Weakly-Supervised and Physics-Guided State Estimation for Distribution Networks

1st Jiachen Jiang
School of Electrical Engineering
Southeast University
Nanjing 210096, China
jiangjiachen.ee@seu.edu.cn

2nd Zhi Wu
School of Electrical Engineering
Southeast University
Nanjing 210096, China
zwu@seu.edu.cn

3rd Wei Gu
School of Electrical Engineering
Southeast University
Nanjing 210096, China
wgu@seu.edu.cn

4th Xiao Zhang
School of Electrical Engineering
Southeast University
Nanjing 210096, China
220232913@seu.edu.cn

5th Yiquan Li
Electric Dispatch and Control Center
Guangdong Power Grid Co., Ltd.
Guangzhou 510308, China
ee_lyq@sina.com.cn

***Abstract*—Accurate and reliable state estimation (SE) is critical for situational awareness in modern distribution networks. Yet the heterogeneity, asynchronicity, and incompleteness of real-world measurements challenge both conventional Weighted Least Squares (WLS) and recent data-driven models. This paper proposes Topology-Aligned Gated Recurrent Unit with Asynchronous Alignment (TAGRU), a reliability-aware and topology-aligned SE framework that integrates heterogeneous measurements through reliability-aware pooling and employs a topology-aligned spatio–temporal encoder to capture electrical dependencies and temporal dynamics. A weakly supervised learning strategy further combines sparse labels with a physics-consistency regularization to enforce physically plausible estimates.**

TAGRU is evaluated on a modified IEEE 33-bus system and IEEE 123-bus system. Under asynchronous and corrupted measurements, TAGRU maintains stable performance across all difficulty tiers, with angle RMSE increasing only from 0.993° to 1.051°.

***Index Terms*—Distribution Network State Estimation; Neural Network; Deep Learning; Physics-Guided Neural Networks; Graph Neural Networks; Weak Supervision**

I. Introduction

With the increasing integration of distributed generation, power flows in distribution networks become more volatile and uncertain. This volatility leads to more frequent fluctuations in operating conditions and imposes stricter requirements on secure operation, monitoring, and control [1], [2]. State Estimation (SE) is a key tool for situational awareness, and its accuracy and efficiency directly determine how well the system operating state can be observed and interpreted.

In response to these deficiencies, a plurality of data-driven approaches has been investigated. Methodologies based on Graph Neural Networks (GNNs) are capable of leveraging network topology[3], [4], [5]; however, their performance is demonstrably attenuated in scenarios characterized by the unavailability of precise topological information [6],[7],[8].

This work is supported by Smart Grid-National Science and Technology Major Project 2024ZD0802200.

To overcome these limitations, this paper proposes a reliability-aware and topology-aligned state estimation framework, denoted as TAGRU. The proposed method introduces a unified architecture that fuses heterogeneous measurements with varying reliability levels. It employs a reliability-aware pooling mechanism to adaptively emphasize trustworthy sensing channels. Furthermore, a topology-aligned spatio–temporal encoder integrates physical network structure with temporal dynamics for robust feature representation. To address the scarcity of ground-truth states, we formulate a weakly supervised learning strategy that combines sparse supervision with physics-consistency regularization, promoting physically plausible and data-efficient estimation. Extensive experiments on the modified IEEE 33-bus system and the IEEE 123-node three-phase unbalanced system demonstrate that TAGRU achieves superior accuracy and robustness compared with conventional and state-of-the-art learning-based baselines.

II. Proposed Method: TAGRU

A. Problem Formulation

We address the state estimation (SE) problem in a distribution network, represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $n = |\mathcal{V}|$ nodes (buses in single-phase systems, or bus-phases in the Y-node formulation for unbalanced three-phase systems). The system state at discrete time t is the vector of voltage magnitudes and angles:

$$x_t = \left[\ |V_1|, \dots, |V_n|,\ \theta_1, \dots, \theta_n\ \right]^\top \in \mathbb{R}^{2n}, \qquad (1)$$

where the slack bus angle is conventionally set to zero. A set of m heterogeneous measurements $z_t \in \mathbb{R}^m$, sourced from SCADA, AMI, and μPMUs, is available. These measurements are related to the system state through the nonlinear AC power-flow model:

$$z_t = h(x_t; \mathcal{G}, \theta_{\mathrm{pf}}) + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t] = 0,\ \mathrm{Cov}(\varepsilon_t) = R_t, \qquad (2)$$

