong Kong. (e-mail: siqi.bu@polyu.edu.hk)

Model-based approaches [2] (e.g., time-domain simulation and direct methods) leverage the mathematical models of power system dynamics to provide accurate results. However, they are too time-consuming to be applied in a real-time manner [3]. To further facilitate computational efficiency, data-driven approaches, such as long short-term memory (LSTM)-based and graph neural networks (GNN)-based approaches [5]–[7], are developed. Although data-driven approaches excel at modeling nonlinear relationships for real-time power system dynamics prediction, their performance is highly sensitive to data processing quality. In particular, the efficient extraction of spatial-temporal features from heterogeneous power system data significantly influences the learning and modeling of dynamic behaviors [4]. A matrix mapping operation is adopted to map the 1-D phasor measurement unit (PMU) data into a 2-D measurement matrix [6]. Then, a CNN module is designed to learn the implicit spatial-temporal features. Another solution is to model a power system with N bus and M transmission lines as a graph with N nodes and M sets of edges [8]. Then, GNNs are employed to process the operation and topology information of the preprocessed graph-structured data.

Although the existing online DSA with spatial-temporal resolution exhibits superior learning performance, the above-mentioned approaches only use single-source data as model inputs. Data from various measurement systems provides more comprehensive system information, which can enhance the situation awareness of the power system dynamic security. However, challenges of implementing multi-source data from different measurement systems, e.g., supervisory control and data acquisition (SCADA) system and wide area measurement system (WAMS), remain unsolved. For example, the SCADA system provides global system measurements using widely deployed remote terminal units at a low sampling rate, while WAMS reports state measurements through PMUs installed on key substations at a high sampling rate. Therefore, different sampling rates and installation schemes pose challenges of integrating and processing hybrid measurements from diverse sources.

Facing the challenge of different sampling rates, a high temporal-spatial resolution using graph convolutional networks (GCNs) is developed [7]. SCADA measurements and PMU measurements at time t are processed independently by two distinct GCNs. Then, the learned features are concatenated to estimate the whole system states at time t . This method provides an efficient solution to integrating measurements

collected at different sampling rates, yet it ignores the internal temporal characteristics of PMU measurements. The dynamic patterns within PMU measurements are critical for predicting post-fault dynamic trajectories and exploring fault information. Therefore, the development of a DSA that can efficiently process multi-source spatial-temporal data and concurrently learn the inherent temporal characteristics of PMU measurements is crucial for achieving comprehensive DSA outcomes.

In addition to processing multi-source spatial-temporal data, learning complex nonlinear system dynamics is another challenge to existing data-driven DSA. An iterative prediction method using LSTM-based recurrent neural network (RNN) is developed for predicting post-fault rotor angle trajectories [5]. This method learns the temporal relationship from the available information about the post-fault system behavior within a certain time window. Although it outputs accurate predictions, the generalization capability of this model is limited. Upon variations in operating conditions, network topology, or system stable conditions, the prediction accuracy of this method diminishes.

Neural operators [9], [10], the powerful learning models that can approximate complex functions and learn the solutions of differential equations, have been developed and employed for power system dynamics trajectory prediction. Deep Operator Network (DeepONet)-based trajectory prediction methods are developed to learn the post-contingency dynamic response of the power grid [11], [12]. A Fourier Neural operator (FNO)-based approach is developed to learn power system dynamics from frequency-domain of measurements [13]. The simulation results show the implementation of neural operators has markedly enhanced the dynamics learning capability of data-driven methods. Despite the powerful dynamics learning capability, these methods can only process single-source data, which limits their ability to identify fault location and consequently, yields inadequate DSA outcomes.

To solve these problems, this paper develops a novel DSA that efficiently processes multi-source spatial-temporal data. Two sets of GNNs are tailored to learn global spatial features from SCADA measurements and local spatial-temporal features from PMU measurements, respectively. The learned spatial-temporal features are used by an FNO-based trajectory predictor and a multilayer perceptron (MLP)-based fault location identifier to achieve the goal of providing comprehensive DSA results. The contributions of this paper are as follows.

1. Considering the single-source data provides limited system information, leading to potentially biased DSA results, a novel online DSA that uses GNNs to learn from multi-source spatial-temporal data, i.e., SCADA measurements and PMU measurements, is developed. Comparison results demonstrate that incorporating multi-source measurements enhances the precision of DSA results.

2. Facing the challenges arising from the disparities in sampling rates and sensor numbers across various measurement systems, a novel spatial-temporal resolution is developed. Various measurement systems are modeled as graphs with different numbers of nodes and edges, so the multi-source spatial-temporal features can be learned by GNNs.

3. To learn various system dynamics trajectories across

diverse operating conditions, a novel dynamics trajectory predictor using FNO, which can learn solutions of differential equations in frequency domain, is developed.

4. In addition to dynamics trajectory prediction, which is the focus of most existing data-driven DSAs, the advanced online DSA developed in this paper can identify fault locations by learning fault characteristics from multi-source spatial-temporal features.

II. PROBLEM FORMULATION

A. Power System Dynamics

The power system dynamics can be described as a set of differential-algebraic equations.

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}_{\xi_f}(\mathbf{x}, \mathbf{y}, t, \xi_p) \\ \mathbf{0} = \mathbf{g}_{\xi_f}(\mathbf{x}, \mathbf{y}, t, \xi_p) \end{cases} \quad (1)$$

where $\mathbf{f}$ indicates the differential equations and $\mathbf{g}$ indicates the algebraic equations. $\mathbf{x}$ and $\mathbf{y}$ are dynamic and algebraic states, respectively. ξ_p indicates the uncertainties from loads, RES, and faults, such as different fault locations. ξ_f indicates different system dynamics stages, i.e., pre-fault, fault-on, and post-fault stages.

Large disturbances, such as three-phase-to-ground faults, can lead to alterations in system dynamics. For instance, the occurrence of a short circuit on the transmission line modifies the susceptance and conductance, leading to changes in the system models across various stages. Apart from the varying system dynamics across different stages, the increasing penetration of RES further complicates the power system dynamics, rendering online DSA a more challenging task.

B. Challenges in Online DSA Application

Although existing DSA methods provide effective solutions, some research gaps remain facing the arising new challenges.

1. *Ability to conduct fast computation for real-time application.* To respond to unexpected faults in a timely manner, it is required for DSA methods to track system dynamics as soon as the fault occurs. Although conventional model-based methods [14], [15] provide accurate DSA results, they are too computationally inefficient for real-time application.

2. *Ability to provide comprehensive system dynamics and fault information to support real-time decision-making of operators.* Most existing DL-based DSAs only perform single tasks such as state estimation [1], [3] or binary security condition classification [16], [17], providing limited system dynamics information. Considering that fault information, such as fault location identification [6], plays an important role in helping operators take countermeasures against potential risks, a DL-based multi-tasking DSA is needed to improve the situation awareness of operators.

3. *Ability to predict system dynamics using multi-source spatial-temporal data.* For real applications, the number of available PMUs is limited, making it difficult to assess the whole system security situation. To address this issue, spatial-temporal solutions using SCADA and PMU measurements have been developed [3], [7]. However, identical graph-based

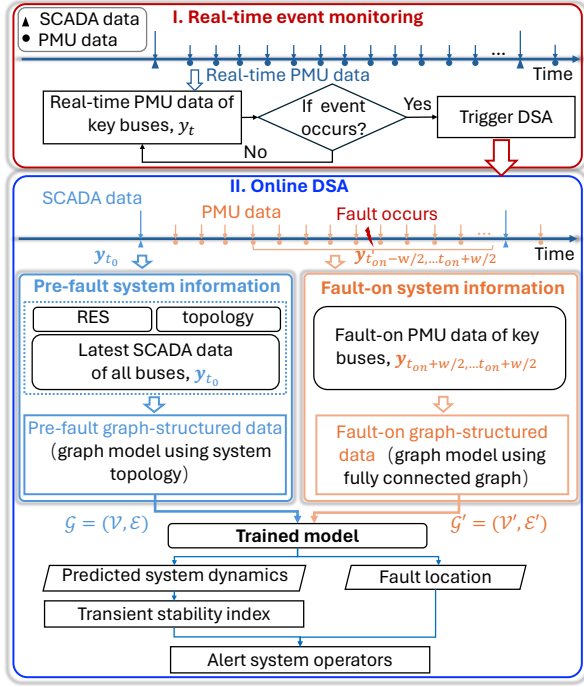


Fig. 1. Online DSA framework.

models have been applied to SCADA and PMU data, disregarding their inherent divergent spatial-temporal characteristics. In addition, the time-varying nature of PMU measurements during the fault-on stage is overlooked, resulting in inadequate capture of fault-related spatial-temporal patterns. Therefore, a method that can extract and utilize the complementary spatial-temporal features is in press.

