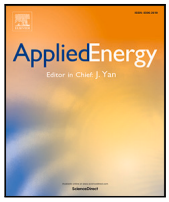
Applied Energy 365 (2024) 123252

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



ST-AGNet: Dynamic power system state prediction with spatial–temporal attention graph-based network

Shiyao Zhang [a], Shuyu Zhang [a,*], James J.Q. Yu [b], Xuetao Wei [c]

[a] Research Institute for Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, 518055, China
[b] Department of Computer Science, University of York, York, YO10 5GH, United Kingdom
[c] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

ARTICLE INFO

Keywords:
Dynamic power system state
Graph-based network
Multi-time-scale prediction
Spatial–temporal relationship

ABSTRACT

Accurate and timely prediction of power system states is one of the most important challenging tasks in modern power systems. Considering the integration of renewable energy sources, recent deep learning-based models have been well studied and found to have benefits in exploiting spatial–temporal relationships in power system data. However, the complexity of different power system topology structures is not substantially captured since the existing models did not fully consider the graph-based information retrieved from power networks. To resolve the problem, a spatial–temporal attention graph-based network (ST-AGNet), an adaptive power system state prediction approach that utilizes graph-based information data to account various typologies of complex power systems, is proposed. Initially, the power flow model is used for generating historical system state data. With the graph-based topology information, the input dataset with the spatial and temporal features is fed into the proposed network for the training and validating process. Meanwhile, the connectivity of the time-varying graph-based information are accounted in the proposed model. Case studies demonstrate the superiority of the ST-AGNet model over the existing baselines under four different scales of complex systems, which can significantly support dynamic power system analysis and operational tasks.

1. Introduction

In modern society nowadays, the emerging smart grid technology is capable of providing innovative solutions to the development of smart cities. In comparison to conventional smart grid systems, the recent smart grid systems are more diversified due to technological advancements, e.g., the increase of the utilization of various renewable energy sources (RESs) and loads. However, the system uncertainty issues still occur due to the erratic and intermittent nature of RESs and loads. Although the use of RESs can help to reduce the environmental effects of global warming, the smart grid system may experience rapid instability status, thereby affecting the system state variables. To address this issue, it is essential to estimate the time-varying changes of the system states so as to tackle the system uncertainties caused by RESs and loads and further support the system operations and services. Thus, a reliable system state prediction is considered an indispensable factor in modern smart grid systems, which can benefit tasks such as power control and energy management [1,2].

There are a number of research studies investigating system state prediction for decades by using traditional statistical methods, e.g., an extended Kalman filter [3], statistical Gaussian mixture model [4], and non-parametric prediction method [5]. Owing to the acquisition of massive dynamic system state data in dynamic power systems, data-driven machine learning approaches become emerging solution to address potential non-linear features in the system. In particular, deep learning approaches, as a subset of machine learning approaches, are capable of extracting latent features through neural networks. The precise prediction results are beneficial to assist the power system operations for the purpose of stability assessment, e.g., [6–9]. Owing to the nature of time-varying RESs and loads, the implementation of long short-term memory (LSTM)-based tools can accurately capture the dynamic temporal features [8,9]. Additionally, to facilitate the spatial measurement data, the use of convolutional neural network (CNN)-based models can help handle spatial features in power systems, e.g., network topology considerations [6]. However, CNNs fail to achieve precise predictions due to the changes in different complex power system graph structures.

To address this point, graph neural networks (GNNs) become better options than CNNs for capturing non-Euclidean spatial characteristics from power networks due to the fact that power networks are generally formed as graph-structured information with high-dimensional

* Corresponding author.
E-mail address: 11712122@mail.sustech.edu.cn (S. Zhang).

https://doi.org/10.1016/j.apenergy.2024.123252
Received 1 February 2024; Received in revised form 14 March 2024; Accepted 12 April 2024
Available online 25 April 2024

characteristics and node inter-dependency, where common buses on power systems are represented as nodes, and transmission lines are represented as edges. To account for the complexity of graph-structured data, deep graph neural networks have been the key tools of numerous recent research publications [10]. As the massive spatial–temporal system state data was acquired, several existing researchers designed the models by combining recurrent neural networks (RNNs) and graph convolutional neural networks (GCNs), e.g., [11–14].

However, the aforementioned research studies face notable challenges in the followings. First of all, the existing graph structure cannot fully represent the adjacency of the common buses in complex power systems. For example, the peaks and valleys of industrial and household electricity usage are different. In this case, the relationship between generator buses and load buses changes periodically over time. Then, the effect of RESs is not sufficiently considered in the time-varying complex power systems, whereas most studies unilaterally implement homogeneous RESs. Furthermore, how to mine and utilize the spatial and temporal dependencies of time-varying system states is still an important question that needs to be answered. Last but not least, the performance cannot be well optimized under complex power systems. To address these challenges, we design a spatial–temporal attention graph-based network (ST-AGNet) model to perform dynamic system state prediction in power systems. The main contributions are shown as follows.

1. Different from [6–9], that fail to capture the complexity of different power system topology structures, we design a generic ST-AGNet model by means of the graph-based information to better excavate power networks with temporal properties.
2. Unlike the existing studies [11–13] that have explicitly integrated the common bus (node) states with fixed topological characteristics and temporal patterns, we propose an adaptive graph to implement the adjacency of the common buses for short-term multi-time-scale power system state predictions with the consideration of heterogeneous RESs.
3. The self-attention mechanism in the proposed network is applied to fuse the hidden spatial and temporal dependency at different scales of feature modules to estimate the power states.
4. We evaluate the prediction results of the designed ST-AGNet model through comprehensive case studies under different power network topologies with other methods, which confirm its optimal performance.

The following structure of this paper is organized below. The related work is introduced in Section 2. Section 3 shows the system model, as well as the utilized power flow model. After that, the ST-AGNet model is then illustrated in Section 4. We assess the performance of the proposed model under various power systems in Section 5. In Section 6, we finally conclude this work.

2. Related work

In this section, we review previous efforts on dynamic power system state prediction. In addition, the related traditional and filtering methods are presented, as well as the recent deep learning approaches.

2.1. Traditional power system state prediction

The application of power system state prediction has been extensively studied in recent years. Some conventional methods are adopted in this research field, such as [3–5,15–17]. For instance, in [3], an extended Kalman filter was proposed to monitor system state dynamics rapidly and accurately. Considering the RESs, the real-time system measurements were estimated by using a statistical Gaussian mixture model developed in [4]. Besides, to identify the state condition that described the system's instability problem, a non-parametric prediction tool was adopted in [5]. For non-linear characteristics of power systems, [15] adopted a multi-area power system with non-linear measurements completely by distributed robust bi-linear state-estimation approach. Then, [16] could address the ambiguity of the power system parameters by a novel proposed mean squared estimator. Furthermore, [17] proposed a distributed online estimation approach for tracking combined heat and power system states. However, these traditional tools only address apparent linear or non-linear system patterns, neglecting the effect of hidden non-linear characteristics in power systems.

