Robust Decentralized Distribution System State Estimation via Message-Aware Graph Network

Shahabodin Afrasiabi
Power System Consultant (PSC)
Power Network North America Group
Burnaby, BC, Canada
Shahabodin.afrasiabi@pscconsulting.com

Sarah Allahmoradi
FortisAlberta
Advanced Power System Group
Airdrie, AB, Canada
Sarah.allahmoradi@fortisalberta.com

Xiaodong Liang
University of Saskatchewan
Department of Electrical and Computer Engineering
Saskatoon, SK, Canada
xil659@mail.usask.ca

***Abstract*—This paper introduces a decentralized state estimation framework for distribution systems using a message-aware graph attention model. Each node processes its own measurements, exchanges short messages with nearby nodes, and updates its state through multi-head attention. This structure captures feeder voltage behavior while preserving local effects from distributed energy resources (DERs). A Huber loss with an overlap-based consensus term improves the resistance to bad data and maintains consistent estimates between adjacent areas. The proposed method is validated using IEEE 123-bus and IEEE 906-bus test feeders, showing accurate and stable state estimation with low communication needs.**

***Index Terms*—Bad data, decentralized distribution state estimation, graph attention networks, robust loss function.**

I. INTRODUCTION

Accurate state estimation is essential for reliable operation, real-time monitoring, and advanced analytics for electrical distribution systems [1]. Rapid growth of distributed energy resources (DERs) has increased the complexity of distribution system state estimation (DSSE) significantly due to bidirectional power flow, the sparse measurements availability, and higher sensitivity to noise and bad data [2],[3]. DSSE methods can be categorized into centralized or decentralized. Centralized DSSE sends all measurements to a control center for estimation using analytical or data-driven models. Analytical DSSE typically uses Weighted Least Squares (WLS) as a baseline [3] and often applies Gauss–Newton iterations for sparse Jacobians [4]. Robust methods, such as LMS [5] and GM formulations [6], are used to reduce the impact of bad/corrupted data. Analytical DSSE depends on accurate topologies and line parameters, sensitive to modeling errors and uncertainty. Centralized DSSE has therefore shifted toward data-driven estimators, from early shallow models, such as Artificial Neural Network (ANN) [7] and kernel methods [8], to deep and graph-based approaches, such as multilayer networks [9], autoencoders [10], and GNNs [11]. However, these models often need large representative datasets, generalize poorly outside the training domain, may miss sharp DER-driven local voltage changes, and their centralized use creates heavy communication and computation burdens that limit scalability.

Decentralized DSSE frameworks have emerged as an alternative to reduce communication burdens, improve scalability, and allow feeders or regions to operate more autonomously. Decentralized analytical methods typically partition the network into local areas, where each region performs its own WLS-based estimation and iteratively exchanges boundary information with neighboring regions through consensus [12] or gradient-tracking schemes [13]. These methods reduce reliance on a single control center, but their accuracy can deteriorate when local regions have sparse measurements or when coordination among areas is insufficient. Decentralized data-driven DSSE has gained significant attention to further improve scalability and robustness. Recent works incorporate graph neural network (GNN) structures into distributed settings, enabling each local estimator to exploit electrical connectivity while avoiding full-system data sharing. The electrical-model-guided graph neural network (EleGNN) [14], for example, integrates electrical-model guidance into a GNN architecture to improve estimation accuracy and robustness across different network configurations. More recently, personalized federated learning (PFL)-based DSSE allow multiple regions to collaboratively train models without exchanging raw measurements, preserving the data privacy while enabling the system-wide learning. Although such decentralized learning methods reduce communication requirements and enhance scalability, they still struggle with localized DER-induced fluctuations, heterogeneous data quality across regions, and the difficulty to jointly model global smoothness and local voltage irregularities.