C. Developed Online DSA and Learning Problems

In the face of the above challenges, this manuscript develops a novel online DSA.

1) Framework Overview: Fig.1 illustrates the flowchart of the online DSA, which adopts SCADA and PMU measurements as model inputs. Considering PMUs have a higher sampling rate, the real-time PMUs are used in **event monitoring** to capture the sudden change of power system, i.e., the sudden change in the measurements y_t .

$$alarm = \begin{cases} 1, & |y_t - y_{t-1}| > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where ε is a pre-defined threshold. Once the alarm is triggered (e.g., at time t_{on}), the previously collected SCADA measurement y_{t_0} will be used in **pre-fault system information**, and PMU measurements $\{y_{t_{on}-w/2}, \dots, y_{t_{on}+w/2}\}$ are taken as **fault-on system information**, where w indicates the length of time window.

In **pre-fault system information**, pre-fault graph-structured data is generated using SCADA measurements. Since SCADA is able to collect global measurements, the pre-fault SCADA measurements of all buses are collected. In addition, RES information ξ_p , such as wind speed, is recorded to calculate the penetration level of wind generation. Given the power

system topology, the operation and topology information of the pre-fault system with M buses and L transmission lines are presented by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ having node attributes h_v, h_u (node $v, u \in \mathcal{V}$) and edge attributes e_{uv} (edge $(v, u) \in \mathcal{E}$). The pre-fault SCADA measurements (i.e., injected active and reactive power) of all buses are used as node attributes $h_v = \{P_v, Q_v\}$. Note that the active power output of the wind turbine is also recorded in node attributes to incorporate the uncertainties from RES. The line impedance $Y_{uv}, (v, u) \in \mathcal{E}$ is used as the edge attribute $e_{uv} = Y_{uv}$. In this way, the pre-fault SCADA measurements and uncertainty information are represented as the graph-structured data $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

As for **fault-on system information**, w number of PMU measurements after the event triggered are collected and used for constructing the fault-on graph-structured data. Considering the scenario where PMUs are installed on N ($N < M$) key buses, the PMU measurements are preprocessed as graph-structured data $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, consisting of a node (bus) set $\mathcal{V}'$ and an edge set $\mathcal{E}'$. The PMU measurements (i.e., injected active and reactive power) of N key buses are used as node attributes $h_{v'} = \{P_{v'}, Q_{v'}\}, v' \in \mathcal{V}'$. Since the fault location is unknown during the fault-on stage, a fully connected graph (i.e., all nodes are connected) is used in learning the relationship of nodes.

SCADA-based graph-structured data focuses on pre-fault steady-state system information. PMU-based graph-structured data captures fast-evolving system dynamics during faults. The hybrid graph fusion combines slow-scale system topology and fast-scale system dynamics, addressing the multi-timescale challenge in processing multi-source power system data.

Finally, the SCADA data and PMU data will be used by the trained model to output **DSA results**, including the predicted dynamics trajectories and the fault location. To determine the system security status, the transient stability index (TSI) is calculated using the predicted rotor angle trajectories.

$$TSI = 100 \times (360 - \Delta\delta_{max}) / (360 + \Delta\delta_{max}) \quad (3)$$

where $\Delta\delta_{max}$ is the maximum rotor angle difference between any two generators. If $TSI > 0$, the system is transient stable; if $TSI < 0$, the system is transient unstable.

2) Learning Problems: Two major DSA tasks are dynamics trajectories prediction and fault location identification. The corresponding learning problems are described as follows.

Dynamics Trajectory Prediction: Most DL-based trajectory predictions assume the states to be predicted are observable. Iteration prediction strategy is commonly adopted to learn temporal relationships from the initial trajectories and predict the future states using a rolling window. However, the complexity of prediction tasks significantly increases when utilizing measurements to forecast unobservable states. This is due to the potential mismatch between the temporal characteristics observed in measurements and those inherent in the states.

Therefore, the learning problem focused in this paper is to predict the trajectories of states $x_{t_0, \dots, t_{end}}$ using the pre-fault system information $[y_{t_0}, \xi_p]$ and PMU measurements

$\mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2}$. The learning problem is given by

$$\mathbf{x}_{t_0, \dots, t_{end}} = G_{\Phi, \Psi, \phi}(\mathbf{y}_{t_0}, \boldsymbol{\xi}_p, \mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2}). \quad (4)$$

The goal is to learn the mapping G from the input space to output states. To solve this problem, this paper designs a novel neural network $G_{\Phi, \Psi, \phi}$, parameterized by Φ, Ψ and ϕ . Graph neural networks $G_{\Phi, \Psi}$ are developed to learn the spatial-temporal representation of hybrid inputs. Then, a Fourier neural operator G_{ϕ} is customized to learn the temporal relationship in the frequency domain. Parameters Φ and ϕ are determined by the model structure and the detailed model introduction is given in Section III. A and B.

Fault Location Identification: Given that various fault locations result in distinct deviations in the measurements $\mathbf{y}$, fault information can be anticipated through analyzing distinct spatial-temporal patterns observed in $\mathbf{y}$. Therefore, the learning problem of locating faulted line is described as

$$P(\text{line}_i) = G_{\Phi, \Psi, \theta}(\mathbf{y}_{t_0}, \boldsymbol{\xi}_p, \mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2}). \quad (5)$$

The goal of this problem is to learn the mapping G from the pre-fault system information $[\mathbf{y}_{t_0}, \boldsymbol{\xi}_p]$ and PMU measurements $\mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2}$ to the probability of faulted lines $P(\text{line}_i)$. The spatial-temporal features learned by GNNs $G_{\Phi, \Psi}$ are used as inputs of classifier G_{θ} . A detailed model introduction is given in Section III.C. The line with the highest failure probability is identified as the faulted line i^* .

$$i^* = \arg \max_i P(\text{line}_i) \quad (6)$$

III. METHODOLOGY

A. GNNs for Spatial-Temporal Representation Learning

To handle hybrid measurements and incorporate diverse system information into the model inputs, two GNNs are constructed to learn distinct spatial-temporal representations. GNN G_{Φ} is developed to learn features from SCADA measurements, while GNN G_{Ψ} learns features from PMU measurements, as shown in Fig.2.

1) GNN G_{Φ} for SCADA measurements: Given the pre-fault SCADA measurements $\mathbf{y}_{t_0}$ of all buses, the system uncertainty $\boldsymbol{\xi}_p$, and the system topology, a pre-fault graph-structured data $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node features $\mathbf{h}_v, \mathbf{h}_u$ (node $v, u \in \mathcal{V}$) and edge features e_{uv} (edge $(v, u) \in \mathcal{E}$) is constructed. Each node represents the bus. The injected active and reactive power of the buses are taken as node features $\mathbf{h}_v = \{P_v, Q_v\}$. Each line represents the transmission line. Line impedance $Y_{uv}, (v, u) \in \mathcal{E}$ is used as the edge features $e_{uv} = Y_{uv}$.