2.2. Time-series deep learning for power system state prediction

As the complexity of power networks increases, deep learning methods can be a good candidate to fully excavate the hidden non-linear spatial and temporal features. There are several existing studies demonstrating the effectiveness of deep learning approaches, e.g., [6–9]. Considering the temporal features in power systems, [7] compared the widely used weighted least squares state estimation approaches and developed a learning-based method that attempted to provide efficient, robust, and extensive observability. Besides, [8] proposed a swarm intelligence fusion model that combined data and physical-driven approaches, which the LSTM network was employed to tackle temporal features. In addition to the temporal features, the spatial features are also necessary for system state prediction tasks. In this case, [6] proposed a multi-level strategy for false data identification so as to improve dynamic state prediction, whereas the three levels used an invention vector, LSTM, and CNN. By implementing the network topology as one crucial spatial feature, [9] developed a Bayesian LSTM network to estimate the system states in various scales of systems. However, these recent studies destroy the non-Euclidean structure of power networks.

2.3. Graph learning-based prediction

Graph-based neural networks, a more modern substitute, offer a superior option for maintaining the connectivity and adjacency information of power networks. Several existing researches have investigated the use of graph-based neural networks on predicting dynamic system states, e.g., [11–13,18,19]. For instance, [11] applied spatial–temporal graph learning method to serve the power system physics, whereas this work designed graph shift operator based on graph recursive neural network (GRN) and GCN. Considering to capture the temporal features, in [12], a recurrent graph convolutional network (RGCN) was proposed to predict the system states, and it explicitly integrated the bus (node) states with the structure information. Additionally, [14,18] adopted graph neural networks to estimate the solar and wind output, respectively, in which these two studies both accounted for local spatial–temporal correlations of the power networks. Furthermore, [13] devised a two-layer algorithm with event-insensitive Graph Attention Network (GAT) and LSTM to monitor the power system state and pinpoint the locations of events. In this work, phasor measurement units (PMUs) data with the magnitude and phase angle of an electrical phasor are converted into two-dimensional images with frequency information for LSTM, while CNN is used to improve the feature gap between various occurrences. Last but not least, a novel graph-learning framework developed in [19], called deep-learning neural representation, was proposed to predict the system state trajectories in real-time. However, existing studies overlook the dynamic adjacency information of common buses from a multi-timescale perspective. Additionally, the utilization of auxiliary information like RESs and power loads as well as effective feature fusion modeling remain open questions. In this work, we utilize the graph-based information by including the adjacency information of the common buses to perform multi-time-scale predictions on dynamic power system states, and consider the impact of RESs and load information.

3. Power flow model

In this section, we introduce and illustrate the mathematical derivations of dynamic system state in a general power system, including the formulated AC power flow model. The details are shown as follows.

3.1. Dynamic power system state

Based on the graph theory, suppose $\mathcal{N}$ be the index set of nodes and $\mathcal{N}$ be the index edges in the system. The entire power system network can be modeled as a directed graph, in which it is modeled as $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. In this case, $\mathcal{N}$ denotes the set of common buses and $\mathcal{E}$ represents the set of transmission lines. By taking into account nodal equations, the nodal admittance matrix can be expressed as $\mathcal{Y}$, and it follows:

$$\mathcal{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1|\mathcal{N}|} \\ y_{21} & y_{22} & \cdots & y_{2|\mathcal{N}|} \\ \vdots & \vdots & \vdots & \vdots \\ y_{|\mathcal{N}|1} & y_{|\mathcal{N}|2} & \cdots & y_{|\mathcal{N}||\mathcal{N}|} \end{bmatrix}, \tag{1}$$

where $|\mathcal{N}|$ is the total number of network nodes. For each element y_{ij}, it can be calculated by $y_{ij} = \mathbf{g}_{ij} + j\mathbf{b}_{ij}$, where $\mathbf{g}_{ij} \geq 0$ and $\mathbf{b}_{ij} \leq 0$, are the conductance and susceptance of the transmission line $(i, j) \in \mathcal{E}$, respectively. Besides, the related shunt capacitance is denoted as $c_{ij} = c_{ji}$ and it is involved when obtaining (1).

In addition, we define the apparent local power inflow of each common bus i as $S_i = P_i + jQ_i$, where it includes both active power inflow P_i and reactive power inflow Q_i. The local power inflows are tightly related to the power generation dispatch. The system operational time period is set as $\mathcal{T}$, which is divided into a set $\mathcal{T} = \{1, \ldots, |\mathcal{T}|\}$ of $|\mathcal{T}|$ time slots. Since the system covers both the conventional and renewable generations for the active power generations, we designate the active and reactive power generation as P_t^G and Q_t^G at time $t \in \mathcal{T}$ while P_t^L and Q_t^L are set as the active and reactive loads, respectively. Thus, we have:

$$P_{i,t} = P_{i,t}^G - P_{i,t}^L, \ \forall i \in \mathcal{N}, t \in \mathcal{T}, \tag{2}$$

$$Q_{i,t} = Q_{i,t}^G - Q_{i,t}^L, \ \forall i \in \mathcal{N}, t \in \mathcal{T}. \tag{3}$$

Furthermore, for any two i, j buses in the system, we denote the complex power flow as $S_{ij} = P_{ij} + jQ_{ij}$, in which it covers both active power flow P_{ij} and reactive power flow Q_{ij}. The relationship between the complex power inflow at common bus i and its nearby power flow is represented as:

$$S_i(\mathcal{T}) = \sum_{j \in \Omega_i} S_{ij}(\mathcal{T}), \ \forall i, j \in \mathcal{N}, \tag{4}$$

where Ω_i denotes the set of nearby transmission lines connected at common bus i.

For the purpose of the power flow analysis, the essential elements are utilized to be determine, including each common bus's voltage magnitude, phase angle, active power inflow, and reactive power inflow. We represent the common bus voltage magnitude at common bus i as $|V_i|$ with its phase angle as θ_i. In this case, a total of $2|\mathcal{N}| - 1$ state variables are explicitly included in the system's common bus voltage, comprising $\mathcal{N}$ voltage magnitudes and $\mathcal{N} - 1$ phase angles, where each common bus's phase angle is specified in regard to the common bus reference. Therefore, we define the state vector of the common bus i as x_i^T, in which it includes the four main essential elements in the system, shown as:

$$x_{i,t}^T = \begin{bmatrix} |V_{i,t}| & \theta_{i,t} & P_{i,t} & Q_{i,t} \end{bmatrix}, \ \forall i \in \mathcal{N}, t \in \mathcal{T}, \tag{5}$$

where $(\cdot)^T$ is the operator for the transpose of the matrix. Based on (5), the dynamic state matrix is shown as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1|\mathcal{T}|} \\ x_{21} & x_{22} & \cdots & x_{2|\mathcal{T}|} \\ \vdots & \vdots & \vdots & \vdots \\ x_{|\mathcal{N}|1} & x_{|\mathcal{N}|2} & \cdots & x_{|\mathcal{N}||\mathcal{T}|} \end{bmatrix}. \tag{6}$$

3.2. Power flow analysis

The AC power flow model is used to obtain the real measurements for the system states. As previously noted, the dynamic state matrix in (6) is developed to aggregate with all features of the input dataset.

The assessment of a general power system state involves the utilization of efficient power flow model. There are three typical AC power flow models utilized, shown as:

- Newton–Raphson power flow model [20];
- Gauss–Seidel power flow model [21];
- Fast decoupled power flow model [22].

By comparing the above three methods, since this work considers a complex power network system, it is demonstrated that the suitable model is Newton–Raphson power flow model because it can compute accurate solutions on large complex systems.

