A Graph based Kolmogorov-Arnold Hybrid Network for Distribution System State Estimation

Tao Xu
State Key Laboratory of Smart Power Distribution Equipment and System, Tianjin University
Tianjin, China
taoxu2011@tju.edu.cn

Ning Yang
State Key Laboratory of Smart Power Distribution Equipment and System, Tianjin University
Tianjin, China

Rujing Wang
State Key Laboratory of Smart Power Distribution Equipment and System, Tianjin University
Tianjin, China

***Abstract*—Distribution System State Estimation (DSSE) faces significant challenges due to increasing renewable integration, measurement noise, inherent nonlinearity and the growing uncertainties of network parameters. While data-driven approaches, particularly graph-based methods, have shown promise, existing models struggle with long-range dependency capture and nonlinear feature representation. To address these challenges, this paper proposes a novel graph-based Kolmogorov-Arnold network (KAN) framework that incorporates an expanded adjacency matrix and a multi-head graph attention mechanism to capture the long-range spatial dependencies. The KAN module further enhances nonlinear regression capability through interpretable, data-driven function approximation. Simulation has been carried out on IEEE 33-bus and 118-bus systems to demonstrate the accuracy and robustness of the proposed model, highlighting its effectiveness and adaptability in realistic, noise-prone distribution network scenarios.**

***Keywords*—Distribution System State Estimation; Kolmogorov–Arnold Network; Graph Attention; Nonlinear Regression**

I. INTRODUCTION

Distribution System State Estimation (DSSE) aims to infer the state information of distribution networks (DNs) using limited measurement data [1]. It provides decision for network operator for supply-demand balance and the stable operation of the grid. Compared to transmission networks, DNs are characterized by poor observability and complex, dynamic topologies, which further increases the difficulty of DSSE. Among various methods, weighted least square (WLS) has been considered as the most established approach for static state estimation in DNs, offering high estimation accuracy and computational efficiency under ideal conditions [2]. However, driven by the global activities on Net Zero and technological advancements, renewable energy sources, i.e. photovoltaics (PVs) and wind turbines (WTs), along with electric vehicles (EVs) are being integrated into DNs on a large scale [3]. The inherent intermittency and volatility of these power sources significantly increase uncertainties of DSSE. Moreover, the presence of measurement noise, communication packet loss, and environmental disturbances further degrade data quality and increase the likelihood of estimation errors. Traditional model-driven state estimation algorithms are sensitive to bad data and faults, failing to accurately capture the true operational state of the DNs.

With the recent development of machine learning studies, data-driven state estimation methods have emerged as effective tools to address the increasing complexity and uncertainties of modern DNs [4]. By leveraging extensive historical measurements, machine learning models can be trained in a supervised or semi-supervised manner to learn the mapping between system inputs and operating states, thereby reducing the reliance on precise physical modeling.

In [5], a real-time state estimation approach based on deep ensemble learning, where the integration of residual neural networks and multivariate linear regression enabled accurate state estimation under limited measurement conditions was proposed. Despite their effectiveness, these models fail to consider the underlying topological structure of the DNs. [6] introduced state estimation methods based on graph neural networks (GNNs), leveraging their strengths in modeling dependencies between nodes in graph-structured data. However, most GNN models rely on fixed local aggregation; simply stacking layers to capture long-range dependencies often leads to the over-smoothing problem [7]. Also, the fixed activation functions limit their ability to model high-order nonlinearities characteristic of power flow equations, reducing effectiveness under complex scenarios.

To address these limitations, this paper proposes a DSSE model that integrates graph structural information with the Kolmogorov–Arnold network (KAN) [8]. The major contributions of this work can be summarized as follows:

1) A DSSE model that integrates graph structural information with the KAN, combining strong nonlinear fitting with graph-based representation is proposed.

2) An adjacency matrix expansion mechanism is designed to capture long-range dependencies consistent with power flow propagation.

3) A hierarchical structure of TransLayer and KAN is developed to enable adaptive global information aggregation and nonlinear regression fitting, achieving a deep fusion of structural perception and functional approximation.