To deal with pre-fault graph-structured data $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the GNN consisting of a graph transformer and a graph isomorphism network is developed. For graph transformer layer [20], source node feature $\mathbf{h}_v$ and distant node feature $\mathbf{h}_u$ are transformed to query vector $\mathbf{Q}_v$ and key vector $\mathbf{K}_u$:

$$\mathbf{Q}_v = W_Q \mathbf{h}_v + b_Q \quad (7)$$

$$\mathbf{K}_u = W_K \mathbf{h}_u + b_K. \quad (8)$$

To take the edge feature $e_{uv}, (u, v) \in \mathcal{E}$ into consideration, the key vector $\hat{\mathbf{K}}_{uv}$ can be calculated as

$$\hat{\mathbf{K}}_{uv} = \mathbf{K}_u + W_{\mathcal{E}} e_{uv} + b_{\mathcal{E}}. \quad (9)$$

The value vector $\hat{\mathbf{V}}_u$ integrates the distant node feature and edge feature:

$$\hat{\mathbf{V}}_u = \mathbf{V}_u + W_{\mathcal{E}} e_{uv} + b_{\mathcal{E}} \quad (10)$$

where

$$\mathbf{V}_u = W_V \mathbf{h}_u + b_V. \quad (11)$$

Hence, the single-attention function is defined as

$$\mathbf{h}_v = \sum_u \text{softmax}(\mathbf{Q}_v^T \hat{\mathbf{K}}_{uv} / \sqrt{d}) \hat{\mathbf{V}}_u \quad (12)$$

where d is the dimension of keys. The multi-head attention mechanism can be expressed as the concatenation of the outputs from individual attention function:

$$\mathbf{h}_v = \parallel_{c=1}^C \left(\sum_u \text{softmax}(\mathbf{Q}_{c,v}^T \hat{\mathbf{K}}_{c,uv} / \sqrt{d}) \hat{\mathbf{V}}_{c,u} \right) \quad (13)$$

where $\parallel(\cdot)$ is the concatenate operator for C head attention. $\mathbf{Q}_{c,v}, \hat{\mathbf{K}}_{c,uv}, \hat{\mathbf{V}}_{c,u}$ are query vector, key vector and value vector of the c th head, respectively.

The layer normalization [21] and ReLU function are implemented after the graph transformer layer.

$$\mathbf{h}_v = \text{ReLU}(\text{LayerNorm}(\mathbf{h}_v)) \quad (14)$$

A graph isomorphism network including edge features (GINE) [22] is adopted after the graph transformer layers. GINE updates node representations as follows.

$$\mathbf{h}_v = \text{MLP}((1 + \epsilon)\mathbf{h}_v + \sum_u \text{ReLU}(\mathbf{h}_u + e_{uv})) \quad (15)$$

where MLP indicates multi-layer perceptron. Then, a global mean pooling layer is used to get the graph-level features.

$$\mathbf{x}^{\Phi} = 1/M \sum \mathbf{h}_v \quad (16)$$

where M is the number of nodes. $\mathbf{x}^{\Phi}$ indicates the output feature of GNN G_{Φ} , i.e., $\mathbf{x}^{\Phi} = G_{\Phi}(\mathbf{y}_{t_0}, \boldsymbol{\xi}_p)$.

2) GNN G_{Ψ} for PMU measurements: Given the fault-on PMU data $\{\mathbf{y}_{t_{on}-w/2}, \dots, \mathbf{y}_{t_{on}+w/2}\}$, measurements at each time point $\mathbf{y}_t$ are processed as a graph-structured data, $\mathcal{G}'_t = (\mathcal{V}'_t, \mathcal{E}'_t)$. Each node represents the PMU and the corresponding PMU measurements are taken as node features, i.e., $\mathbf{h}_{v'} = \{P_{v'}, Q_{v'}\}$ (node $v' \in \mathcal{V}'_t$). Considering the fault location is unknown during fault-on stage, it is assumed that all nodes are connected and the edge features are identical, i.e., $e_{u'v'} = 1$. Node features are used in the message aggregation of the topology adaptive graph convolutional network (TAGConv) [23]. GNN G_{Ψ} consists of two TAGConv layers, each of which performs convolution by adopting a set of size-1 up to size- D filters.

$$\mathbf{h}_{v'} = \sigma(\sum_{d=1}^D g_{f,d} \mathbf{A}^d \mathbf{h}_{v'} + b_f) \quad (17)$$

where σ is a nonlinear operation, e.g., rectified linear unit (ReLU). b_f is a learnable bias, $g_{f,d}$ is the graph filter polynomial coefficients of f -th filter with size d , and $\mathbf{A}$ is the normalized adjacency matrix of graph.

$$\mathbf{A} = \mathbf{D}^{-1/2} \bar{\mathbf{A}} \mathbf{D}^{-1/2} \quad (18)$$

where $\bar{\mathbf{A}}$ is the weighted adjacency matrix, $\mathbf{D} = \text{diag}[d]$ with i -th component being $d(i) = \sum_j A_{i,j}^3$. A layer normalization is adopted between the two TAGConv layers. For a set of PMU measurements with w instances, i.e., $\mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2}$, the features learned from GNN G_{Ψ} is denoted as $\mathbf{x}^{\Psi} = G_{\Psi}(\mathbf{y}_{t_{on}-w/2, \dots, t_{on}+w/2})$.

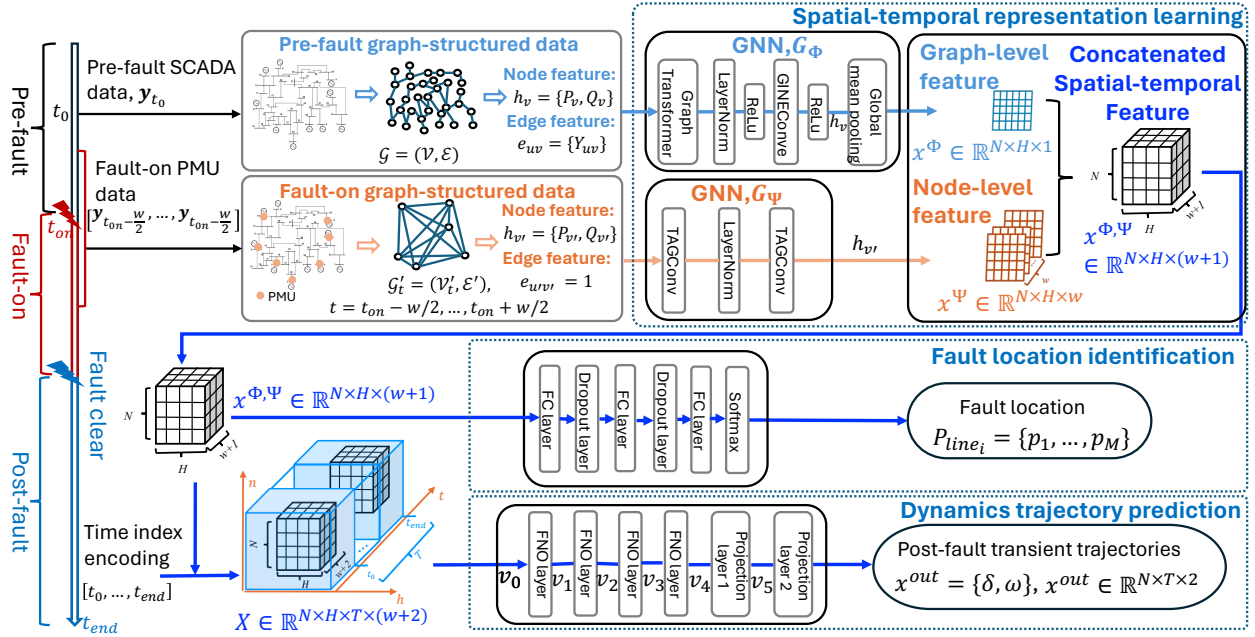


Fig. 2. The architecture of the developed model.

3) *Spatial-temporal feature concatenation*: As illustrated in the upper part of Fig.2, GNN G_Φ takes the node features h_v and edge features e_{uv} as inputs and learns the graph-level feature x^Φ . GNN G_Ψ uses node features $h_{v'}$ and outputs the node-level feature x^Ψ . In this way, the graph-level feature x^Φ captures the global pre-fault system information while the node-level feature x^Ψ learns the local fault-on system information. Finally, the graph-level feature x^Φ is reshaped and concatenated with the node-level features x^Ψ . The spatial-temporal feature $x^{\Phi, \Psi} = [x^\Phi; x^\Psi] = G_{\Phi, \Psi}(y_{t_0}, \xi_p, y_{t_{on}-w/2}, \dots, y_{t_{on}+w/2})$ is used as the inputs of FNO and the classifier.

B. FNO for System Dynamics Learning

FNO [24], a powerful neural network capable of learning mappings between infinite function spaces, exhibits exceptional performance in learning operator solutions for partial differential equations, thereby making it suitable for learning and predicting power system dynamics.