By using the power flow model, the non-linear equation is used to represent the fulfillment of the AC power balance condition, shown as

$$g(\mathbf{X}) = \begin{bmatrix} \Delta\mathbf{P} \\ \Delta\mathbf{Q} \end{bmatrix} = \mathbf{0}, \tag{7}$$

where $\mathbf{0}$ is the matrix with zero values.

The overview of solving Newton–Raphson power flow problem is introduced in the followings. For each iteration k, $\Delta\mathbf{P}^k$ and $\Delta\mathbf{Q}^k$ are gained by following [23]. Then, the system Jacobian matrix is calculated so as to compute $\Delta|\mathbf{V}^k|$ and $\Delta\boldsymbol{\theta}^k$. Lastly, the dynamic state matrix (6) is obtained. During the iteration procedure, the tolerance of the convergence is checked. Once the stopping criterion $\epsilon \leq 10^{-5}$ is met, the final solution $\mathbf{X} = \mathbf{X}^{k^*}$ is updated when k^* is the value of the last iteration number.

3.3. Problem definition

We define the state prediction as a spatial–temporal prediction task, considering the spatial structure and temporal dependency of the power system. As mentioned in Section 3.1, the power system network is modeled as a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and the related problem is formulated as

$$\left[X_{t-\mathcal{T}+1}, \ldots, X_t; \mathcal{G}; \mathcal{R}; \mathcal{Y}\right] \xrightarrow{f(\cdot)} \left[\hat{X}_{t+1}, \ldots, \hat{X}_{t+\mathcal{H}}\right] \tag{8}$$

where $\mathcal{H}$ is the length of the predicted data, and $f(\cdot)$ is the prediction function.

4. Spatial–Temporal Attention Graph-based Network (ST-AGNet)

In this section, we begin by introducing the overall structure of the proposed model, referred to as the ST-AGNet. The primary focus of ST-AGNet lies in leveraging temporal convolutional operations and adaptive graph neural networks to extract temporal and spatial features from the states in the power system. Following that, we delve into the design principles and roles of each module within ST-AGNet, alongside an explanation of the training approach employed for the proposed model.

4.1. ST-AGNet

This study introduces an advanced model named the Spatial-Temporal Attention Graph-based Network (ST-AGNet) to represent the spatial–temporal relationship of the power system state. The model effectively captures the changing dynamics of the system states in the power system, taking into account the influence of various renewable energy sources (RESs) and the fixed relationship between common buses in complex power systems. The overall structure of ST-AGNet is shown in Fig. 1.

In each layer, the feature module takes the power system states as input, and then pairs the feature dimensions through convolution

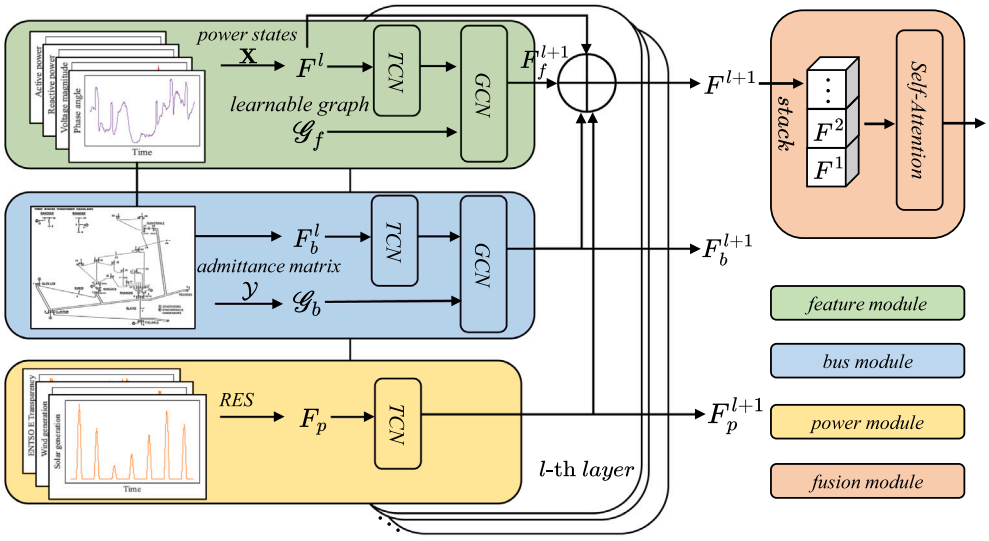


Fig. 1. Overall structure of the proposed ST-AGNet. The proposed ST-AGNet has a stacked structure. In each layer, there are three main modules, namely, the feature module, bus module, and power module, for each kind of statistical data, including historical states, the admittance matrix, and RES from the power system. In addition, a fusion module for combining the hidden features is used to predict the states in the future.

operations. After that, it performs feature extraction using Temporal Convolutional Network (TCN) [24] and GCN operations for temporal and spatial feature extraction, respectively. To better capture changing system states in a day, we construct a learnable adjacency matrix as input to the GCN by building separate relation matrices for different times of the day and back-propagating to obtain the relation matrix across all times. Similar to the feature module, the bus module also uses TCN and GCN for spatial–temporal feature extraction, but the input to its graph convolution is the bus relation matrix that is derived from the transmission line matrix $\mathcal{Y}$. Given the timing dependence of RESs and loads, the power module performs feature extraction solely using TCN. After extracting the hidden states from the above three modules, we proceed to combine the resultant feature sets, namely F_f^{l+1}, F_b^{l+1}, and F_p^{l+1}, in order to derive the output for the layer, denoted by F^{l+1}. We also apply skip connections to avoid vanishing gradients when the proposed network is sufficiently deep, which is formulated as:

$$F^{l+1} = F_f^{l+1} + F_b^{l+1} + F_p^{l+1} + F^l. \tag{9}$$

Subsequently, a fusion module, equipped with a self-attention mechanism and linear layers, combines the hidden features from each layer to make predictions about future states.

4.2. Feature module

In power systems, the power flows between buses can vary based on the time of day. For instance, the peaks and valleys in industrial and household loads follow different patterns. As a result, the relationships between generator and load buses change periodically over time. To enable multi-timescale predictions of power system states, we construct a adaptive graph to model the time-varying adjacency between common buses.

In the feature module, we first apply TCN to extract temporal features from the historical states. TCN is a variant of CNN that can capture long-term dependencies in time series data. Compared to recurrent neural networks used on traditional time-series tasks [25], the TCN layer with dilated operation [26] is able to reduce time complexity significantly. Besides, the gating mechanism demonstrates proficiency in modeling sequence data, as it is employed in the temporal convolution layer to enhance the capability of the model. In particular, the formulation of TCN for the lth layer in the feature module is in the form:

$$H_f^l = \tanh\left(W_f^l * F^l\right) \odot \sigma\left(W_g^l * F^l\right), \tag{10}$$

where F^l is the output feature of the lth layer, which combines the feature from three modules from previous layer, W_f^l and W_g^l are the convolutional kernels of the temporal convolution layer, $\odot$ symbolizes the element-wise multiplication, $*$ represents the dilated convolution operation, and σ denotes the sigmoid function.