979-8-3315-8138-1

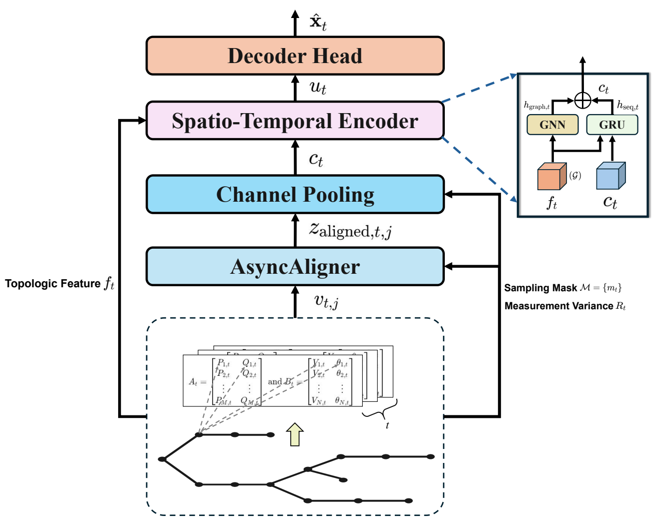

Fig. 1. Overall architecture of the proposed TAGRU framework.

where $h(\cdot)$ represents the physical power flow equations, θ_{pf} denotes network parameters, and ε_t is the measurement noise with covariance R_t.

The classical Weighted Least Squares (WLS) approach, $\min_x (z_t - h(x))^\top R_t^{-1}(z_t - h(x))$, is often compromised in modern distribution grids due to (i) low observability from sparse metering, (ii) high non-Gaussian noise, and (iii) uncertainty in network parameters. Furthermore, two critical challenges emerge from sensor heterogeneity:

1) **Asynchronicity:** Measurements arrive at different reporting rates (e.g., μPMUs at 30 Hz, SCADA every 2-10s). Standard sequential models require temporally aligned data, making pre-processing a significant challenge.
2) **Heterogeneous Reliability:** The reliability of z_t, captured by the variance R_t, varies drastically between sensor types and over time.

These challenges motivate a data-driven estimator that (1) explicitly models and compensates for sensor asynchronicity and reliability, (2) integrates grid topology, and (3) can be trained effectively under **weak supervision** (i.e., with sparse or absent ground-truth state labels x_t).

B. TAGRU: Architecture Overview

To address these challenges, we propose TAGRU (Topology-Aligned Gated Recurrent Unit with Asynchronous Alignment), a unified end-to-end architecture that maps a window of raw, asynchronous measurements $\mathcal{Z} = \{z_t\}_{t=t-W+1}^{t}$, their sampling masks $\mathcal{M} = \{m_t\}$, and covariances $\mathcal{R} = \{R_t\}$ to state estimates $\hat{\mathcal{X}} = \{\hat{x}_t\}$. As illustrated in Figure 1, the forward pass consists of four key stages:

1) **Asynchronous Alignment Module:** A learnable, local time-window processor that uses z_t, m_t, and R_t to compute a temporally aligned and smoothed measurement feature, $z_{\text{aligned},t}$.
2) **Reliability-Aware Channel Pooling:** An attention-based module that compresses the high-dimensional measurement vector into a low-dimensional, reliability-weighted embedding.
3) **Spatio-Temporal Encoder:** A dual-branch backbone (GNN + GRU) that processes c_t and external features f_t to capture graph-spatial and temporal dependencies, producing a fused hidden representation u_t.
4) **Decoder Head:** A final Multi-Layer Perceptron (MLP) that maps u_t to the system state estimate $\hat{x}_t$.

C. Detailed Model Architecture

a) Asynchronous Alignment Module (‘AsyncAligner‘): This module addresses sensor asynchronicity by learning a data-driven, reliability-aware interpolation kernel. For each channel j and time t, we define a 3-dimensional feature

$$v_{t,j} = \left[z_{t,j},\, m_{t,j},\, \log(1/(R_{t,j} + \epsilon))\right], \tag{3}$$

where $z_{t,j}$ is the raw value, $m_{t,j}$ is the binary sampling mask, and $\log(1/R_{t,j})$ encodes measurement reliability. A temporal feature extractor (MLP plus 1D depthwise convolution) maps $v_{t,j}$ to logits $w_{t,j} \in \mathbb{R}^k$, which are normalized by Softmax to obtain alignment weights $\alpha_{t,j}$. The aligned measurement is then

$$z_{\text{aligned},t,j} = \sum_{i \in \mathcal{N}_k(t)} \alpha_{t,j,i} \cdot z_{i,j}, \tag{4}$$

implementing a learned local interpolation that adaptively selects informative time neighbors.