II. METHODOLOGIES

A. Kolmogorov-Arnold Networks

KAN is inspired by the Kolmogorov–Arnold theorem[9], which states that any multivariate continuous function can be represented as a composition of several univariate functions, as shown below:

$$f(x_1, x_2, ..., x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \psi_{pq}(x_p) \right) \tag{1}$$

where both ϕ_q and ψ_{pq} represent univariate functions.

Based on this theory, KAN replaces the conventional combination of activation functions and linear weights in neural networks with an explicit additive structure. Each weighting parameter is replaced by a univariate function, thereby enhancing both interpretability and fitting capacity. A typical KAN layer approximates complex functions through basis function expansion and local spline fitting:

$$y = f(x) = \sum_{i=1}^{n} w_i \cdot \mathrm{Spline}_i(x_i) + b \tag{2}$$

where x_i denotes the i^{th} input variable, $\mathrm{Spline}(\cdot)$ is a spline function defined on x_i, w_i is the learned weighting, and β is the bias term.

979-8-3315-8138-1

Compared with the fitting function of a traditional multilayer perceptron (MLP), which is given by:

$$y = f(x) = \phi(Wx + b) \tag{3}$$

where σ is a predefined activation function (commonly ReLU, etc.).

Unlike MLPs constrained by fixed, piecewise linear activations, KAN employs learnable spline functions on edges. This is critical for DSSE, as it enables the model to efficiently capture complex, highly nonlinear voltage-power mappings. By replacing rigid approximations with data-driven splines, KAN achieves superior estimation accuracy and enhanced interpretability.

B. *DSSE Based on Graph Neural Network*

DSSE aims to infer unobservable internal state variables based on available measurements (e.g., nodal power injections), in order to ensure system operational security and accuracy. The objective of the WLS method is to minimize the weighted squared error between the measured and calculated values, thereby obtaining an optimal estimation of the system state. The objective function is formulated as:

$$\hat{x} = \arg\min_x (z - h(x))^T R^{-1} (z - h(x)) \tag{4}$$

where x denotes the state vector, z represents the measurable quantities, $h(x)$ is the nonlinear measurement function that models the mapping from x to z, and R is the covariance matrix of measurement errors. However, WLS relies on accurate modeling of the DNs, making it difficult to adapt to the dynamic and uncertain characteristics introduced by the high penetration of distributed energy resources (DERs). GNNs, as deep learning models designed for graph-structured data, have been widely applied to power system state estimation in recent years. By stacking multiple layers, GNNs can progressively aggregate multi-hop neighbor information to learn more expressive node representations. However, as the number of layers increases, GNNs often suffer from issues such as over-smoothing and vanishing gradients, which further hinder their ability to capture global structural dependencies, especially in tasks that require capturing long-range dependencies or global graph structures.

III. Proposed Model

A. *Model Architecture*

The overall architecture of the model is illustrated in Figure 1. It is based on the graph structure of the DNs where the nodal active and reactive power injections are inputs and voltage magnitude and phase angle estimation are outputs. First, the adjacency matrix is expanded to enlarge each node's receptive field, enabling the capture of long-range dependencies among nodes. Then, the model employs a layer-wise alternating structure where TransLayer blocks and KAN modules are stacked in sequence. Each TransLayer implements a multi-head graph attention mechanism that dynamically aggregates global information based on the graph structure. Through residual connections and feature flattening, the graph attention outputs are transformed into high-dimensional representations suitable for the regression module. These refined features are then fed into KAN modules, which perform nonlinear regression through their dual-path architecture to model complex voltage-power relationships.

B. *Initial Feature Embedding*

Before feeding the input into the graph attention layers, an initial linear projection is applied to map the raw input features (active and reactive power injections) to a high-dimensional latent space. This embedding enhances the model's expressive power and enables downstream layers to operate in a feature-rich representation space:

$$\mathbf{X}_p = \mathrm{Linear}\left(P_{\mathrm{inj}}, Q_{\mathrm{inj}}\right) \tag{5}$$

where $X_p \in \mathbb{R}^{B \times N}$ is the embedded feature tensor for batch size B, number of nodes N.