In this paper, a message-aware graph attention estimator (MAGAE) is proposed for decentralized DSSE. The method targets the major DSSE challenges, including the localized operating variability, sensitivity to noisy/uneven measurements (including varying pseudo-measurement penetration), and limited inter-area communication budgets. The key novelty is the message-aware inter-area learning that selectively prioritizes the exchanged boundary information and enforces the overlap consistency. A message-aware graph-attention network is designed for multi-area DSSE on the feeder graph with overlapping partitions. Compact, explicit boundary messages are designed for exchange between adjacent areas. The attention mechanism is designed to weight these boundary messages and suppress non-informative neighbor exchanges, reducing communication burden and improving the boundary-state consistency. An encoder–decoder mapping is developed for heterogeneous measurements, including real measurements and pseudo-measurements. A robust Huber-based learning objective is developed with adaptive pseudo-measurement weighting, and an overlap-aware soft-consensus penalty is designed to align shared boundary states under decentralized operation. The main contributions of this paper include:

- A decentralized MAGAE-DSSE architecture is proposed. A message-aware graph-attention network is designed with explicit boundary-message exchange between adjacent areas. The attention operator is conditioned on the boundary message to enable selective neighbor

979-8-3315-8138-1

aggregation under the limited inter-area communication.
- A robust training objective is proposed. A Huber-based loss is developed with adaptive pseudo-measurement weighting and an overlap-aware soft-consensus penalty, improving tolerance to bad data and measurement heterogeneity while maintaining consistency across overlapping regions.

The paper is organized as follows: Section II presents the DSSE formulation and graph-based system modeling; Section III describes the proposed MAGAE-DSSE architecture; Section IV provides numerical evaluations for IEEE 123-bus and IEEE 906-bus test systems under multiple operating scenarios using the proposed method; and Section VI concludes the paper and discusses the future work.

II. Preliminaries

A. Graph-based State Vector and Measurement

A distribution network is modeled by a graph $G_{net} = (\mathcal{V}, E_{net})$, where $U = \{(n,\phi): v \in \mathcal{V}, \phi \in \Phi_n\}$ show nodes. Each node $(n,\phi) \in U$ has a complex voltage $V_{n\phi} = v_{n\phi}^{re} + j\, v_{n\phi}^{im}$. The state vector collects real and imaginary parts of all phase voltages as follows:

$$x = \left[v_{1,a}^{re}, v_{1,a}^{im}, v_{1,b}^{re}, v_{1,b}^{im}, \dots, v_{n,\phi}^{re}, v_{n,\phi}^{im}, \dots\right]^{\mathrm{T}} \in \mathbb{R}^{2|U|} \quad (1)$$

Measurements z_m are indexed by $m \in \{1, \dots, M\}$. Each measurement satisfies a nonlinear model below:

$$z_m = h_m(x) + e_m \quad (2)$$

where $h_m(\cdot)$ characterizes the relationship between the state and the measurement, and e_m shows the noise associated with the measurement. The vector form in (1) reads $z = h(x) + e$. In a diagonal weight matrix $W = \mathrm{diag}(w_1, \dots, w_M)$, $w_m = 1/\sigma_m^2$. It is worthwhile noting that pseudo-measurements are assigned large σ_m^2(small w_m) to limit the bias.

B. The Objective Function of DSSE

DSSE estimates the system state from noisy measurements using the following weighted least-squares objective:

$$\min_{x \in \chi} F(x) = 1/2 \ \| W^{1/2}\big(z - h(x)\big) \|_2^2 \quad (3)$$

This objective penalizes squared residuals of measurements weighted by their inverse variances W.

To guard against bad data, a Huber loss is often used. With a threshold $\delta > 0$, the Huber penalty for residual r is

$$\rho_\delta(r) = \begin{cases} 1/2\, r^2, & |r| \le \delta, \\ \delta\, |r| - 1/2\, \delta^2, & |r| > \delta, \end{cases} \quad (4)$$

The common robust choice is $\delta \approx 1.345\, \sigma_m$ for a high efficiency under noisy conditions. The robust objective is

$$F_{Hub}(x) = \Sigma_{\mathrm{m}} \rho_\delta\big(w_m^{1/2}[z_m - h_m(x)]\big) \quad (5)$$

To ensure the estimated states are feasible, the following constraints are applied:

$$P_{inj.n\phi}^{\min} \le \mathrm{P}_{\mathrm{inj,n\phi}}(x) \le \mathrm{P}_{\mathrm{inj,n\phi}}^{\max}, \forall (n,\phi) \in U \quad (6)$$

$$Q_{inj.n\phi}^{\min} \le \mathrm{Q}_{\mathrm{inj,n\phi}}(x) \le \mathrm{Q}_{\mathrm{inj,n\phi}}^{\max}, \forall (n,\phi) \in U \quad (7)$$

$$\mathrm{v}_{\min} \le \big|V_{n\phi}(x)\big| \le \mathrm{v}_{\max}, \forall (n,\phi) \in U \quad (8)$$

$$\big|I_{ij,\phi}(x)\big| \le \mathrm{I}_{\mathrm{ij,\phi}}^{\max} \quad (9)$$

These constraints restrict the estimated states to physically feasible operating regions and serve as lightweight physics-informed regularization for the proposed framework.