The spatial relationship between the buses is also crucial for the power prediction task. In this way, we adopt GCN to extract spatial features from the states. Different from the traditional GCN, we construct a learnable adjacent matrix to be used as the input for GCN, i.e., we build a learnable relation matrix $\mathcal{G}_f = \{\mathcal{G}_{f1}, \mathcal{G}_{f2}, \ldots, \mathcal{G}_{fT}\}$ for T moments of a day separately. For example, if the sampling interval of the sensor is 15 min, there are $T = 96$ time slots in a day, and we will train 96 adjacency matrices $\mathcal{G}_f$ for different time slots for the subsequent task. Then, we update the relation matrix $\mathcal{G}_f$ for different moments via back-propagate. The formulation of GCN for the lth layer in the feature module is in the form:

$$F_f^{l+1} = \sum_{u=0}^{U} \left(\mathcal{G}_{ft}\right)^n H_l^t W^u, \tag{11}$$

where F_{l+1}^t is the output of the lth layer in feature module, $\mathcal{G}_{f,t}$ is the relation matrix at time t, W^u is the weight matrix of the uth order, and U is the order of the graph convolution.

4.3. Bus module

Similarly, the bus module uses TCN and GCN for spatial–temporal feature extraction. However, the input to its graph convolution is the bus relation matrix $\mathcal{G}_b = \{\alpha_{ij}\}_{\mathcal{N}\times\mathcal{N}}$ derived from the transmission line matrix $\mathcal{Y}$. Specifically, different from the feature module, spatial feature extraction is mainly based on the physical connection relationship between different buses in the power system. And the connectivity between ith bus and jth bus α_{ij} is defined as:

$$\alpha_{ij} = W_{ij} * (\mathbf{g}_{ij}, \mathbf{b}_{ij}) + b_{ij}, \tag{12}$$

where W_{ij} and b_{ij} are the weight and bias of the embedding layer of the ij-th element in $\mathcal{Y}$. Note that $\mathbf{g}_{ij}$ and $\mathbf{b}_{ij}$ are the conductance and susceptance of the ij-th transmission line, respectively. Using this adjacency matrix, a GCN is then applied to extract spatial features from the states.

4.4. Power module

For the dynamic power system state prediction, power information shall provide auxiliary information. Given the timing dependence of RESs, the power module first performs feature expansion through linear operations. The formulation is:

$$H_l^l = W_l F_l^l + b_l, \tag{13}$$

where $W_l \in \mathbb{R}^{f_p \times N}$ and b_l are the weight and bias of the lth layer in the power module, f_p is the feature dimension of the power information, and N is the number of buses. Then, the TCN is used to extract temporal features from the hidden states H_l^l.

4.5. Fusion module

In order to combine the spatial–temporal features obtained from the feature module, bus module, and power module, we provide a fusion module. This module incorporates a self-attention mechanism and linear layers to merge the hidden features from each layer and make predictions about future states. Infusion module, to leverage the hidden feature in various levels, we first concatenate the output hidden features from different layers as $\mathcal{H} = [F_f^1, F_f^2, \ldots, F_f^l]$, l is the number of layers, and then apply a self-attention mechanism to fuse the hidden features. The self-attention was first proposed in [27] for sequence data. It has been widely used in different areas, like natural language processing, computer vision, and time-series data. In this work, we apply the self-attention mechanism to fuse the hidden features of each layer. The computation of the self-attention mechanism involves three components: the query matrix $\mathcal{Q}$, the key matrix $\mathcal{K}$, and the value matrix $\mathcal{V}$. These matrices are obtained by multiplying the input matrix $\mathcal{H}$ with the corresponding weight matrices W_q, W_k, and W_v, respectively. The output $\mathcal{Z}$ of self-attention mechanism can be computed by:

$$\mathcal{Z} = \mathrm{softmax}(\frac{\mathcal{Q}\mathcal{K}^T}{\sqrt{d_k}})\mathcal{V}, \tag{14}$$

where d_k is the dimension of the key matrix $\mathcal{K}$. After that, we apply the linear layers to fuse the output and predict the dynamic system states in the future.

4.6. Loss function

When training a deep learning model, its parameters are refined to optimize its performance by minimizing the loss function. This process is facilitated by optimization algorithms like gradient descent, which iteratively adjust the model's parameters in accordance with the gradient of the loss function. Moreover, the loss function serves as a quantitative metric of our model's performance. It steers the optimization process towards reducing the gap between the model's predictions and the actual states, thereby improving the accuracy of the model's predictions.

Following the previous deep learning-based models [9,11] for the dynamic power system state prediction task, we apply the mean squared error (MSE) as the loss function to optimize the parameters of the proposed ST-AGNet model, which is defined as

$$\mathcal{L} = \frac{1}{\mathcal{N}}\frac{1}{\mathcal{H}}\sum_{i=1}^{\mathcal{N}}\sum_{t=1}^{\mathcal{H}}(\hat{x}_{i,t} - x_{i,t})^2, \tag{15}$$

where $\mathcal{N}$ is the number of buses, $\mathcal{H}$ is the number of predicted time slots, $\hat{x}_{i,t}$ is the predicted state of the ith bus at the tth time slot, and $x_{i,t}$ is the corresponding ground truth state.

5. Case studies

This section presents case studies demonstrating the proposed approach. First, the configurations of the test systems are introduced. Next, we present the baseline as well as the prediction performance. To examine the roles of different modules, ablation experiments were conducted to evaluate their impacts. Furthermore, given practical applications, we assessed the time costs of the proposed model. Finally, the influence of noise on prediction accuracy is also discussed.

5.1. Test system configurations

In the simulation, we adopt four different scales of power systems [28] for case studies, including IEEE 30-Bus, IEEE 57-Bus, IEEE 118-Bus, and IEEE 145-Bus systems. For each power system, we extract the information of buses and bus lines following the IEEE Common Data Format.[1] The details of the four test systems are as follows:

- **IEEE 30-Bus System:** The system comprises a total of 30 buses and 41 transmission lines. The solar and wind generators are deployed at Bus 2 and 10 in the test instances, respectively.
- **IEEE 57-Bus System:** The system comprises a total of 57 buses and 80 transmission lines. The solar and wind generators are deployed at Bus 13 and 37 in the test instances, respectively.
- **IEEE 118-Bus System:** The system comprises a total of 118 buses and 186 transmission lines. The solar and wind generators are deployed at Bus 49 and 69 in the test instances, respectively.
- **IEEE 145-Bus System:** The system comprises a total of 145 buses and 153 transmission lines. The solar and wind generators are deployed at Bus 17 and 69 in the test instances, respectively.

For the four systems, buses with the most transmission lines connected were selected for installing solar and wind generators from different areas. Additionally, historical power system data from [29] was utilized. Specifically, time-varying system data over one year from 2016-01-01 to 2016-12-31 was used. The data was partitioned into three distinct sets — training, validation, and test sets, comprising 60%, 20%, and 20% of the data by time, correspondingly. This separating procedure is evaluated through cross-validation. The time interval of sampling period of the system state is 15 min. In addition, sample construction is performed separately on each dataset individually by a sliding window. After that, the data is normalized to the range of $[0, 1]$ by Z-score. Furthermore, the adopted time-varying generations and loads are scaled to fit the above four test systems.

For comparison purposes, unless stated otherwise, all case studies utilized the following configuration. The proposed ST-AGNet and baseline models were implemented in Python with PyTorch. For deep learning methods, models were trained for 100 epochs with a batch size of 50, using a learning rate of $1e^{-3}$. We utilized the Adam optimizer for model optimization, and implemented early stopping to mitigate the risk of overfitting. All tests were run on servers with an Intel(R) Xeon(R) E5-2620 v4 CPU and nVidia GeForce RTX 2080 Ti GPUs.