C. *Construction of the Expanded Adjacency Matrix*

The DNs inherently exhibit a graph structure, where its physical configuration can be modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The node set $\mathcal{V}$ represents the load nodes in the network topology, and the edge set $\mathcal{E}$ denotes the connections between distribution lines. During modeling, the system topology is typically represented by an adjacency matrix A, where $A_{ij} = 1$ if node i is connected to node j, and 0 otherwise.

In real DNs, although some nodes are not directly connected in the physical topology, long-range dependencies often exist due to the electrical coupling and propagation characteristics of power flow. Particularly in the presence of DERs or load fluctuations, such influences may propagate through multiple paths within the network, affecting distant nodes. To address this issue, an adjacency matrix expansion mechanism is introduced. By incorporating the K-hop neighbors, the original adjacency matrix is expanded as:

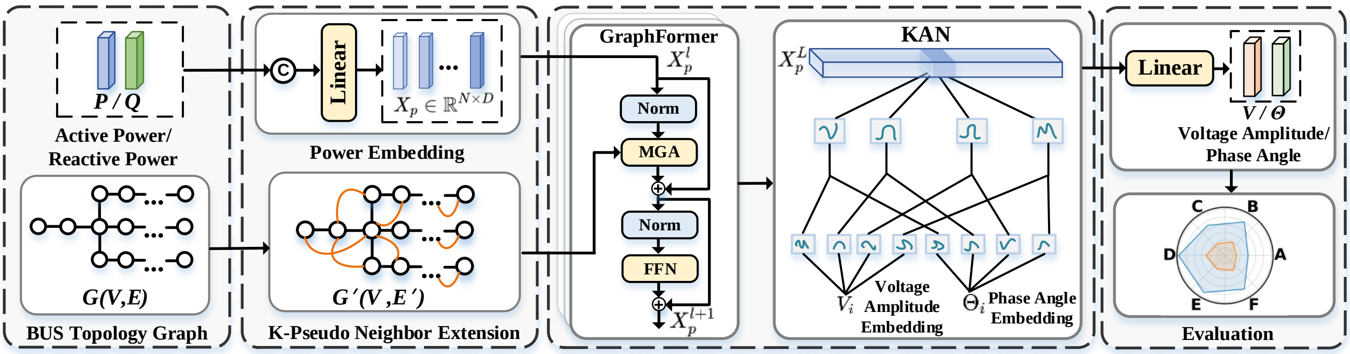

Fig. 1. Schematic diagram of the overall architecture of the proposed model

979-8-3315-8138-1

$$\mathbf{A}_{ij}^{(K)} = \begin{cases} 1, & \text{if } d(i,j) \le K \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

By connecting nodes within K hops, this method acts as a topological proxy for electrical distance to capture power flow influences among electrically close nodes. To implement this, a topology-based expansion mechanism is adopted. The expanded adjacency matrix $\mathbf{A}^{(K)}$ is calculated using matrix power series to determine K hop reachability:

$$\mathbf{A}^{(K)} = \min\left((\mathbf{A}+\mathbf{I})^{K}, 1\right) \tag{7}$$

where I denotes the identity matrix representing self-connections, and the term $(\mathbf{A}+\mathbf{I})^K$ aggregates all paths of length up to K. The $\min(\cdot\ ,1)$ operation serves as an element-wise clipping function to ensure the resulting matrix remains binary.

D. TransLayer with Graph Attention Mechanism

After expanding the adjacency matrix, a multi-head attention layer based on the graph structure is constructed. Traditional GNNs typically aggregate neighbor information through averaging or weighted summation, lacking the ability to distinguish the relative importance of different neighbors. TransLayer addresses this limitation by introducing a graph attention mechanism with learnable attention weightings, thereby enhancing the model's representation capability over graph-structured data.

In each TransLayer, the feature representation of a node is updated by attending to its neighboring nodes through a graph-based multi-head attention mechanism. For node i, the attention score with respect to node j is first computed using a scaled dot-product attention:

$$e_{ij} = \frac{(\mathbf{W}_Q \mathbf{x}_i)^{\top} (\mathbf{W}_K \mathbf{x}_j)}{\sqrt{d_k}} \tag{8}$$

where $\mathbf{W}_Q$ and $\mathbf{W}_K$ are learnable projection matrices for queries and keys, d_k is the dimension of the key vectors, and α_{ij} denotes the attention weight from node i to j.