C. Decentralized DSSE

Let the feeder be portioned into K areas $\mathcal{K} = \{1, \dots, K\}$ with index sets $\bigcup_i U_i = U$ (overlap allowed). Define selection matrices S_i that extract local states:

$$x_i = S_i x \ \in \ \mathbb{R}^{2|U_i|} \quad (10)$$

Let $\mathcal{N}_i$ be the set of neighboring areas that share boundary with the area i. With overlap selectors, C_{ij} and C_{ji} enforce consensus on the shared components as follows:

$$\min_{\{x_i\}_{i \in A}} f_i(x_i) \quad (11)$$

$$s.t.\ C_{ij} x_i = C_{ji} x_j, \forall i \in \mathcal{K}, \forall j \in \mathcal{N}_i \quad (12)$$

where $f_i(x_i)$ is the local WLS objective by restricting (3) to the measurement subset assigned to the area i, i.e., Eq. (3) with (z, W, h) is replaced by (z_i, W_i, h_i)). For the robust variant, replace each f_i by the Huber form induced by (4)–(5) on the local subset.

III. The Proposed Decentralized Message-Aware Graph Attention-based DSSE

A. Message-Aware Graph Attention-based DSSE

Each node exchanges local information over a communication graph, and selectively aggregates messages from its neighboring nodes through an attention mechanism. At each iteration, the hidden representation of the node $v = (n,\phi) \in U$ is denoted as $s_v \in \mathbb{R}^d$, derived from its local measurement features z_v and system metadata (e.g., phase type, measurement class, device type). The overall node feature matrix is:

$$S = \left[s_{v_1}, s_{v_2}, \cdots, s_{v_{|U|}}\right]^T \in \mathbb{R}^{|U| \times d} \quad (13)$$

The DSSE inference process performs graph convolutions with message-aware attention weights as

$$H = \sigma\big((E \odot A) S W\big) \quad (14)$$

where A is the adjacent matrix of the communication graph, E is the attention matrix representing the importance of neighboring messages, $\odot$ denotes element-wise multiplication, W is the trainable transformation matrix, and $\sigma(\cdot)$ is an activation function. The normalized attention coefficients between nodes i and j are defined by

$$E_{ij} = \frac{\exp\big(LeakyReLU\ (e_{ij})\big)}{\Sigma_{k \in \mathcal{N}_i} \exp(LeakyReLU\ (e_{ik}))} \quad (15)$$

where the unnormalized attention score is computed by the following key–query operation:

$$e_{ij} = \big(W_q s_i\big)^T \big(W_k s_j\big) \quad (16)$$

where W_q and W_k are trainable query and key matrices, respectively. W, W_q, and W_k are learned end-to-end jointly with the remaining model parameters by minimizing the overall training objective using the standard backpropagation. This formulation allows each node to adaptively weigh the information received from neighbors based on the message's relevance and context.

B. Multi-Head Attention and Aggregation

The multi-head attention mechanism is used in the proposed framework to improve stability and representation of richness. Considering P independent attention heads used to process parallel feature subspaces, the output features from all heads are concatenated as follows:

$$S' = f^{concat}\left[\sigma\left((E^p \odot A)SW^{(p)}\right)\right], \forall \mathrm{p} = 1, \cdots, \mathrm{P} \quad (17)$$

The resulting S' is then mapped to the estimated state vector $\hat{x}$ using a feed-forward network as follows:

$$\hat{x} = f_\theta(\acute{S}) \quad (18)$$

where $f_\theta(\cdot)$ represents the decoder with the parameters θ. The network is trained to minimize the robust Huber objective defined in (5), incorporating pseudo-measurements with adaptive weights.

C. Overal Structure of the Poposed Framework

The proposed framework is a decentralized estimator that runs at each node and proceeds in three stages per iteration: 1) local feature encoding, 2) message-aware graph attention aggregation, and 3) state decoding with overlap consensus. The design minimizes communication, scales to large feeders, and remains robust to heterogeneous measurements.