5.2. Baseline models

In this work, we aim to develop sophisticated neural network architectures tailored for power systems. In this way, except for the advanced models proposed in the past years for states prediction, like LSTM, LSTM-CNN, STGCN. We also have selected the most exemplary models for spatio-temporal prediction and system state forecasting from a variety of sectors such as power, energy, and transportation, which have shown promise in recent years. Specifically, we compare the proposed method with the following baselines:

[1] http://labs.ece.uw.edu/pstca/formats/cdf.txt

- **GW-Net (2019)** dynamically captures spatial dependencies by learning an adaptive dependency matrix through node embedding [30].
- **Long Short-Term Memory (LSTM, 2021)** employs a stacked LSTM network to take the time-series input and predict the system states [9].
- **Dynamic and Multi-faceted SpatioTemporal Graph Convolution Network (DMSTGCN, 2021)** is a variant of GCN that uses a dynamic graph convolutional network to capture the time-varying relationship and estimate the states of a graphed system [31].
- **Multivariate Time-series Graph Neural Networks (MTGNN, 2021)** is a deep spatial–temporal model for multivariate time series forecasting based on graph neural networks [32].
- **LSTM-CNN (2021)** employs a multi-level identification method combining innovation vectors, Long Short-Term Memory, and Convolutional Neural Network to detect false data and enhance state estimation accuracy in power systems. [6]
- **Spatial–Temporal Graph Convolutional Network (STGCN, 2022)** applies vanilla GCN and GRN layers to predict the states. It takes account of the fixed connectivity between buses in the power system [11].
- **Spatial and Temporal IDentity (STID, 2022)** is a simple yet effective method for multivariate time series forecasting based on spatial and temporal identity information. It applies an identity matrix to capture the spatial and temporal dependencies of the system states [33].
- **Spatial and Temporal Normalization (STNorm, 2022)** applies spatial and temporal normalization layers to improve the performance of the spatial–temporal network [34].

5.3. Compared approaches

To assess the performance of our proposed ST-AGNet method, we compare it with the above approaches on the following metrics: root mean square error (RMSE) and mean absolute error (MAE), considering there are negative values of the power systems' states. The RMSE and MAE are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathcal{N}}\frac{1}{\mathcal{H}}\sum_{i=1}^{\mathcal{N}}\sum_{t=1}^{\mathcal{H}}(\hat{x}_{i,t} - x_{i,t})^2}, \tag{16}$$

$$\mathrm{MAE} = \frac{1}{\mathcal{N}}\frac{1}{\mathcal{H}}\sum_{i=1}^{\mathcal{N}}\sum_{t=1}^{\mathcal{H}}\left|\hat{x}_{i,t} - x_{i,t}\right|, \tag{17}$$

where $\mathcal{N}$ stands for the count of buses, while $\mathcal{H}$ denotes the quantity of predicted time slots. $\hat{x}_{i,t}$ represents the forecasted state of the ith bus at the tth time slot, and $x_{i,t}$ corresponds to the actual state.

5.4. Prediction results

To assess the model comprehensively, we compare the proposed ST-AGNet model on the four different scales of power systems with the mentioned the-state-of-arts with RMSE and MAE. Besides, to evaluate the performance of the proposed model on different time scales, we predict the system states in the next 1, 2, and 3 hours; that is, the predicted time slot $\mathcal{H}$ is $\{4, 8, 12\}$.

Tables 1, 2, 3, and 4 present the prediction accuracy of various methods across the four datasets. As demonstrated in these tables, the proposed model outshines the other baselines on all four datasets. As one might expect, the task of forecasting grows incrementally more challenging as the prediction window extends, subsequently leading to a dip in the prediction accuracy across all models. Hence, in real-world applications, a careful trade-off consideration between the prediction accuracy and the prediction window is necessary.

In addition, most of the spatio-temporal models achieved better predictive structure compared to LSTM and LSTM-CNN, indicating that spatial information is crucial for the state prediction problem It is noteworthy that our selection predominantly consisted of deep learning methods, including other spatial–temporal prediction models and prior system state prediction models. Our model attained the most superior prediction outcomes, illustrating that our network structure is particularly well-suited for the state prediction task of power systems.

In addition, the accuracy of the proposed models in predicting the power system decreases as the size of the power system increases. This difference can be attributed to their more complex topology, which makes it more challenging to predict future system states. As a result, the effectiveness of different models decreases to some extent. However, our proposed ST-AGNet still achieves the best prediction accuracy, with the main model being able to capture the spatial and

Table 1
Comparison with different methods on IEEE 30-Bus system.

Methods	1h		2h		3h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
GW-Net	0.0017	0.0008	0.0025	0.0012	0.0033	0.0015
LSTM	0.0033	0.0020	0.0040	0.0024	0.0048	0.0027
DMSTGCN	0.0018	0.0009	0.0024	0.0012	0.0029	0.0014
MTGNN	0.0019	0.0009	0.0026	0.0012	0.0034	0.0016
LSTM-CNN	0.0025	0.0014	0.0030	0.0016	0.0036	0.0019
STGCN	0.0022	0.0013	0.0028	0.0014	0.0037	0.0019
STID	0.0143	0.0091	0.0173	0.0106	0.0188	0.0115
STNorm	0.0020	0.0010	0.0026	0.0013	0.0032	0.0016
ST-AGNet	**0.0014**	**0.0006**	**0.0022**	**0.0009**	**0.0030**	**0.0013**

Table 2
Comparison with different methods on IEEE 57-Bus system.

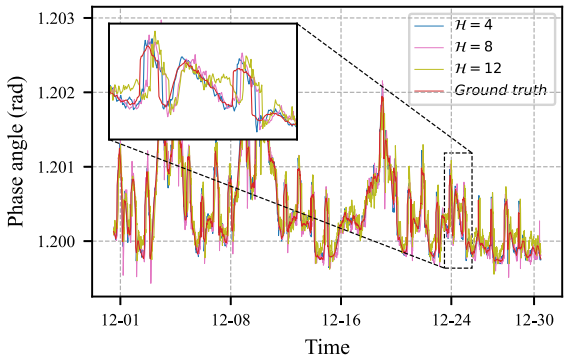
Methods	1h		2h		3h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
GW-Net	0.0073	0.0033	0.0075	0.0035	0.0082	0.0039
LSTM	0.0219	0.0128	0.0185	0.0109	0.0238	0.0142
DMSTGCN	0.0081	0.0041	0.0089	0.0044	0.0090	0.0046
MTGNN	0.0075	0.0036	0.0086	0.0045	0.0095	0.0048
LSTM-CNN	0.0109	0.0071	0.0114	0.0075	0.0118	0.0078
STGCN	0.0133	0.0072	0.0147	0.0080	0.0128	0.0079
STID	0.0292	0.0199	0.0327	0.0222	0.0351	0.0239
STNorm	0.0076	0.0041	0.0075	0.0042	0.0122	0.0048
ST-AGNet	**0.0039**	**0.0011**	**0.0046**	**0.0014**	**0.0050**	**0.0017**

Table 3
Comparison with different methods on IEEE 118-Bus system.