An FNO layer transforms the inputted time-domain data into the frequency domain via Fourier Transform $\mathcal{F}$, applies a linear transform R on the lower frequency modes and filters out the higher modes, then recovers the time-domain data using inverse transform $\mathcal{F}^{-1}$. To compensate for truncated high-frequency information, a residual connection applies a linear operator W directly on the time-domain data preserving local details. By combining frequency-domain processing (for global low-frequency patterns) and time-domain residuals (for high-frequency details), each FNO layer hierarchically learns multiscale features. In each FNO layer, the update to the representation $v_i(x) \mapsto v_{i+1}(x)$ is calculated as

$$v_{i+1}(x) = \sigma(Wv_i(x) + \mathcal{F}^{-1}(R \cdot \mathcal{F}(v_i))(x)), \quad (19)$$

where σ is the activation function, i.e., GeLU. W and R are linear transforms. $\mathcal{F}$ denotes Fourier transform and $\mathcal{F}^{-1}$ is

inverse Fourier transform.

$$(\mathcal{F}v)(k) = \int v(x)e^{-2i\pi \langle x, k \rangle} dx, \quad (20)$$

$$(\mathcal{F}^{-1}v)(x) = \int v(k)e^{2i\pi \langle x, k \rangle} dk, \quad (21)$$

where $i = \sqrt{-1}$ is the imaginary unit, k is frequency mode, and $\langle \cdot, \cdot \rangle$ denotes the inner product. Fast Fourier Transform (FFT) [10] are utilized to compute Eqs.(20) and (21). Typically, 4 FNO layers are adopted.

To predict the system states, the time indexes $t, t \in [t_0, \dots, t_{end}]$ are encoded into the learned features $x_t = [x^{\Phi, \Psi}; t]$. Therefore, the inputs of FNO is denoted as $v_0 = X = \{x_{t_0}, \dots, x_{t_{end}}\}, X \in \mathbb{R}^{N \times H \times T \times (w+2)}$, as shown in Fig.2. N is the number of PMU nodes, H is the number of hidden features, and $T = t_{end} - t_0$ is the number of encoded time stamps. In each FNO layer, the FFT $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ are calculated as

$$(\mathcal{F}v)_{k_n, k_h, k_t, l}(k) = \sum_{n=0}^{N-1} \sum_{h=0}^{H-1} \sum_{t=0}^{T-1} v(n, h, t, l) e^{-2i\pi(\frac{nk_n}{N} + \frac{hk_h}{H} + \frac{tk_t}{T})}, \quad (22)$$

$$(\mathcal{F}^{-1}v)_{n, h, t, l}(x) = \sum_{k_n} \sum_{k_h} \sum_{k_t} v(k_n, k_h, k_t, l) e^{2i\pi(\frac{nk_n}{N} + \frac{hk_h}{H} + \frac{tk_t}{T})}. \quad (23)$$

To enhance implementation efficiency, the Fourier series is truncated at a predefined maximum number of modes k_{max} . Therefore, the weight tensor $R \in \mathbb{R}^{N \times H \times T \times (w+2) \times (w+2)}$ is used to truncate the higher modes.

$$(R \cdot (\mathcal{F}v))_{k_n, k_h, k_t, l} = \sum_{j=1}^{w+2} R_{k_n, k_h, k_t, l, j} (\mathcal{F}v)_{k_n, k_h, k_t, j} \quad (24)$$

Therefore, the inverse Fourier transform $\mathcal{F}^{-1}(\mathbf{R} \cdot \mathcal{F}(\mathbf{v}))(x)$ is calculated as

$$\mathcal{F}^{-1}(\mathbf{R} \cdot \mathcal{F}(\mathbf{v}))_{n,h,t,l}(x) = \sum_{k_n=0}^{k_{max}} \sum_{k_h=0}^{k_{max}} \sum_{k_t=0}^{k_{max}} (\mathbf{R} \cdot (\mathcal{F}\mathbf{v}))_{k_n,k_h,k_t,l} e^{2i\pi(\frac{nk_n}{N} + \frac{hk_h}{H} + \frac{tk_t}{T})}. \quad (25)$$

After four FNO layers, the output feature $\mathbf{v}_4 \in \mathbb{R}^{N \times H \times T \times (w+2)}$ will be sent to two projection layers. Projection layer 1 is designed to aggregate $w+2$ temporal features, which consist of features learned from w PMU measurements, 1 pre-fault instance, and 1 encoded time stamp. Two fully connected (FC) layers followed by activation functions (i.e., GeLU) are used. After the projection layer 1, the feature $\mathbf{v}_4 \in \mathbb{R}^{N \times H \times T \times (w+2)}$ becomes $\mathbf{v}_5 \in \mathbb{R}^{N \times H \times T \times 1}$.

Since the goal is to predict the state, the dimension of hidden feature, i.e., H , is different from the dimension of states. After reshaping the feature, three FC layers are used to learn the mapping: $\mathbb{R}^{N \times T \times H} \rightarrow \mathbb{R}^{N \times T \times 2}$. As a result, the output is the post-fault system state $\mathbf{x}_{n,t}^{out} = \{\delta, \omega\}_{n,t}$ of n^{th} ($n = 1, \dots, N$) generator at time t .

To train the GNN-FNO model $G_{\Phi, \Psi, \phi}$, the mean squared error loss is adopted. Given that S sets of trajectories are used for training, the loss is computed as below.

$$L_{G_{\Phi, \Psi, \phi}} = \frac{1}{S \times N \times T \times 2} \sum_{i=1}^{S \times N \times T \times 2} (x_i^{out} - x_i^{real})^2 \quad (26)$$

where x_i^{out} and x_i^{real} are predicted and real states, respectively.

C. Classifier for Fault Location Identification

To identify fault location, a classifier consisting of three FC layers is developed, as shown in Fig.2. The spatial-temporal features $\mathbf{x}^{\Phi, \Psi}$ learned from two GNNs are used as inputs of the classifier. Since the fault location is treated as a multi-classification problem, cross entropy loss is used for optimizing the classifier parameters θ .

$$L_{G_{\theta}} = -1/S \sum_{i=1}^S \sum_{m=1}^M (p_{i,m}^{real} \log(p_{i,m}^{out})) \quad (27)$$

where $p_{i,m}^{out}$ is the predicted probability that fault of the i th sample occurs at transmission line m ($m = 1, \dots, M$). $p_{i,m}^{real}$ denotes the real probability and is calculated using the real label $l_{i,m}^{real}$:

$$p_{i,m}^{real} = \exp(l_{i,m}^{real}) / \sum_{m=1}^M \exp(l_{i,m}^{real}). \quad (28)$$

If the fault of the i th sample occurs at transmission line m , $l_{i,m}^{real} = 1$; otherwise, $l_{i,m}^{real} = 0$.

IV. CASE STUDY

A. Simulation and Hyper-Parameter Setup

1) Test System and Dataset: The developed online DSA is tested on IEEE 39 Bus System. Synchronous generators 5, 7, and 8 are replaced by wind generators. Five penetration levels (i.e., 10%, 20%, 30%, 40%, and 50%) are used to simulate the renewable energy uncertainties. For load uncertainty, it is assumed that the active power load of each bus follows a normal distribution $\mathcal{N}(1, 0.1^2)$ (i.e., mean is 1 and standard deviation is 0.1) of the base case, and the reactive power load is calculated by multiplying the active power load with

a factor uniformly drawn from the range $[0.25, 0.55]$. For N-1 contingency, it is assumed that 16 lines (i.e., Line 04 - 05, Line 06 - 11, Line 10 - 13, Line 16 - 19, Line 16 - 24, Line 26 - 29, Line 23 - 24, Line 21 - 22, Line 08 - 09, Line 02 - 25, Line 25 - 26, Line 26 - 27, Line 16 - 17, Line 14 - 15, Line 03 - 04, Line 01 - 39) have been randomly tripped. After 0.1s, the fault is cleared by tripping the line according to protection schedule.

SCADA measurements include the active power and reactive power of each bus. 6 PMUs are deployed on the bus directly connected to generators (i.e., Bus 30, Bus 32, Bus 33, Bus 35, Bus 38, Bus 39). It is assumed that fault occurs at $t_{on} = 0.1$ s. SCADA measurements collected at $t = 0$ and PMU measurements collected from $t \in [0.05, 0.15]$ are used ($w = 10$ data points are collected during 0.1s). A total number of 3482 samples (70% for training and 30% for testing) are generated, each of which contains SCADA measurements $[P, Q]_{39 \times 2}$, PMU measurements $[P, Q]_{6 \times 2 \times 10}$, the system transient dynamics trajectories to be predicted $[\delta, \omega]_{6 \times 2 \times 300}$, and the fault location.