1) Local feature encoding: Each node $v = (n, \phi)$ collects its local measurement z_v, which are processed by a lightweight encoder $g(\cdot)$ to produce a compact latent representation:

$$s_v = g(z_v) \quad (19)$$

2) Message-aware graph attention aggregation: Encoded features are exchanged among neighboring nodes according to the communication graph defined by A. Each node aggregates messages from its neighbors through the attention mechanism introduced in (14)–(17), which adaptively weighs incoming information based on its relevance. This process yields the refined representation S' that embeds both local measurement content and network-topological relationships.

3) State decoding and overlap: The aggregated features are mapped to the estimated node states through the decoder $f_\theta(\cdot)$ as formulated in (18), producing $\hat{x}$ for all nodes. In feeders divided into multiple areas, shared boundary states are aligned via a soft consensus constraint, ensuring consistent estimates across overlaps. The consensus term is incorporated into the total loss as in (20), balancing local accuracy with global coordination.

$$L_{total} = L_{Huber} + \lambda_{cons} L_{cons} \quad (20)$$

where λ_{cons} controlling the trade-off between local accuracy and global agreement. In the decentralized framework, local estimators exchange boundary states with their neighboring areas. The overlapping states are enforced to reach consensus through the following soft penalty term:

$$L_{cons} = \sum_{(i,j)\in\varepsilon} \left\| C_{ij} x_i - C_{ji} x_j \right\|_2^2 \quad (21)$$

where C_{ij} and C_{ji} are selection matrices extracting the shared states. Fig. 1 shows the three-stage pipeline of the proposed MAGAE-based DSSE, showing the flow from measurement inputs to decentralized state reconstruction.

IV. Numerical Results

The proposed framework is evaluated using the IEEE 123-bus and IEEE 906-bus test systems. The IEEE 123-bus feeder includes two 500-kVA PV units and two 100-kVA wind turbines, and it is divided into five zones. The IEEE 906-bus low voltage (LV) test system, supplied by an 11 kV/0.4 kV transformer, is partitioned into twenty-one zones. Time-series load and renewable energy profiles are generated in OpenDSS using historical smart meter, solar irradiation, and wind speed data in [15], with 75% real measurements and 25% pseudo-measurements. The model is trained using Adam with a learning rate of 0.005, a batch size of 64, 4 attention heads, and $\lambda_{cons} = 1e^{-4}$. Training and test datasets are generated via OpenDSS time-series power flow simulations. Load and DER profiles are applied to generate operating points. Measurement vectors are constructed from simulated states and then corrupted with zero-mean Gaussian noise according to the assumed sensor variances. Samples are down-sampled to 30-minute resolution and split into 70% training and 30% testing sets. In all initial simulation scenarios, $T = 4$. The proposed MAGAE-based DSSE is compared with DSS² in [11], EleGNN in [14], and PFL-DSSE in [16], representing centralized, physics-guided, and federated learning-based DSSE methods. These methods are implemented according to the original publications and publicly available configurations. Fig. 2 shows stable loss convergence across all areas for both test systems.

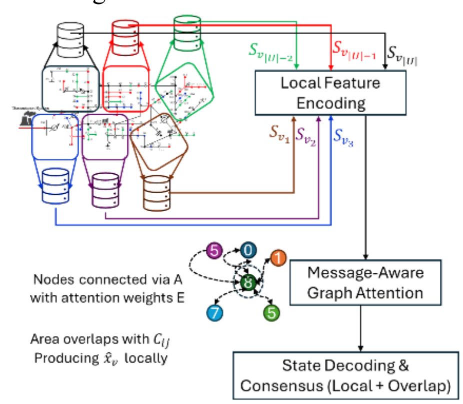

Fig. 1. The overall architecture of the proposed framework.

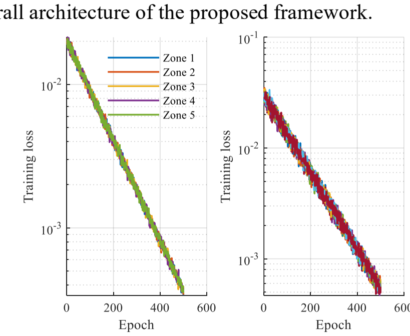

Fig. 2. Training loss convergence for: (a) IEEE 123-bus test system with 5 zones, (b) IEEE 906-bus test system with 21 zones.