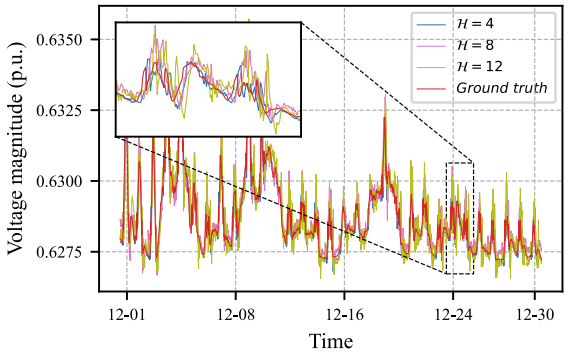
Methods	1h		2h		3h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
GW-Net	0.0109	0.0058	0.0148	0.0069	0.0209	0.0092
LSTM	0.1113	0.0552	0.1303	0.0645	0.1186	0.0595
DMSTGCN	0.0150	0.0072	0.0172	0.0085	0.0200	0.0094
MTGNN	0.0251	0.0123	0.0335	0.0163	0.0342	0.0159
LSTM-CNN	0.0517	0.0270	0.0528	0.0273	0.0573	0.0294
STGCN	0.0387	0.0194	0.0376	0.0188	0.0426	0.0209
STID	0.0535	0.0354	0.0586	0.0392	0.0651	0.0427
STNorm	0.0261	0.0140	0.0291	0.0153	0.0354	0.0184
ST-AGNet	**0.0100**	**0.0041**	**0.0146**	**0.0059**	**0.0204**	**0.0085**

Table 4
Comparison with different methods on IEEE 145-Bus system.

Methods	1h		2h		3h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
GW-Net	0.0298	0.0130	0.0271	0.0131	0.0357	0.0149
LSTM	0.1084	0.0486	0.1258	0.0569	0.1248	0.0580
DMSTGCN	0.0445	0.0167	0.0457	0.0197	0.0468	0.0184
MTGNN	0.0333	0.0152	0.0336	0.0160	0.0388	0.0170
LSTM-CNN	0.0415	0.0254	0.0521	0.0285	0.0504	0.0285
STGCN	0.0514	0.0193	0.0449	0.0232	0.0523	0.0234
STID	0.1300	0.0742	0.1399	0.0804	0.1484	0.0858
STNorm	0.2226	0.0609	0.2333	0.0610	0.2154	0.0656
ST-AGNet	**0.0155**	**0.0094**	**0.0236**	**0.0147**	**0.0286**	**0.0144**

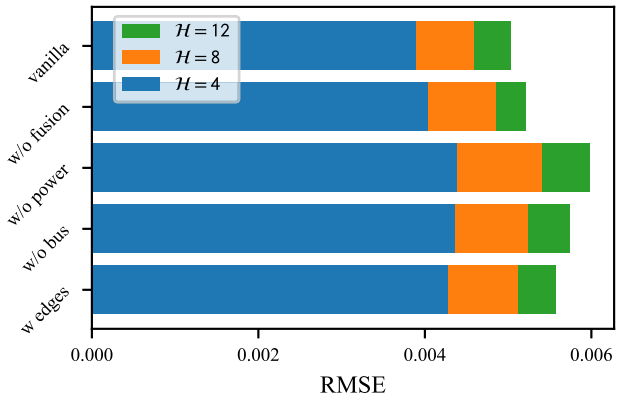


(a) The phase angle.

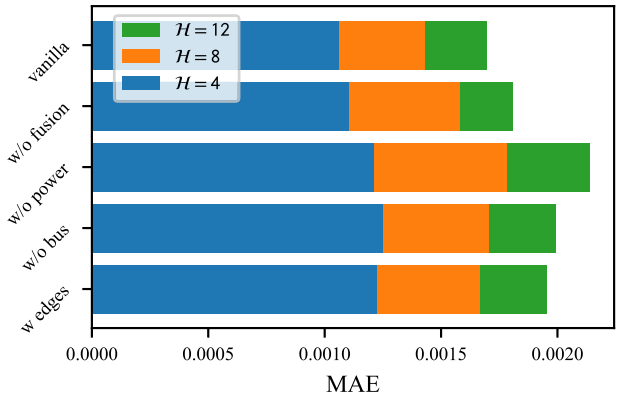


(b) The voltage magnitude.

Fig. 2. The predicted states curve of IEEE 57-Bus System on bus 37.



(a) The RMSE performance of different variants.



(b) The MAE performance of different variants.

Fig. 3. The performance of ablation test on the IEEE 57-Bus System.

temporal dependencies of the system state from different dimensions, which are fused and predicted.

5.5. Prediction result visualization

In this section, we delve deeper into the monthly dynamic system state variables of each common bus, with a particular emphasis on bus 37 where the solar power generation is integrated. We visualize the data for the entire month of December and three days, which encompasses both the real data and the prediction outcomes with prediction windows of $4, 8, 12$, respectively. Concurrently, we document both the phase angle and the voltage magnitude of this bus. The normalized results are depicted in Fig. 2.

As can be discerned from the fluctuating curves, when the prediction time window is set at 4, the model more accurately represents the trend of real values for the phase angle at bus 37. This superior fit can largely be attributed to the longer forecast time window, which inevitably results in an accumulation of errors and aligns with the accuracy result illustrated in Table 2. Moreover, in Fig. 2, a majority of the predicted values closely align with the real values for both the voltage magnitude and phase angle at bus 37. Similar findings can also be observed at bus 13, where wind power generation is integrated.

Upon further scrutiny of Fig. 2, it is evident that the prediction outcomes effectively mirror the daily changes in the states of power system, despite the occasional mismatches between sudden fluctuations in predicted and actual values at a detailed level, which are particularly noticeable at $\mathcal{H} = 12$. This effectively illustrates the proficiency of the ST-AGNet network in capturing spatial–temporal features within the power system.

5.6. Ablation study

In the subsequent subsection, we conduct comprehensive ablation studies to assess the influence of each module on the overall performance of the model. To ensure objectivity, the configurations used for training and testing in these case studies mirror those outlined in 5.1. For ease of understanding, we refer to the original ST-AGNet model as 'vanilla', and scrutinize the following four ST-AGNet variants:

- 'w/o fusion': The ST-AGNet model is without fusion module.
- 'w/o power': The ST-AGNet model without power module.
- 'w/o bus': The ST-AGNet model without bus module.
- 'w edges': Replace the graph $\mathcal{G}_b$ in bus module by the connections of buses in real-world.

Fig. 3 shows the prediction accuracy of different variants of the ST-AGNet model on the IEEE 57-Bus System with all predicted time windows. The results of the ablation study clearly show that each module of the ST-AGNet model contributes to its prediction accuracy. The fusion module, power module, and bus module all play a critical role in the model's ability to forecast states in the power system accurately. The reduction in prediction accuracy when any of these modules are removed emphasizes their importance in the design of the ST-AGNet model and performance.

In addition, by comparing the ST-AGNet model with the 'w edges', it can be observed that the use of a real-world bus connection graph (w edges variant) instead of the graph $\mathcal{G}_b$ in the bus module leads to a decrease in prediction accuracy. This suggests that the transmission matrix $\mathcal{Y}$ utilized in the ST-AGNet model is more effective at capturing

Table 5
Time consumption (s) comparison with different methods.

Methods	30-Bus		145-Bus	
	Training	Inference	Training	Inference
GW-Net	71	5	503	48
LSTM	**5**	**2**	**11**	**3**
DMSTGCN	165	15	360	40
MTGNN	102	8	540	32
STGCN	36	3	195	17
STID	26	3	57	5
STNorm	102	8	576	56
ST-AGNet	114	9	886	89

the spatial relationships and dependencies between the buses than the real-world bus connections. This demonstrates the importance of an accurately constructed graph in graph-based models like ST-AGNet, as it plays a crucial role in capturing the topological structure and dependencies of the system.