To ensure the attention weights are comparable and sum to one, the softmax function is applied to normalize the scores across the expanded neighbor set $\mathcal{N}(i)$:

$$\alpha_{ij} = \text{softmax}_j\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathcal{N}(i)} \exp\left(e_{ik}\right)} \tag{9}$$

Next, node i aggregates the value-transformed features of its neighbors based on the learned attention weights:

$$\mathbf{x}_i = \sum_{j\in\mathcal{N}(i)} \alpha_{ij} \mathbf{W}_V \mathbf{x}_j \tag{10}$$

where $\mathbf{W}_V$ is the learnable value transformation matrix. This formulation enables the model to adaptively focus on informative neighbors while suppressing noisy or less relevant nodes.

To further enhance representational expressiveness, a multi-head attention mechanism, where the above attention is computed independently across H heads is proposed. The outputs from each head are averaged and fused with the original input via a residual connection:

$$\mathbf{Z}_i = \mathbf{x}_i + \frac{1}{H} \sum_{h=1}^{H} \tilde{x}_i^{(h)} \tag{11}$$

where $\tilde{x}_i^{(h)}$ denotes the aggregated features from the h^{th} attention head. This design allows the model to jointly capture diverse structural patterns from multiple representation subspaces.

Following the attention module, the output $\mathbf{Z}_i$ is passed through a two-layer feed-forward network (FFN) with a residual connection:

$$\mathbf{x}_i = \mathbf{Z}_i + \text{FFN}(\text{Norm}(\mathbf{Z}_i)) \tag{12}$$

where FFN $(\cdot)$ denotes the nonlinear transformation composed of linear mappings and activation functions, and Norm $(\cdot)$ represents RMS normalization. This step improves the nonlinear modeling capacity while preserving stable gradients during training.

E. KAN Regression Module

Following the graph attention module, KAN is introduced as a nonlinear regression layer. The high-dimensional embedding features of each node, obtained from the graph attention module, are flattened into vector representations and used as inputs to the KAN:

$$\mathbf{x}_{\text{flat}} = \text{Reshape}(\mathbf{X}) \in \mathbb{R}^{B\times(N\cdot D)} \tag{13}$$

For an input feature vector $\mathbf{x}$, the output of each KAN layer is given by:

$$\mathbf{y} = W_{\text{base}} \cdot \sigma(\mathbf{x}_{\text{flat}}) + W_{\text{spline}} \cdot \Phi(\mathbf{x}_{\text{flat}}) \tag{14}$$

where $\sigma(\mathbf{x})$ denotes the activation-based path, and $\Phi(\mathbf{x})$ represents the nonlinear transformation path based on spline function expansion. The corresponding weights for the two branches are W_{base} and W_{spline}, respectively.

The dual-path structure combines the stability of traditional activation functions with the nonlinear modeling capacity of spline-based transformations, enhancing the model's ability to fit complex relationships while preserving interpretability. This makes KAN well-suited for DSSE, where nonlinearities from system dynamics, DERs, and load variations pose challenges. By integrating KAN after the graph attention layers, the model jointly exploits structural information and nonlinear regression capability, achieving more accurate estimation.

IV. Case Studies

A. Settings

To evaluate the proposed DSSE model, simulations are performed on a modified IEEE 33-bus system. Load data are based on typical daily profiles from the New York Independent System Operator (NYISO) [10], capturing realistic urban load variations. PV units are integrated at buses 5, 10, and 15. The capacity of each unit is randomly sampled within a predefined range to simulate the variability and uncertainty of actual PV output. Power flow calculations are performed using Pandapower, and voltage, power, as well as other time-series operational data are collected as samples.

The model takes active power P and reactive power Q at each node as input, and outputs the voltage magnitude V and phase angle θ. The dataset is split into training and validation sets at a ratio of 8:2. All simulations are conducted on an NVIDIA RTX 4060 Ti GPU. The AdamW optimizer is used with a learning rate of 2×10^{-4}, and the model is trained for 100 epochs.