2) Hyper-parameter Setup: The developed model architecture is shown in Fig.2. (*input size, output size*) of GNN model components are as follows. G_{Φ} : graph transformer (2,8), GINEConve (8,8). G_{Ψ} : TAGConv (2,8) $\rightarrow$ TAGConv (8,8). MLP-based fault location identification: MLP (64 $\rightarrow$ 64 $\rightarrow$ 16). FNO-based dynamics trajectory prediction: $k_n = 7$, $k_h = 9$, $k_t = 13$, and 4 FNO layers. In the training, 3-fold cross-validation is adopted. The batch size and learning rate are 128, and 0.01, respectively. We use Pytorch and a single Nvidia GeForce RTX 3090 GPU with 24GB memory.

B. Ablation Study

In this part, ablation studies are designed to compare the impact of graph structure and feature selection strategies on model performance. In comparison, the mean squared error (MSE) and mean absolute percentage error (MAPE) is used to assess the prediction performance over the period K :

$$MAPE = 100\% \times 1/K \sum_{k=0}^K (|x_k^{true} - x_k^{pred}|) / \max(1, x_k^{true}) \quad (29)$$

where x_k^{true} and x_k^{pred} are the true and predicted value.

The developed method G_{Φ} - G_{Ψ} -FNO(4) using a GNN G_{Φ} , a GNN G_{Ψ} , and an FNO block with 4 layers. The inputs of G_{Φ} are SCADA measurements and the graph structure is designed based on power system topology. The inputs of G_{Ψ} are PMU measurements and a fully connected graph is adopted to model the spatial relationship between PMU nodes. To explore the impacts of SCADA measurements on DSA, G_{Ψ} -FNO(4) is used for comparison. It can be observed that the trajectory prediction error increase and the fault location classification accuracy decreases when only PMU measurements are used.

Then, G_{Φ} - G_{Ψ} -FNO(6), G_{Φ} - G_{Ψ} -FNO(2), and G_{Φ} - G_{Ψ} -WT-CNN are designed to compare the impacts of frequency-domain feature selection on model performance. G_{Φ} - G_{Ψ} -FNO(6) and G_{Φ} - G_{Ψ} -FNO(2) represent the developed method with 6 FNO layers and 2 FNO layers, respectively. The comparison between G_{Φ} - G_{Ψ} -FNO(4) and G_{Φ} - G_{Ψ} -FNO(2)

indicates that 2 FNO layers are insufficient to capture complex features, thus leading to degraded model performance. Although G_Φ - G_Ψ -FNO(6) outperforms G_Φ - G_Ψ -FNO(2), its dynamics prediction performance is not improved compared to G_Φ - G_Ψ -FNO(4). Therefore, a 4-layer FNO block is adopted in the developed method. For G_Φ - G_Ψ -WT-CNN, the FNO block is replaced by a Wavelet Transform and a CNN module [25]. It can be seen that the 2-layer FNO module performs better than using Wavelet Transform, which indicates the superior frequency-domain learning capability of FNO layers.

The impacts of different graph structures are compared, as illustrated in the last two rows of TABLE I. ' $\mathcal{G}$, $\mathcal{G}'$ ': All connected' indicates that the fully connected graphs are used in modeling both SCADA measurements and PMU measurements. ' $\mathcal{G}$, $\mathcal{G}'$ ': Physical topology' indicates that the nodes in SCADA measurements and PMU measurements are connected following the power system topology. It can be observed that when PMU measurements are modeled using physical topology, the classification performance has an obvious degradation. Therefore, when the available PMU nodes are limited, a fully connected graph has better capability to capture hidden spatial-temporal features compared to using physical topology.

C. Trajectory Prediction Performance Validation

To validate the trajectory prediction performance of the developed method, different models, including LSTM-RNN [5], FNO [13], and structure-informed graph learning [4] are utilized for comparison. The structure-informed graph learning adopts a GNN to process spatial data and a gated-recurrent unit to predict system trajectories. TABLE II illustrates the comparison results. It can be observed that the FNO-based methods (i.e., G_Φ - G_Ψ -FNO(4) and FNO) outperform the other two trajectory prediction methods. This validates the effectiveness of FNO learning complex power system dynamics. By comparing the model performance of G_Φ - G_Ψ -FNO(4) and FNO, it can be found that the utilization of GNNs (i.e., G_Φ and G_Ψ) can further reduce the prediction errors (e.g., $\text{MAPE}(\omega)$ is reduced from 0.22% to 0.18%). This is because the multi-source spatial-temporal features learned by G_Φ and G_Ψ enrich the inputs of FNO to help it learn the complex nonlinear system dynamics.

D. Fault Location Identification Performance Validation

To validate the fault location identification performance of the developed method, decision tree (DT) [26], support vector machine (SVM) [27], LSTM with self-attention mechanism (LSTM-SAF) [28], and CNN-LSTM [6] are used for comparison. PMU measurements are used as inputs of DT, SVM, and LSTM-SAF. CNN-LSTM adopts matrix mapping to extract spatial relationship of PMU data and use CNN to learn spatial features. The developed methods use GNNs (i.e., G_Φ and G_Ψ) to learn the multi-source spatial-temporal features. The comparison results are shown in TABLE III. It can be observed that methods that learn spatial features (i.e., G_Φ - G_Ψ -FNO(4) and CNN-LSTM) outperform the other three methods. In addition, among 5 learning models, the developed

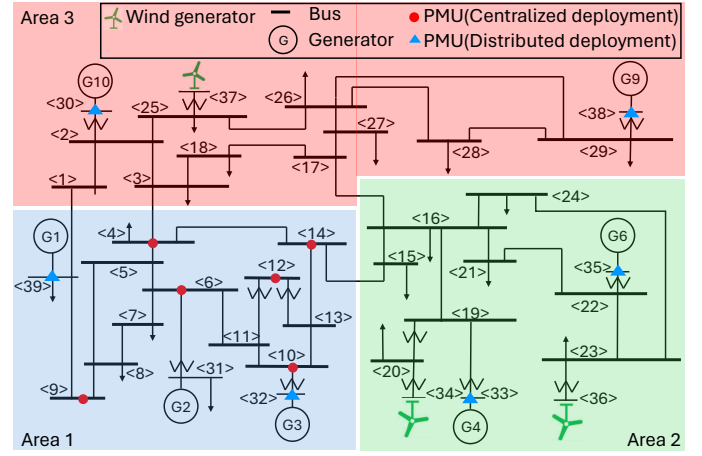


Fig. 3. IEEE 39-Bus System with 3 areas.

method achieves the highest fault location identification accuracy. These experimental findings demonstrate that explicitly modeling spatial-temporal features improves neural network performance in power system data analytics.

E. Discussion on Different PMU Deployments

In real applications, the system configurations (e.g., the deployment of PMUs and the number of available PMUs) can be diverse. In the previous demonstrations, it is assumed that PMUs are distributed throughout the system, i.e., installed on generator buses (Bus 30, 32, 33, 35, 38, 39). For comparison, a more centralized distribution is adopted in the IEEE 39-Bus System with 3 areas (see Fig.3). PMUs are installed in Area 1 (Bus 4, 6, 9, 10, 12, 14). Considering the scenario where the number of available PMUs is reduced due to economic reasons or unexpected events, measurements from 3 PMUs are input to the developed DSA to predict the dynamics trajectory of 6 generators and identify fault locations. For distributed deployment, measurements from Bus 30, Bus 32, and Bus 33 are used. For centralized deployment, it is assumed that available PMUs are on Bus 4, Bus 9, and Bus 14.

1) *Dynamics Trajectory Prediction*: Prediction errors for rotor angle prediction and rotor speed prediction are shown in Fig.4 and Fig.5. The average MAPEs of 6 generators are given in the 3rd and 4th columns in TABLE IV. By comparing the prediction results of distributed deployment and centralized deployment, e.g., the 2nd and 4th rows of TABLE IV, the developed DSA demonstrates comparable trajectory prediction performance (i.e., the $\text{MAPE}(\delta)$ of the 4th row is slightly higher than that of the 2nd row). However, with the reduction of available PMU numbers, the MAPEs have larger increases. For example, the average $\text{MAPE}(\delta)$ of distributed deployment with 3 PMUs is twice the average $\text{MAPE}(\delta)$ of distributed deployment with 6 PMUs.