5.7. Time consumption

Considering real-world applications, we compare the time overhead of the different methods, including training and inference time. The data in the Table 5 are the results on 30-Bus and 145-Bus, where the training time is under per epoch and the inference time is around 7000 test samples. As can be seen from the table, LSTM has the shortest training time and inference time, mainly due to its simple structure, but the same leads to its lower prediction accuracy. In contrast, the networks with complex structures, such as DMSTGCN, GW-Net, etc., have more extended training and inference times, but their prediction accuracies are also relatively high.

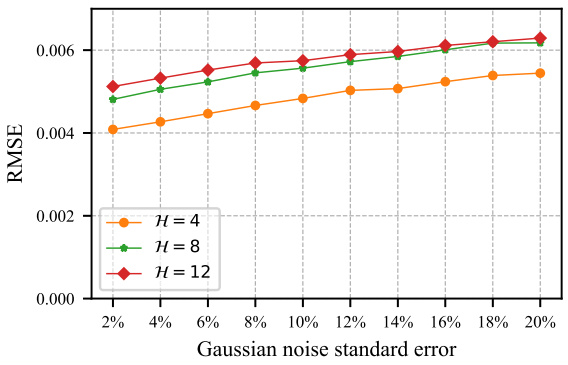
The training time and inference time of the proposed ST-AGNet are also longer, especially on 145-Bus, mainly due to its complexity as it contains multiple graph structures, thus the time increase is more significant. However, considering that our data sampling interval is 15 min, this time overhead is acceptable in practical applications. In addition, the prediction accuracy of ST-AGNet is also the highest. Therefore, the proposed ST-AGNet is feasible in practical applications.

However, in terms of time complexity, for a graph of n nodes, the time complexity of LSTM is $O(n)$, while the time complexity of GCN is $O(n^2)$, so the increase in the time consumed by the model based on the GCN network like the proposed ST-AGNet will be more pronounced as the number of nodes increases. This needs to be taken into account in large-scale power systems, even though the number of nodes in a power system is limited in practical applications.

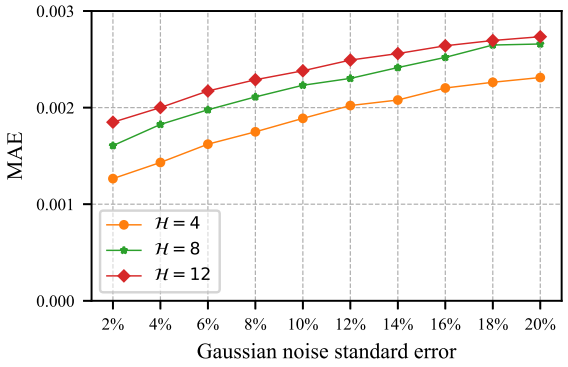
5.8. Noise analysis

In preceding case studies, we leverage the ground truth from four power system datasets to assess and scrutinize the proposed model. Nevertheless, the act of sampling unavoidably exposes the gathered data to the influence of unwanted noise. In this subsection, we intentionally introduce varying levels of noise to the input of the model to evaluate its robustness within the context of the IEEE 57-Bus System. Specifically, drawing on the methodology from previous work [35], we generate Gaussian noise with a mean of 1 and standard errors ranging among $\{2\%, 4\%, \ldots, 20\%\}$. This noise is then multiplied by the original input. This implies that the average deviations from the ground truth are 1%, 2%, ..., and 10%, respectively.

Fig. 4 shows the prediction performance of the proposed ST-AGNet models with three predicted time windows. The findings suggest that higher levels of noise often result in increased MAE and RMSE, while the affected data distribution patterns become more difficult to learn. But it can be seen that even at 10% of noise, the proposed ST-AGNet model still achieves a low MAE and RMSE under three scenarios, which demonstrates the robustness of the proposed model to noise interference.



(a) The RMSE result by noising influence.



(b) The MAE result by noising influence.

Fig. 4. The predicted performance with noise on IEEE 57-Bus System.

Table 6
The performance of ST-AGNet with different hyper-parameters.

Learning rate	Batch Size	RMSE	MAE
1e−2	25	0.0048	0.0019
	50	0.0046	0.0020
	100	0.0042	0.0017
1e−3	25	0.0042	0.0013
	50	0.0039	0.0011
	100	0.0044	0.0016
1e−4	25	**0.0037**	0.0010
	50	0.0039	**0.0009**
	100	0.0038	0.0010

5.9. Hyperparameter analysis

While hyperparameter selection is often deemed critical for model performance. We conduct a case study based on the IEEE 57-Bus System with a 1-hour prediction window to evaluate the impact of learning rate and batch size on model performance. The results are shown in the Table 6.

The table demonstrates a relatively stable performance across a variety of learning rate and batch size configurations. Despite altering learning rates by orders of magnitude, from 1e-2 down to 1e-4, and adjusting batch sizes from 25 to 100, the RMSE and MAE metrics exhibit only slight variations. The RMSE values remain within a tight range of 0.0037 to 0.0048, while MAE values fluctuate minimally between 0.0009 and 0.0020. These findings suggest that the early stop mechanism and the robustness of the proposed model are effective in maintaining performance despite the tested hyperparameter variations. Avoiding the obvious impact of hyperparameter selection on model results.

6. Conclusion

In this paper, we propose a generic ST-AGNet for predicting dynamic power system states. Initially, the AC power flow model is used for generating historical system state data. Then, both the spatial and temporal features are aggregated into the input dataset of the proposed neural network. Incorporating the graph-based information, the complete input dataset is trained and tested via the proposed ST-AGNet model in order to get accurate prediction results. Meanwhile, the time-varying complex system topology structures can be sufficiently captured, as well as the adjacency nodal information of the dynamic graphs. The comprehensive experiments are conducted based on four different scales of the systems. Compared with the baseline approaches, more accurate results are obtained by the proposed ST-AGNet model. In addition, we conduct an ablation test to verify the effectiveness of the constituting elements in the proposed ST-AGNet. Furthermore, empirical findings demonstrate that the proposed approach successfully fulfills the time and robustness criteria for practical implementations. Although the proposed ST-AGNet model is effective in predicting the dynamic power system states, there are still some limitations. For example, with the increase of nodes, the time consumed will increase dramatically. In the future, we will focus on improving the computational efficiency of the model. Besides, how to utilize the seasonal and yearly patterns to improve long-term prediction is also an important direction for future research.

CRediT authorship contribution statement

Shiyao Zhang: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Shuyu Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **James J.Q. Yu:** Writing – review & editing, Supervision, Investigation. **Xuetao Wei:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported in part by the Stable Support Plan Program of Shenzhen Natural Science Fund 20220815111111002, and in part by Key Talent Programs of Guangdong Province under Grant 2021QN02X166.

References

[1] Zhao J, Gómez-Expósito A, Netto M, Mili L, Abur A, Terzija V, et al. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. IEEE Trans Power Syst 2019;34(4):3188–98. http://dx.doi.org/10.1109/TPWRS.2019.2894769.

[2] Zhao J, Netto M, Huang Z, Yu SS, Gómez-Expósito A, Wang S, et al. Roles of dynamic state estimation in power system modeling, monitoring and operation. IEEE Trans Power Syst 2021;36(3):2462–72. http://dx.doi.org/10.1109/TPWRS.2020.3028047.