To comprehensively evaluate the estimation performance, three commonly used metrics are adopted: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). MAE measures the average deviation between predicted and true values, RMSE emphasizes the impact of larger errors, and MAPE reflects relative error, although for phase angles, MAE and RMSE serve as the primary metrics due to the numerical instability of MAPE near zero.

B. Results and Analysis

To evaluate the contribution of each component in the proposed model, ablation experiments with three variants are carried out: (1)TransLayer, which applies standard graph attention without neighbor expansion or KAN; (2)TransLayer-Expand, which incorporates K neighborhood expansion to capture broader structural dependencies; and (3) Proposed method, the full model combining graph attention with neighbor expansion and KAN nonlinear regression.

Simulation results in Table I show that both the neighborhood expansion and KAN module improve the model's estimation accuracy, validating the effectiveness of the proposed architecture.

Table II presents the performance comparison between the proposed model and other representative models on the state estimation task for the IEEE 33-bus system. The comparison benchmarks against three graph-based models (GCN [11], GGNN [12], and GAT [13]) as well as the traditional non-graph model (MLP). Specifically, GAT is included to validate the effectiveness of the proposed architecture against existing attention-based graph methods, providing a comprehensive evaluation across different model structures. As shown, the proposed model achieves superior estimation accuracy across all evaluation metrics. demonstrating its capability in capturing the complex nonlinear relationships among state variables. Compared to traditional GCN, GGNN, and even attention-based GAT models, the introduced neighbor expansion mechanism and the KAN module effectively alleviate issues related to limited long-range dependency capture and the constrained expressiveness of fixed activation functions, thereby improving the model's overall fitting capacity.

TABLE I.
ABLATION STUDY ON THE EFFECT OF NEIGHBORHOOD EXPANSION AND KAN MODULE ON IEEE 33-BUS SYSTEM

Case	V_m/θ	Metrics	Translayer	Translayer - Expand	Proposed Method
IEEE 33-Bus system	V_m	MAE	2.04E-3	1.07E-3	5.11E-5
		MAPE(%)	2.32E-1	1.21E-1	4.93E-3
		RMSE	2.69E-3	1.51E-3	2.16E-4
	θ	MAE	2.26E-1	2.43E-2	2.35E-3
		MAPE(%)	5.47E+0	6.25E-1	5.63E-2
		RMSE	2.91E-1	3.71E-2	3.57E-3

TABLE II.
COMPARISON OF STATE ESTIMATION PERFORMANCE OF DIFFERENT MODELS ON IEEE 33-BUS SYSTEM

Case	V_m/θ	Metrics	MLP	GAT	GCN	GGNN	Proposed Method
IEEE 33-Bus system	V_m	MAE	7.16E-5	1.64E-3	5.28E-3	1.68E-4	5.11E-5
		MAPE(%)	6.92E-3	1.37E-1	5.12E-1	1.64E-2	4.93E-3
		RMSE	2.31E-4	1.70E-3	7.10E-3	3.47E-4	2.16E-4
	θ	MAE	4.34E-3	7.38E-3	3.14E-1	6.23E-2	2.35E-3
		MAPE(%)	2.05E-1	6.27E-1	7.84E+0	1.64E+0	5.63E-2
		RMSE	5.89E-3	9.71E-3	4.01E-1	1.11E-1	3.57E-3

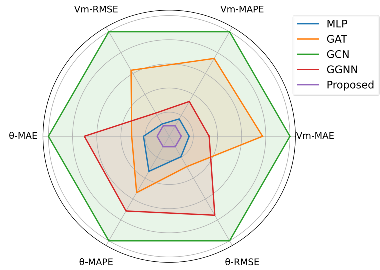

Fig. 2. Comparison of Model Performance on IEEE 33-Bus System

To provide a more intuitive comparison of the performance across different models, a radar chart is plotted using log-transformed values of each evaluation metric. The values are normalized using log-based Min-Max scaling, mapping them into the range [ε,1] (with ε=0.01).

To further evaluate the performance of the proposed model under real-world scenarios, Gaussian noise is added to the input data on the IEEE 33-bus system to simulate disturbances. Specifically, zero-mean Gaussian noise is applied to each node's active and reactive power with standard deviations set to 5% and 10% of their respective original standard deviations.