It can be concluded that the developed online DSA is robust to different deployment patterns, but the prediction accuracy decreases with the reduction of available PMU numbers.

2) *Fault Location Identification*: The fault location identification results are illustrated in the 5th to 8th columns of TABLE IV. The average accuracy of the four deployments is

TABLE I
DSA RESULTS FOR DIFFERENT GRAPH STRUCTURES AND LEARNING MODEL STRUCTURES

Graph structure	Methods	Dynamics trajectory prediction				Fault location identification			
		MAPE(δ)	MAPE(ω)	MSE(δ)	MSE(ω)	F1	Precision	Recall	Acc
$\mathcal{G}$: Physical topology $\mathcal{G}'$: All connected	G_{Φ} - G_{Ψ} -FNO(4)	10.68%	0.18%	0.0377	$2.07e-5$	0.98	0.98	0.98	0.98
	G_{Φ} -FNO(4)	16.54%	0.25%	0.0795	$4.12e-5$	0.79	0.85	0.82	0.82
	G_{Φ} - G_{Ψ} -FNO(6)	11.29%	0.22%	0.0394	$3.22e-5$	0.94	0.95	0.94	0.94
	G_{Φ} - G_{Ψ} -FNO(2)	12.95%	0.21%	0.0500	$2.08e-5$	0.87	0.89	0.87	0.87
$\mathcal{G}$, $\mathcal{G}'$: All connected $\mathcal{G}$, $\mathcal{G}'$: Physical topology	G_{Φ} - G_{Ψ} -WT-CNN	16.94%	0.26%	0.0728	$4.16e-5$	—	—	—	—
	G_{Φ} - G_{Ψ} -FNO(4)	11.10%	0.20%	0.0380	$2.22e-5$	0.94	0.95	0.94	0.94
		11.39%	0.18%	0.0402	$1.77e-5$	0.87	0.90	0.87	0.87

TABLE II

DYNAMICS TRAJECTORY PREDICTION PERFORMANCE COMPARISON

Methods	Dynamics trajectory prediction			
	MAPE(δ)	MAPE(ω)	MSE(δ)	MSE(ω)
G_{Φ} - G_{Ψ} -FNO(4)	10.68%	0.18%	0.0377	$2.07e-5$
Structure-informed graph learning	13.32%	0.28%	0.0546	$6.70e-5$
FNO	11.35%	0.22%	0.0418	$2.75e-5$
LSTM-RNN	22.59%	0.30%	0.1351	$8.52e-5$

TABLE III

FAULT LOCATION IDENTIFICATION COMPARISON

Methods	Weighted average metrics			
	F1	Precision	Recall	Acc
G_{Φ} - G_{Ψ} -FNO(4)	0.98	0.98	0.98	0.98
DT	0.90	0.92	0.90	0.90
SVM	0.89	0.90	0.91	0.91
LSTM-SAF	0.84	0.86	0.85	0.85
CNN-LSTM	0.93	0.94	0.93	0.93

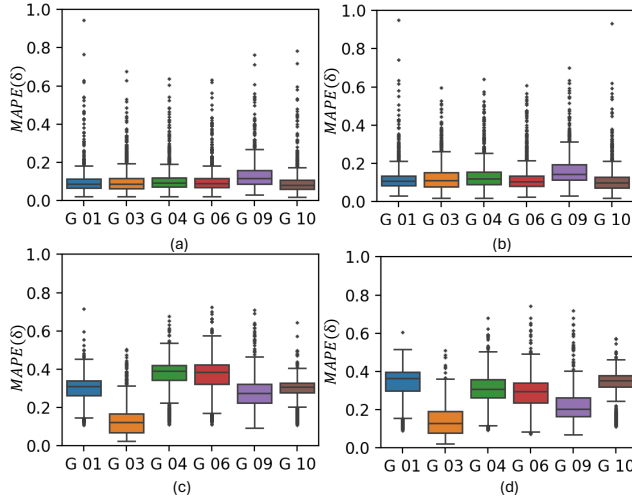


Fig. 4. Boxplots for rotor angle trajectory prediction error of (a) Distributed deployment-6 PMUs; (b) Centralized deployment-6 PMUs; (c) Distributed deployment-3 PMUs; (d) Centralized deployment-3 PMUs.

0.96. It demonstrates the superior fault location identification performance of the developed DSA under different deployments. Among the four deployments, the developed DSA with distributed deployment using 3 PMUs has the poorest performance. 44 samples belonging to the fault location 5 (Line 16 - 24) are misclassified as the fault location 13 (Line 16 - 17). This is because fault locations 5 and 13 are very close, i.e., both lines are connected to Bus 16, thus making fault location identification a more difficult task.

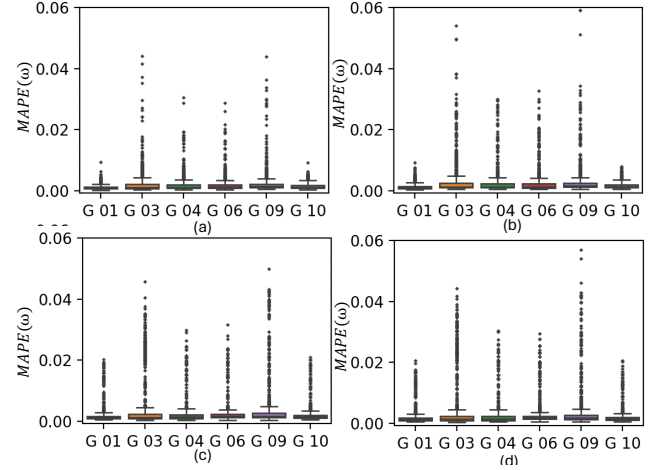


Fig. 5. Boxplots for rotor speed trajectory prediction error of (a) Distributed deployment-6 PMUs; (b) Centralized deployment-6 PMUs; (c) Distributed deployment-3 PMUs; (d) Centralized deployment-3 PMUs.

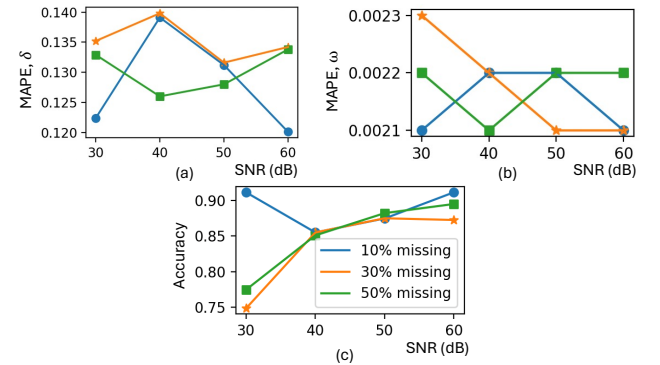


Fig. 6. Comparative assessment of model robustness to noise and missing data. (a) Prediction error $MAPE(\delta)$; (b) Prediction error $MAPE(\omega)$; (c) Fault location classification accuracy.