A. Metrics

979-8-3315-8138-1

Two metrics are used to evaluate the proposed MAGAE-based DSSE framework: the mean absolute percentage error (MAPE), and the normalized root mean square error (NRMSE). For each measurement index $m \in \{1, \dots, M\}$ and each time step t, the MAPE and NRMSE are defined by

$$MAPE = \frac{1}{M} \sum_{m=1}^{M} \sum_{t=1}^{T} \left| \frac{\tilde{Y}_t^m - Y_t^m}{Y_t^m} \right| \tag{22}$$

$$NRMSE = \frac{1}{M.|\max(Y_t^m) - \min(Y_t^m)|} \sqrt{\left(\tilde{Y}_t^m - Y_t^m\right)^2} \tag{23}$$

B. The Baseline Scenario

The baseline scenario evaluates the performance of the proposed MAGAE-based DSSE framework under normal operation with full measurements and no bad data. Fig. 3 shows a 24-hour voltage estimation at bus 72, phase A. The estimated values track smart meter readings throughout the day. Table I reports the baseline accuracy for IEEE 123-bus and IEEE 906-bus test systems. MAPE improvements compared with DSS², EleGNN, and PFL-DSSE are 72.5%, 85.6%, and 86.9% for IEEE 123-bus test system, and 71.7%, 85.1%, and 86.6% for IEEE 906-bus test system.

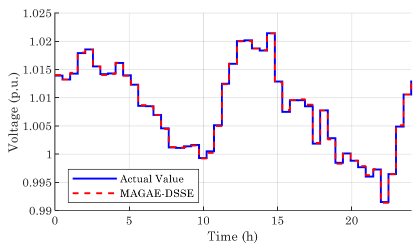

Fig. 3. A 24-hour voltage estimation at bus 72, phase A in IEEE 123-bus test system.

TABLE I. BASELINE MAPE AND NRMSE COMPARISON USING IEEE 123-BUS AND IEEE 906-BUS TEST SYSTEMS.

Methods	MAPE		NRMSE	
	IEEE 123-bus test system	IEEE 906-bus test system	IEEE 123-bus test system	IEEE 906-bus test system
MAGAE-DSSE (proposed)	0.92	1.04	0.17	0.2352
DSS² [11]	3.344	3.684	0.58	0.6827
EleGNN [14]	6.389	6.994	0.9009	0.7838
PFL-DSSE [16]	7.054	7.764	1.030	0.7917

C. Pseudo-Measurement Penetration

This subsection evaluates the effect of pseudo-measurement penetration using the IEEE 123-bus test system. Ten different penetration levels, ranging from 10% to 100%, are tested for the three selected benchmark methods in [11], [14], [16]. Fig. 4(a) shows the MAPE distribution heatmap. The proposed method keeps a narrow spread across all levels, while the three benchmark methods show wider variations and higher errors as the penetration level increases. Fig. 4(b) shows the NRMSE distribution heatmap with similar patterns. The heatmaps provide a clear view of how performance changes with penetration. The proposed method shows a slow and controlled rise in MAPE and NRMSE, while DSS², EleGNN, and PFL-DSSE show faster and larger increases.

D. Bad data Injection

Bad data are generated by randomly corrupting 5–20% of measurements with large additive noise spikes modeled as zero-mean Gaussian noise with an inflated variance, while the remaining measurements retain their nominal noise levels. Figs. 5(a)–5(b) show that the proposed method consistently achieves the lowest MAPE and NRMSE, while the three benchmark methods deteriorate as the data corruption increases. The proposed method shows superior performance, which down-weight outliers and maintain stable estimates for bad data.

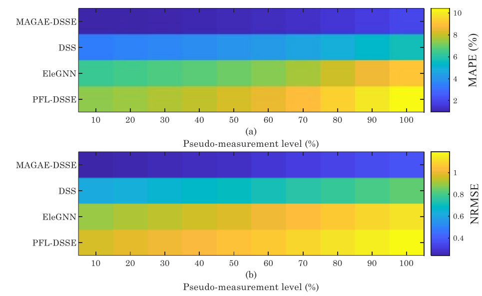

Fig. 4. The proposed method with various pseudo-measurement penetration levels for IEEE 123-bus test system: (a) MAPE distribution heatmap, (b) NRMSE distribution heatmap.

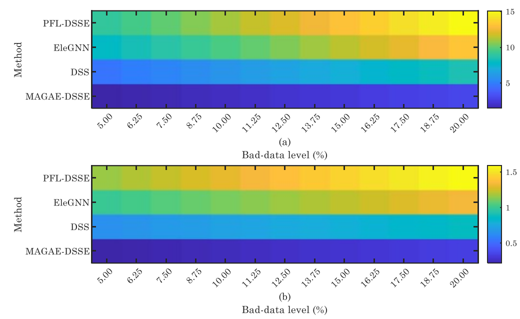

Fig. 5. The proposed method with increasing bad data for IEEE 123-bus test system: (a) MAPE distribution heatmap, (b) NRMSE distribution heatmap.