It can be observed from Tables III and IV that the estimation accuracy of all models decreases as the level of input noise increases. In particular, GCN exhibit the highest errors among all models under 10% Gaussian noise, in both voltage magnitude and phase angle estimation. This indicates that traditional graph-based models are prone to information degradation when subjected to measurement noise. In contrast, the proposed model consistently achieves the best performance across all three metrics, demonstrating its robustness and suitability for DNs with noisy measurements.

To evaluate the adaptability and generalization capability of the proposed model in DNs of different scales, additional simulation is conducted on the IEEE 118-bus test system. The model is retrained using the same training strategy and hyperparameter settings. As shown in Table V, although the estimation errors of all models increase compared to those on the IEEE 33 system, the proposed model still achieves the best performance across all evaluation metrics which indicate that the proposed model can maintain high estimation accuracy in larger and complex network topologies.

To evaluate the practicality of the proposed method, the average inference time of each model on the IEEE 33-bus system and 118-bus system are compared in Figure 3. Due to the added KAN module and neighbor expansion, the proposed method involves higher computational complexity. Compared with other methods, the proposed model runs slightly slower but remains within a reasonable range,

979-8-3315-8138-1

demonstrating a good trade-off between accuracy and efficiency.

TABLE III.
COMPARISON OF STATE ESTIMATION PERFORMANCE OF DIFFERENT MODELS ON IEEE 33-BUS SYSTEM UNDER 5% GAUSSIAN NOISE

Case	V_m/θ	Metrics	MLP	GAT	GCN	GGNN	Proposed Method
IEEE 33-Bus system	V_m	MAE	9.16E-4	1.89E-3	2.64E-2	2.01E-3	7.60E-4
		MAPE(%)	1.05E-1	2.51E-1	3.97E+0	2.32E-1	8.77E-2
		RMSE	1.33E-3	2.32E-3	2.28E-2	2.88E-3	1.21E-3
	θ	MAE	5.34E-2	6.19E-2	3.04E-1	1.28E-1	5.08E-2
		MAPE(%)	1.31E+0	1.46E+0	7.59E+0	3.23E+0	1.23E+0
		RMSE	7.38E-2	8.23E-2	3.96E-1	1.88E-1	7.06E-2

TABLE IV.
COMPARISON OF SATE ESTIMATION PERFORMANCE OF DIFFERENT MODELS ON IEEE 33-BUS SYSTEM UNDER 10% GAUSSIAN NOISE

Case	V_m/θ	Metrics	MLP	GAT	GCN	GGNN	Proposed Method
IEEE 33-Bus system	V_m	MAE	1.47E-3	1.53E-3	2.64E-2	2.33E-3	1.23E-3
		MAPE(%)	1.68E-1	1.82E-1	3.98E+0	2.68E-1	1.41E-1
		RMSE	2.08E-3	2.13E-3	4.47E-2	3.45E-3	1.88E-3
	θ	MAE	9.89E-2	1.01E-1	3.14E-1	1.61E-1	9.13E-2
		MAPE(%)	2.39E+0	2.43E+0	7.78E+0	4.03E+0	2.22E+0
		RMSE	1.32E-1	1.38E-1	4.05E-1	2.30E-1	1.23E-1

TABLE V.
COMPARISON OF STATE ESTIMATION PERFORMANCE OF DIFFERENT MODELS ON IEEE 118-BUS SYSTEM

Case	V_m/θ	Metrics	MLP	GAT	GCN	GGNN	Proposed Method
IEEE 118-Bus system	V_m	MAE	5.49E-4	7.52E-4	1.61E-2	5.50E-4	2.71E-4
		MAPE(%)	5.65E-2	7.80E-2	1.67E+0	5.69E-2	2.79E-2
		RMSE	7.16E-4	9.33E-4	2.17E-2	9.56E-4	3.62E-4
	θ	MAE	1.11E-2	1.56E-2	7.30E-1	4.91E-2	5.91E-3
		MAPE(%)	3.37E-1	5.25E-1	1.13E+1	1.13E+0	9.05E-2
		RMSE	1.59E-2	2.59E-2	1.26E+0	1.19E-1	8.04E-3