F. Discussion on Model Robustness

This part considers the data imperfection scenarios and the system malfunction scenario. For data imperfection scenarios, different levels (10%, 30%, and 50%) of data are randomly missing and White Gaussian Noise (60 dB, 50 dB, 40 dB, and 30 dB) is added to the original data. The level of noise is indicated by signal-to-noise ratio (SNR). The comparative assessment of model robustness to different noises and missing data are illustrated in Fig. 6. It can be observed that the transient dynamics trajectory prediction performance has little deviation ($\Delta MAPE(\delta) = |\max MAPE(\delta) - \min MAPE(\delta)| = 1.97\%$, $\Delta MAPE(\omega) = 0.02\%$) with the

TABLE IV
DSA RESULTS FOR DIFFERENT PMU DEPLOYMENTS

Deployment pattern	PMU number	Dynamics trajectory prediction				Fault location identification			
		MAPE(δ)	MAPE(ω)	MSE(δ)	MSE(ω)	F1	Precision	Recall	Acc
Distributed deployment	6 (Bus 30,32,33,35,38,39)	10.68%	0.18%	0.0377	$2.07e-5$	0.98	0.98	0.98	0.98
	3 (Bus 30,32,33)	28.52%	0.28%	0.1592	$6.18e-5$	0.89	0.92	0.90	0.90
Centralized deployment	6 (Bus 4,6,9,10,12,14)	13.02%	0.23%	0.0507	$4.18e-5$	0.97	0.98	0.97	0.97
	3 (Bus 4,9,14)	26.79%	0.29%	0.1668	$6.74e-5$	0.98	0.98	0.98	0.98

TABLE V
MODEL PERFORMANCE UNDER SYSTEM MALFUNCTION SCENARIOS

MAPE(δ)	MAPE(ω)	MSE(δ)	MSE(ω)
13.10%	0.24%	0.0560	$3.83e-5$
F1	Precision	Recall	Acc
0.92	0.92	0.92	0.92

increase of noise and missing data. This is thanks to the strong learning capability of FNO, which can learn in frequency domain to obtain global frequency patterns rather than learning from individual points, thus enabling robust predictions with imperfect data. When the range of SNR values of noise is 60dB–40dB, which is a reasonable range in the practical PMU data [5], the fault location accuracy is above 0.85 even when 50% of data is missing, which indicates the robustness of the developed model under data imperfection scenarios.

To simulate the scenario where system topology change is caused by system malfunction, one line (*Line 17 – 18*) is tripped in addition to N-1 contingency. A total number of 751 samples are generated. 651 samples are used for training and 100 samples for testing. The model performance under system malfunction scenarios is shown in TABLE V. Compared to baseline scenario, the model performance has a little degradation. For example, $MAPE(\delta)$ is increased from 10.68% to 13.10% and accuracy is decreased from 0.98 to 0.92. However, it is comparable to other methods employed in baseline scenario, such as structure-informed graph learning ($MAPE(\delta) = 13.32\%$ in TABLE II) and SVM ($Acc = 0.91$ in TABLE III).

G. Model Scalability Demonstration

To validate the model scalability, the IEEE 118 Bus System, which represents an approximation of the American Electric Power System (in the U.S. Midwest), is adopted. Synchronous generators 10, 69, and 80 are replaced by wind generators. Five penetration levels are used to simulate the RES uncertainties. For load uncertainty, the active power load of each bus follows a normal distribution $\mathcal{N}(1, 0.1^2)$ of the base case, and the reactive power load is calculated by multiplying the active power load with a factor uniformly drawn from the range [0.25, 0.55]. For N-1 contingency, 16 lines (i.e., Line 12-117, Line 23-32, Line 34-43, Line 45-49, Line 83-85, Line 51-58, Line 60-62, Line 100-106, Line 70-75, Line 80-99, Line 92-102, Line 105-108, Line 5-11, Line 27-115, Line 47-69, Line 68-116) have been randomly tripped. After 0.1s, the fault is cleared.

Two PMU deployment patterns are considered. For distributed deployment, 6 PMUs are deployed on the bus directly connected to generators (i.e., Bus 12, Bus 31, Bus 59, Bus

65, Bus 87, Bus 111). The transient dynamics trajectories of these 6 generators are predicted using the developed method. For centralized deployment, 6 PMUs are deployed on Bus 75, Bus 79, Bus 85, Bus 98, Bus 103, and Bus 108.

It is assumed that fault occurs at $t_{on} = 0.1s$. SCADA measurements collected at $t = 0$ and PMU measurements collected from $t \in [0.05, 0.15]$ are used. A total number of 1236 samples (70% for training and 30% for testing) are generated, each of which contains SCADA measurements $[P, Q]_{118 \times 2}$, PMU measurements $[P, Q]_{6 \times 2 \times 10}$, the system transient dynamics trajectories to be predicted $[\delta, \omega]_{6 \times 2 \times 300}$, and the fault location.

To validate model performance, learning models that performed well in the IEEE 39 bus system are used as comparison models for the IEEE 118 bus system test. For transient dynamics trajectory prediction, FNO is adopted. For fault location identification, CNN-LSTM is utilized for comparison. The comparison results are illustrated in TABLE VI. Compared to FNO and CNN-LSTM, the developed method has the lowest prediction errors and the highest fault location identification accuracy when applied to distributed deployment with 6 PMUs, demonstrating the model scalability to larger systems. In addition, it can be observed that all methods show better performance when applied distributed deployment with 6 PMUs. For example, $MSE(\delta)$ of the developed method has increased from 0.0561 to 0.0881 when it is applied to centralized deployment with 6 PMUs. When using 3 PMUs, $MAPE(\delta)$ of the developed method doubles for both distributed and centralized deployments. A similar situation occurs in the comparison method FNO. Compared to transient dynamics trajectory prediction, the impact of changes in deployment patterns on fault location identification is more pronounced. For example, the accuracy of the developed method is higher than 0.83 when applied to distributed deployment, but lower than 0.8 when applied to centralized deployment. For the comparison method CNN-LSTM, the accuracy in all deployment patterns is lower than 0.8. Based on the above analysis, we found that the prediction accuracy of centralized deployment is lower than distributed deployment, so when conducting DSA for large-size systems, we recommend using distributed distribution and the number of PMUs should not be too small.

V. CONCLUSION

This paper develops an advanced online DSA, which is capable of processing multi-source spatial-temporal data and generating comprehensive DSA results, i.e., transient dynamics trajectory prediction and fault location identification. The SCADA and PMU data are modeled as different graphs,

TABLE VI
DSA RESULTS FOR DIFFERENT PMU DEPLOYMENTS

Deployment pattern	PMU number	Methods	Transient Dynamics trajectory prediction				Fault location identification			
			MAPE(δ)	MAPE(ω)	MSE(δ)	MSE(ω)	F1	Precision	Recall	Acc
Distributed deployment	6 (Bus 12,31,59,65,87,111)	G_Φ - G_Ψ -FNO(4)	7.23%	0.16%	0.0561	$5.60e-5$	0.98	0.99	0.98	0.98
		FNO	7.76%	0.19%	0.0842	$6.36e-5$	-	-	-	-
		CNN-LSTM	-	-	-	-	0.77	0.78	0.77	0.77
	3 (Bus 31,59,87)	G_Φ - G_Ψ -FNO(4)	13.51%	0.16%	0.0712	$6.08e-5$	0.81	0.82	0.83	0.83
		FNO	16.14%	0.21%	0.1034	$7.40e-5$	-	-	-	-
		CNN-LSTM	-	-	-	-	0.61	0.63	0.61	0.61
Centralized deployment	6 (Bus 75,79,85,98,103,108)	G_Φ - G_Ψ -FNO(4)	7.26%	0.16%	0.0881	$5.00e-5$	0.71	0.70	0.73	0.73
		FNO	9.41%	0.16	0.1207	$4.95e-5$	-	-	-	-
		CNN-LSTM	-	-	-	-	0.64	0.66	0.64	0.64
	3 (Bus 75,79,103)	G_Φ - G_Ψ -FNO(4)	14.47%	0.15%	0.0683	$5.51e-5$	0.78	0.78	0.80	0.80
		FNO	14.63%	0.28%	0.0984	$1.08e-5$	-	-	-	-
		CNN-LSTM	-	-	-	-	0.57	0.59	0.58	0.58

which are processed by two GNNs to extract distinct spatial-temporal features. Then, an FNO-based transient dynamics trajectory predictor and an MLP-based fault location identifier are developed to learn system dynamics and fault patterns from spatial-temporal features, respectively.

Case studies on the IEEE 39 Bus System and IEEE 118 Bus System considering uncertainties from RES, loads, and unexpected faults validate the effectiveness of the developed online DSA. Comparisons of the developed model, LSTM-RNN, FNO, and structure-informed graph learning, reveals that the developed online DSA exhibits superior transient dynamics trajectory performance, characterized by the minimum prediction errors. Another comparison demonstrates that the developed method outperforms DT, SVM, LSTM-SAF, and CNN-LSTM in terms of f1-score, recall, and accuracy, thereby validating its superior fault location identification capability. Considering different system configurations exist in real applications, the developed method is tested using different deployment distributions and available PMU numbers. The results show that the developed model is robust to different PMU deployments. Data imprecise scenarios have been simulated to test the model robustness. The results show that transient dynamics trajectory performance has little deviation (e.g., $\Delta MAPE(\delta) = 1.97\%$) under data missing and noise scenarios. Model scalability has been validated using IEEE 118 Bus System.