E. Communication Budget

This scenario studies the effect of limiting message-passing iterations in the proposed framework. The communication budget is set by the number of iterations T exchanged between neighboring areas. Lower T reduces the communication cost, but may also reduce the estimation accuracy. Table II shows the MAPE and NRMSE for $T = 1, 2,$ and 4 on IEEE 123-bus and IEEE 906-bus test systems., Reducing T from 4 to 1, MAPE increases from 0.92% to 1.35%, and NRMSE increases from 0.17 to 0.24, respectively, for IEEE 123-bus test system; while MAPE increases from 1.04% to 1.55% and NRMSE increases from 0.2352 to 0.31, respectively, for IEEE 906-bus test system. Although synchronous communication is assumed in this work, the proposed message-aware attention and overlap-consensus mechanisms provide inherent robustness to partial message loss and delayed updates.

TABLE II. COMMUNICATION BUDGET IMPACT ON THE PROPOSED MAGAE-BASED DSSE.

	IEEE 123-bus test system	IEEE 906-bus test system	IEEE 123-bus test system	IEEE 906-bus test system
T	MAPE	MAPE	NRMSE	NRMSE
1	1.35	1.55	0.24	0.31
2	1.05	1.0	0.19	0.26
4	0.92	1.04	0.14	0.2352

F. Computational Complexity

979-8-3315-8138-1

The training time, inference speed, and communication overhead are evaluated here. The proposed method converges faster than the three benchmarks due to message-aware attention and compact node features, and its decentralized design limits communication to boundary nodes. Table III shows that the proposed method uses fewer parameters and achieves a shorter training and inference time for IEEE 123-bus and IEEE 906-bus test feeders, while requiring fewer consensus iterations and lower computation and communication load.

TABLE III. TRAINING AND INFERENCE EFFICIENCY

Metric	MAGAE-DSSE (Proposed)	DSS² [11]	EleGNN [14]	PFL-DSSE [16]
Train / Epoch (s)	0.82	1.96	2.44	3.12
Train Time (min) – IEEE 123-bus	14.8	42.3	51.7	63.4
Train Time (min) – IEEE 906-bus	31.4	88.1	104.5	128.8
Inference (ms) – IEEE 123-bus	4.8	7.9	12.3	16.5
Inference (ms) – IEEE 906-bus	11.4	19.2	28.7	34.9
Trainable Params (×10³)	185	510	640	720
Communication Cost / Iter.	5×boundary nodes; 3.2 kB/msg	Centralized	Full-graph; 7.8 kB/msg	Client–server; 12.4 kB/msg
Convergence Iters.	8–10	–	18–22	25–30

G. Number of Zones

The impact of partition granularity is evaluated by varying the number of zones, K. As K increases, the local problem size decreases, but the boundary coupling and reliance on inter-area messages increase, which can reduce the estimation accuracy. Fig. 6 shows that when K increases from 5 to 10 on IEEE 123-bus test system, MAPE increases from 0.94% to 1.17%, and NRMSE increases from 0.171 to 0.204, respectively; Similarly, when K increases from 16 to 25, MAPE increases from 1.05% to 1.28%, and NRMSE increases from 0.241 to 0.278, respectively, for IEEE 906-bus test system.

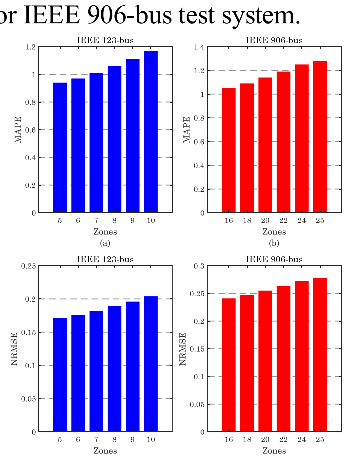

Fig. 6. Impact of partition granularity on the proposed method: (a) MAPE vs. number of zones on IEEE 123-bus test system, (b) MAPE vs. number of zones on IEEE 906-bus test system, (c) NRMSE vs. number of zones on IEEE 123-bus test system, (d) NRMSE vs. number of zones on IEEE 906-bus test system.