[3] Zhao J, Netto M, Mili L. A robust iterated extended Kalman filter for power system dynamic state estimation. IEEE Trans Power Syst 2017;32(4):3205–16.

[4] Cao X, Stephen B, Abdulhadi IF, Booth CD, Burt GM. Switching Markov Gaussian models for dynamic power system inertia estimation. IEEE Trans Power Syst 2016;31(5):3394–403.

[5] Lv J, Pawlak M, Annakkage UD. Prediction of the transient stability boundary based on nonparametric additive modeling. IEEE Trans Power Syst 2017;32(6):4362–9.

[6] Gao Z, Hu S, Sun H, Liu J, Zhi Y, Li J. Dynamic state estimation of new energy power systems considering multi-level false data identification based on LSTM-CNN. IEEE Access 2021;9:142411–24. http://dx.doi.org/10.1109/ACCESS.2021.3121420.

[7] Tian G, Gu Y, Shi D, Fu J, Yu Z, Zhou Q. Neural-network-based power system state estimation with extended observability. J Mod Power Syst Clean Energy 2021;9(5):1043–53. http://dx.doi.org/10.35833/MPCE.2020.000362.

[8] Xu L, Li L, Wang M, Wang X, Li Y, Li W, et al. Online prediction method for power system frequency response analysis based on swarm intelligence fusion model. IEEE Access 2023;11:13519–32. http://dx.doi.org/10.1109/ACCESS.2023.3242557.

[9] Zhang S, Yu JJ. Bayesian deep learning for dynamic power system state prediction considering renewable energy uncertainty. J Mod Power Syst Clean Energy 2022;10(4):913–22. http://dx.doi.org/10.35833/MPCE.2020.000939.

[10] Liao W, Bak-Jensen B, Pillai JR, Wang Y, Wang Y. A review of graph neural networks and their applications in power systems. J Mod Power Syst Clean Energy 2022;10(2):345–60. http://dx.doi.org/10.35833/MPCE.2021.000058.

[11] Wu T, Carreño IL, Scaglione A, Arnold D. Spatio-temporal graph convolutional neural networks for physics-aware grid learning algorithms. IEEE Trans Smart Grid 2023;14(5):4086–99. http://dx.doi.org/10.1109/TSG.2023.3239740.

[12] Huang J, Guan L, Su Y, Yao H, Guo M, Zhong Z. Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system. IEEE Access 2020;8:93283–96. http://dx.doi.org/10.1109/ACCESS.2020.2991263.

[13] Ma H, Lei X, Li Z, Yu S, Liu B, Dong X. Deep-learning based power system events detection technology using spatio-temporal and frequency information. IEEE J Emerg Sel Top Circuits Syst 2023;13(2):545–56. http://dx.doi.org/10.1109/JETCAS.2023.3252667.

[14] Zhang M, Zhen Z, Liu N, Zhao H, Sun Y, Feng C, et al. Optimal graph structure based short-term solar PV power forecasting method considering surrounding spatio-temporal correlations. IEEE Trans Ind Appl 2023;59(1):345–57. http://dx.doi.org/10.1109/TIA.2022.3213008.

[15] Zheng W, Wu W, Gomez-Exposito A, Zhang B, Guo Y. Distributed robust bilinear state estimation for power systems with nonlinear measurements. IEEE Trans Power Syst 2017;32(1):499–509. http://dx.doi.org/10.1109/TPWRS.2016.2555793.

[16] Bilil H, Gharavi H. MMSE-based analytical estimator for uncertain power system with limited number of measurements. IEEE Trans Power Syst 2018;33(5):5236–47. http://dx.doi.org/10.1109/TPWRS.2018.2801121.

[17] Zhang T, Zhang W, Zhao Q, Du Y, Chen J, Zhao J. Distributed real-time state estimation for combined heat and power systems. J Mod Power Syst Clean Energy 2021;9(2):316–27. http://dx.doi.org/10.35833/MPCE.2020.000052.

[18] Liao W, Wang S, Bak-Jensen B, Pillai JR, Yang Z, Liu K. Ultra-short-term interval prediction of wind power based on graph neural network and improved bootstrap technique. J Mod Power Syst Clean Energy 2023;11(4):1100–14. http://dx.doi.org/10.35833/MPCE.2022.000632.

[19] Zhao T, Yue M, Wang J. Structure-informed graph learning of networked dependencies for online prediction of power system transient dynamics. IEEE Trans Power Syst 2022;37(6):4885–95. http://dx.doi.org/10.1109/TPWRS.2022.3153328.

[20] Le Nguyen H. Newton-Raphson method in complex form [power system load flow analysis]. IEEE Trans Power Syst 1997;12(3):1355–9.

[21] Chatterjee S, Mandal S. A novel comparison of Gauss-Seidel and Newton-Raphson methods for load flow analysis. In: 2017 international conference on power and embedded drive control. 2017, p. 1–7. http://dx.doi.org/10.1109/ICPEDC.2017.8081050.

[22] Zimmerman R, Chiang H-D. Fast decoupled power flow for unbalanced radial distribution systems. IEEE Trans Power Syst 1995;10(4):2045–52. http://dx.doi.org/10.1109/59.476074.

[23] Wood A, Wollenberg B. Power generation, operation, and control. New Jersey: Wiley-Interscience; 2014, 252–253, p. 252–3.

[24] Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018, arXiv preprint arXiv:1803.01271.

[25] Zhang S, James J. Bayesian deep learning for dynamic power system state prediction considering renewable energy uncertainty. J Mod Power Syst Clean Energy 2021;10(4):913–22.

[26] Oord Avd, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, et al. Wavenet: A generative model for raw audio. 2016, arXiv preprint arXiv:1609.03499.

[27] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 6000–10.

[28] Power Systems Test Case Archive, University of Washington. [Online]. Available: http://www.ee.washington.edu/research/pstca/.

[29] Data platform: time series, Open Power System Data. [Online]. Available: http://data.open-power-system-data.org/.

[30] Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph WaveNet for deep spatial-temporal graph modeling. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence. International Joint Conferences on Artificial Intelligence Organization; 2019, p. 1907–13. http://dx.doi.org/10.24963/ijcai.2019/264.

[31] Han L, Du B, Sun L, Fu Y, Lv Y, Xiong H. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. New York, NY, USA: Association for Computing Machinery; 2021, p. 547–55. http://dx.doi.org/10.1145/3447548.3467275.

[32] Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. New York, NY, USA: Association for Computing Machinery; 2020, p. 753–63. http://dx.doi.org/10.1145/3394486.3403118.

[33] Shao Z, Zhang Z, Wang F, Wei W, Xu Y. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In: Proceedings of the 31st ACM international conference on information & knowledge management. New York, NY, USA: Association for Computing Machinery; 2022, p. 4454–8. http://dx.doi.org/10.1145/3511808.3557702.

[34] Deng J, Chen X, Jiang R, Song X, Tsang IW. ST-norm: Spatial and temporal normalization for multi-variate time series forecasting. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. New York, NY, USA: Association for Computing Machinery; 2021, p. 269–78. http://dx.doi.org/10.1145/3447548.3467330.

[35] Zhang S, Zhang C, Zhang S, Yu JJQ. Attention-driven recurrent imputation for traffic speed. IEEE Open J Intell Transp Syst 2022;3:723–37. http://dx.doi.org/10.1109/OJITS.2022.3215621.