While the developed method demonstrates promising performance in baseline scenarios, two key limitations are identified through case studies: (1) Performance degradation under high-noise conditions (e.g., about 20% fault location identification accuracy drop at SNR = 30 dB) due to error propagation in spatial-temporal feature extraction; (2) Sensitivity to PMU numbers, with transient dynamics trajectory prediction errors increasing by 17.8% when 3 PMU utilized. Additionally, the current framework does not account for cybersecurity threats inherent to power system, such as adversarial data manipulation. Future research will focus on integrating physics-informed noise filters to enhance robustness, developing topology-aware data augmentation for sparse PMU scenarios, and embedding graph-based anomaly detection modules to mitigate cyber-physical attack risks.

REFERENCES

- [1] G. Lu and S. Bu, "Online Dynamic Security Assessment: Using Hybrid Physics-Guided Deep Learning Models," *IEEE Trans. Ind. Inform.*, vol. 20, no. 11, pp. 13190-13200, Nov. 2024.
- [2] M. Kabiri and N. Amjadi, "A new hybrid state estimation considering different accuracy levels of PMU and SCADA measurements," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3078-3089, Sep. 2019.
- [3] J. Hu, Q. Wang, Y. Ye and Y. Tang, "High-Resolution Real-Time Power Systems State Estimation: A Combined Physics-Embedded and Data-Driven Perspective," *IEEE Trans. Power Syst.*, vol. 40, no. 2, pp. 1532-1544, March 2025.
- [4] T. Zhao, M. Yue and J. Wang, "Structure-Informed Graph Learning of Networked Dependencies for Online Prediction of Power System Transient Dynamics," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4885-4895, Nov. 2022.
- [5] X. Ye, A. Radovanović and J. V. Milanović, "The Use of Machine Learning for Prediction of Post-Fault Rotor Angle Trajectories," *IEEE Trans Power Syst.*, vol.39, no.5, pp.6496-6507, Sept. 2024.
- [6] Q. Wang, S. Bu, Z. He and Z. Y. Dong, "Toward the Prediction Level of Situation Awareness for Electric Power Systems Using CNN-LSTM Network," *IEEE Trans. Ind. Inform.*, vol.17, no.10, pp.6951-6961, 2021.
- [7] J. Hu et al., "A High Temporal-Spatial Resolution Power System State Estimation Method for Online DSA," *IEEE Trans Power Syst.*, vol. 39, no. 1, pp. 877-889, Jan. 2024.
- [8] G. Lu and S. Bu, "Adaptive Stability Contingency Screening for Operational Planning Based on Domain-Adversarial Graph Neural Network," *IEEE Trans Power Syst.*, vol. 39, no. 1, pp. 1503-1516, Jan. 2024.
- [9] L. Lu et al., "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators". *Nature mach. intell.*, vol.3, no.3, pp.218-229,2021.
- [10] Z. Li et al., "Fourier neural operator for parametric partial differential equations". *arXiv preprint arXiv:2010.08895*, 2020.
- [11] C. Moya, S. Zhang, M. Yue, and G. Lin, "Deepnet-grid-ug: A trustworthy deep operator framework for predicting the power grid's post-fault trajectories," *Neurocomputing*, vol.535, pp. 166-182, 2023.
- [12] Y. Sun, C. Moya, G. Lin, and M. Yue, "DeepGraphONet: A Deep Graph Operator Network to Learn and Zero-Shot Transfer the Dynamic Response of Networked Systems," *IEEE Syst. J.*, vol. 17, no. 3, pp. 4360-4370, Sept. 2023.
- [13] W. Cui, W. Yang and B. Zhang, "A Frequency Domain Approach to Predict Power System Transients," *IEEE Trans Power Syst.*, vol. 39, no. 1, pp. 465-477, Jan. 2024.
- [14] P. Kundur, *Power system stability and control*. New York: McGraw-Hill, 1994
- [15] S. Xia, S. Bu, and J. Hu, "Efficient transient stability analysis of electrical power system based on a spatially paralleled hybrid approach," *IEEE Trans. Ind. Inform.*, vol.15, no.3, pp.1460-1473, 2019.
- [16] S. Cheng et al., "Power System Transient Stability Assessment Based on the Multiple Paralleled Convolutional Neural Network and Gated Recurrent unit," *Protection Control Modern Power Syst.*, vol. 7, no. 3, pp. 1-16, July 2022
- [17] T. Huang, S. Gao and L. Xie, "A Neural Lyapunov Approach to Transient Stability Assessment of Power Electronics-Interfaced Networked Microgrids," *IEEE Trans Smart Grid*, vol.13, no.1, pp.106-118, Jan. 2022.
- [18] J. Zhao et al., "Power System Real-Time Monitoring by Using PMU-Based Robust State Estimation Method," *IEEE Trans Smart Grid*, vol. 7, no. 1, pp. 300-309, Jan. 2016
- [19] C. Ren and Y. Xu, "A Fully Data-Driven Method Based on Generative Adversarial Networks for Power System Dynamic Security Assessment With Missing Data," *IEEE Trans Power Syst.*, vol.34, no 6, pp.5044-5052, Nov. 2019

- [20] Y. Shi et al. "Masked label prediction: Unified message passing model for semi-supervised classification". arXiv preprint arXiv:2009.03509. 2020.
- [21] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization". arXiv preprint arXiv:1607.06450. 2016.
- [22] W. Hu, et al., "Strategies for pre-training graph neural networks". arXiv preprint arXiv:1905.12265. 2019.
- [23] J. Du et al., "Topology adaptive graph convolutional networks". arXiv preprint, arXiv:1710.10370, 2017.
- [24] N. Kovachki et al., "Neural operator: learning maps between function spaces with applications to PDEs". *J. Mach. Learn. Res.*, vol.24, no.89, pp.1-97, 2023.
- [25] X. Peng et al., "Short-Term Wind Power Prediction Based on Wavelet Feature Arrangement and Convolutional Neural Networks Deep Learning," *IEEE Trans. Ind. Appl.*, vol.57, no.6, pp.6375-6384, Nov.-Dec. 2021.
- [26] T. Guo and J. V. Milanovic, "Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 377–385, Jan. 2014.
- [27] L. Moulin, A. A. Da Silva, M. El-Sharkawi, and R. J. Marks, "Support vector machines for transient stability analysis of large-scale power systems," *IEEE Trans. Power Syst.*, vol.19, no.2, pp.818–825, May 2004.
- [28] Z. Shao et al., "A Novel Data-Driven LSTM-SAF Model for Power Systems Transient Stability Assessment," *IEEE Trans. Ind. Inform.*, vol. 20, no. 7, pp. 9083-9097, July 2024.



Genghong Lu received the Ph.D. degree in Control Science and Engineering in Zhejiang University, Hangzhou, China, in 2021. From 2018 to 2019, she worked as a visiting doctoral scholar at University of Alberta, Edmonton, Canada. She is currently a postdoctoral fellow in Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong. Her research interest includes data analytics, deep learning, and power system stability analysis.



Siqi Bu (S'11-M'12-SM'17) received the Ph.D. degree from the electric power and energy research cluster, The Queen's University of Belfast, Belfast, U.K., where he continued his postdoctoral research work before entering industry. Then he was with National Grid UK as an experienced UK National Transmission System Planner and Operator. He is currently a Professor and Associate Head with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Kowloon,

Hong Kong, Associate Director of Research Centre for Grid Modernisation, and a Chartered Engineer with UK Royal Engineering Council, London, U.K.. His research interests include power system stability, operation and economics considering renewable energy integration, smart grid application and transport electrification.

Dr Bu is an Editor of IEEE Transactions on Power Systems, IEEE Transactions on Consumer Electronics, IEEE Power Engineering Letters, IEEE Open Access Journal of Power and Energy, CSEE Journal of Power and Energy Systems, Protection and Control of Modern Power Systems, Journal of Modern Power Systems and Clean Energy, and Advances in Applied Energy. He is a Fellow of IET, Chairman of IET HK Power and Energy Section, Co-Chairman of IET DPSP 2025 and APSCOM 2025, and technical chairman of IEEE PESIM 2026.