V. CONCLUSION AND FUTURE WORK

This paper proposes a decentralized DSSE method based on message-aware graph attention. The proposed MAGAE-based DSSE framework combines local measurements with neighbor information through cooperative attention and improves robustness using Huber loss with an overlap-consensus term. IT reduces MAPE and NRMSE substantially compared to the three benchmark methods, converges faster, and requires fewer message-passing rounds, cutting communication needs by more than half, while maintaining stable estimates with up to 20% bad-data injections. For future work, the proposed framework can be extended by incorporating explicit power-flow–based or physics-informed regularization beyond current feasibility constraints; relaxing the synchronous communication assumption to account for delays, packet loss, and asynchronous message passing; and further strengthening evaluation rigor through unified baseline tuning, uncertainty quantification, and enhanced reproducibility and presentation.

REFERENCES

[1] S. Afrasiabi, S. Allahmoradi, and X. Liang, "Pseudo-Measurement Models in Distribution Networks: A Review," *IET Smart Energy Systems,* vol. 1, no. 1, pp. 56-72, 2025.

[2] X. Lai, T. Xiao, L. Ding, Y. Chen, and S. Huang, "Robust State Estimation for Distribution Systems via Graph Signal Processing," IEEE Transactions on Power Systems, pp. 1-15, 2025.

[3] W. Tian, H. D. Chiang, and H. Mori, "Theoretical Foundation of Weight Adjustment for Weighted Least Squares State Estimator," IEEE Transactions on Power Systems, pp. 1-12, 2025.

[4] M. Cosovic and D. Vukobratovic, "Distributed Gauss–Newton Method for State Estimation Using Belief Propagation," IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 648-658, 2019.

[5] S. Park, R. Mohammadi-Ghazi, and J. Lavaei, "Nonlinear Least Absolute Value Estimator for Topology Error Detection and Robust State Estimation," IEEE Access, vol. 9, pp. 137198-137210, 2021.

[6] H. Zhao and B. Tian, "Robust Power System Forecasting-Aided State Estimation With Generalized Maximum Mixture Correntropy Unscented Kalman Filter," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-10, 2022.

[7] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurement Modeling," IEEE Transactions on Power Systems, vol. 27, no. 4, pp. 1888-1896, 2012.

[8] G. Hong and Y. S. Kim, "Supervised Learning Approach for State Estimation of Unmeasured Points of Distribution Network," IEEE Access, vol. 8, pp. 113918-113931, 2020.

[9] M. R. Dehbozorgi, M. Rastegar, and M. F. M. Arani, "False Data Injection Attack Detection and Localization Framework in Power Distribution Systems Using a Novel Ensemble of CNNs and Explainable Artificial Intelligence," IEEE Transactions on Industry Applications, vol. 61, no. 3, pp. 4801-4811, 2025.

[10] P. Sundaray and Y. Weng, "Alternative Auto-Encoder for State Estimation in Distribution Systems With Unobservability," IEEE Transactions on Smart Grid, vol. 14, no. 3, pp. 2262-2274, 2023.

[11] B. Habib, E. Isufi, W. v. Breda, A. Jongepier, and J. L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," IEEE Transactions on Power Systems, vol. 39, no. 2, pp. 4039-4050, 2024..

[12] A. Sagan, Y. Liu, and A. Bernstein, "Decentralized Low-Rank State Estimation for Power Distribution Systems," IEEE Transactions on Smart Grid, vol. 12, no. 4, pp. 3097-3106, 2021.

[13] X. Zhou, Z. Liu, Y. Guo, C. Zhao, J. Huang, and L. Chen, "Gradient-Based Multi-Area Distribution System State Estimation," IEEE Transactions on Smart Grid, vol. 11, no. 6, pp. 5325-5338, 2020.

[14] H. Lin and Y. Sun, "EleGNN: Electrical-Model-Guided Graph Neural Networks for Power Distribution System State Estimation," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 4-8 Dec. 2022 2022, pp. 5292-5298.

[15] G. L. Authority, "London datastore," URL: https://data. london. gov. uk, vol. 1, 2016.

[16] H. Wu, Z. Xu, J. Ruan, and X. Sun, "PFL-DSSE: A Personalized Federated Learning Approach for Distribution System State Estimation," CSEE Journal of Power and Energy Systems, vol. 10, no. 5, pp. 2265-2270, 2024.

979-8-3315-8138-1