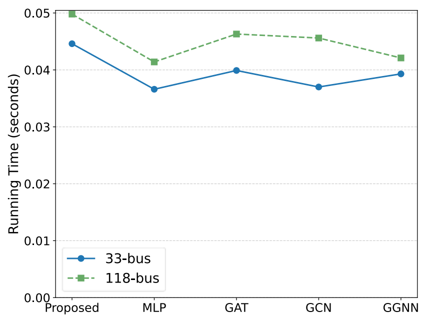

Fig. 3. Evaluation of Model Efficiency on IEEE 33-Bus and 118-Bus Systems

V. CONCLUSION

This study proposes a DSSE method that integrates graph-based model with KAN. To address the limitations of traditional methods in handling nonlinearities and exploiting graph structural information, particularly under high-penetration renewable energy scenarios, the model incorporates a neighbor expansion mechanism, graph attention layers, and an explicitly nonlinear KAN module. These components enable accurate estimation of complex voltage states in DNs.

Simulation results on the IEEE case study systems demonstrate that the proposed model significantly outperforms conventional approaches in predicting both voltage magnitudes and phase angles. Furthermore, robustness tests under Gaussian noise confirm the model's strong tolerance to measurement errors, highlighting its generalization ability and applicability in noise-affected conditions. Future work will focus on extending the framework to handle more complex scenarios, accounting for three-phase unbalance and single-phase laterals, to further validate its applicability in real DNs.

ACKNOWLEDGMENT

This work is supported by the Smart-Grid National Science and Technology Major Project (No. 2025ZD0804604) and the National Natural Science Foundation of China (No. 52277119).

REFERENCES

[1] J. Vijaychandra, B. R. V. Prasad, V. K. Darapureddi, B. V. Rao, and Ł. Knypiński, "A review of distribution system state estimation methods and their applications in power systems," *Electronics*, vol. 12, no. 3, p. 603, 2023.

[2] A. S. Debs and R. E. Larson, "A dynamic estimator for tracking the state of a power system," *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-89, no. 7, pp. 1670–1678, 1970.

[3] N. M. Manousakis, P. S. Karagiannopoulos, G. J. Tsekouras, and F. D. Kanellos, "Integration of renewable energy and electric vehicles in power systems: A review," *Processes*, vol. 11, no. 5, p. 1544, 2023.

[4] X.B. Jin, R. J. Robert Jeremiah, T.L. Su, Y.T. Bai, and J.L. Kong, "The new trend of state estimation: From model-driven to hybrid-driven methods," *Sensors*, vol. 21, no. 6, 2021.

[5] N. Bhusal, R. M. Shukla, M. Gautam, M. Benidris, and S. Sengupta, "Deep ensemble learning-based approach to real-time power system state estimation," *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106806, 2021.

[6] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Graph neural network-based distribution system state estimators," *IEEE Trans. on Industrial Informatics*, vol. 19, no. 12, pp. 11630–11639, Dec. 2023.

[7] W. Ju, Z. Fang, Y. Gu, *et al.*, "A comprehensive survey on deep graph representation learning," *Neural Networks*, vol. 173, p. 106207, 2024.

[8] Z. Liu, Y. Wang, S. Vaidya, *et al.*, "KAN: Kolmogorov-Arnold networks," in *Proc. International Conference on Learning Representations*, Singapore, Apr. 2025.

[9] J. Braun and M. Griebel, "On a constructive proof of Kolmogorov's superposition theorem," *Constructive Approximation*, vol. 30, pp. 653–675, 2009.

[10] The New York Independent System Operator, Inc[US]. https://www.nyiso.com/

[11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. International Conference on Learning Representations*, Toulon, France, Apr. 2017.

[12] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," in *Proc. International Conference on Learning Representations*, San Juan, Puerto Rico, May 2016.

[13] P. Veličković, G. Cucurull, A. Casanova, *et al.*, "Graph attention networks," in *Proc. International Conference on Learning Representations*, Vancouver, Canada, Apr. 2018.