b) Reliability-Aware Channel Pooling: This module compresses the m-dimensional aligned measurements $z_{\text{aligned},t} \in \mathbb{R}^m$ into a fixed-size embedding c_t. For each channel j at time t, we construct an aligned, reliability-aware feature vector

$$\tilde{v}_{t,j} = \left[z_{\text{aligned},t,j},\, m_{t,j},\, \log(1/(R_{t,j} + \epsilon))\right]. \tag{5}$$

An attention mechanism is applied to compute a channel importance score

$$s_{t,j} = \text{MLP}_{\text{score}}(\tilde{v}_{t,j}), \tag{6}$$

which is normalized across channels by

$$\beta_{t,j} = \text{Softmax}_j(s_{t,j}). \tag{7}$$

In parallel, each channel feature is projected as

$$v'_{t,j} = \text{MLP}_{\text{proj}}(\tilde{v}_{t,j}). \tag{8}$$

The final context vector is obtained by a reliability-weighted aggregation:

$$c_t = \sum_{j=1}^{m} \beta_{t,j} \cdot v'_{t,j}. \tag{9}$$

This mechanism prioritizes information from reliable and actively sampled sensors when forming the sequence embedding.

979-8-3315-8138-1

c) Spatio-Temporal Encoder and Decoder: The core estimator follows a standard spatio-temporal design:

- **Temporal Encoder:** A multi-layer GRU processes the sequence of context vectors c_t and any external features f_t, capturing temporal dynamics: $h_{\text{seq},t} = \text{GRU}([c_t, f_t], h_{\text{seq},t-1})$.
- **Spatial Encoder:** A stack of L_g Topology-Adaptive Graph Convolutional (TAGConv) layers operates on node-level features (derived from f_t or a static node embedding). This captures the spatial correlations defined by $\mathcal{G}$: $g_t = \text{TAGConvStack}(f_{\text{node},t}, \mathcal{G})$. The node features are pooled to create a graph-level sequence $h_{\text{graph},t}$.
- **Gated Fusion:** The temporal $h_{\text{seq},t}$ and spatial $h_{\text{graph},t}$ representations are fused using a learned gate α_t: $u_t = \alpha_t \odot h_{\text{seq},t} + (1-\alpha_t) \odot h_{\text{graph},t}$.
- **Decoder:** A final MLP decoder maps the fused representation u_t to the state vector: $\hat{x}_t = \text{MLP}_{\text{head}}(u_t)$. The angular components $[\hat{\theta}_n]$ are passed through an angle-wrapping function to ensure they lie in $(-\pi, \pi]$.

D. Weakly-Supervised Training Objective

A key challenge in distribution system state estimation is the scarcity of ground-truth state vectors x_t. To overcome this, we employ a weakly-supervised training objective that combines a sparse supervised loss (for accuracy, where available) with a dense physics-consistency loss (for physical plausibility, everywhere).

$$\mathcal{L}_{\text{total}} = \lambda_{\text{sup}} \cdot \mathcal{L}_{\text{sup}} + \lambda_{\text{phys}} \cdot \mathcal{L}_{\text{phys}} + \lambda_{\text{temp}} \cdot \mathcal{L}_{\text{temp}} \quad (10)$$

where λ terms are scalar weights.

1) **Sparse Supervised Loss ($\mathcal{L}_{\text{sup}}$):** This is a standard Mean Squared Error (MSE) loss, but it is applied *only* at time steps $t \in \mathcal{T}_L$ where high-fidelity ground-truth labels are available. Let $\mathcal{T}_L \subseteq \{1, \ldots, W\}$ be the set of labeled time steps (e.g., only the last step, or every k-th step).

$$\mathcal{L}_{\text{sup}} = \frac{1}{|\mathcal{T}_L|} \sum_{t \in \mathcal{T}_L} \left(w_\theta \cdot \|\hat{\theta}_t - \theta_t\|^2 + w_v \cdot \||\hat{V}|_t - |V|_t\|^2 \right) \quad (11)$$

This term anchors the model to known-good states, while w_θ and w_v balance the angle and magnitude errors.

2) **Physics-Consistency Loss ($\mathcal{L}_{\text{phys}}$):** This loss serves as the primary regularizer, applied at all time steps t. It enforces physical laws by ensuring the state estimate $\hat{x}_t$ is consistent with the measurements z_t. It minimizes the weighted squared residual between the actual measurements z_t and the measurements recalculated from the estimate, $h(\hat{x}_t)$:

$$\mathcal{L}_{\text{phys}} = \mathbb{E}_t \left[\left\| z_t - h(\hat{x}_t) \right\|^2_{R_t^{-1}} \right] \quad (12)$$

where $\|\cdot\|^2_{R_t^{-1}}$ is the squared Mahalanobis distance. This loss requires no state labels, only the physical model $h(\cdot)$ and measurement covariances R_t.

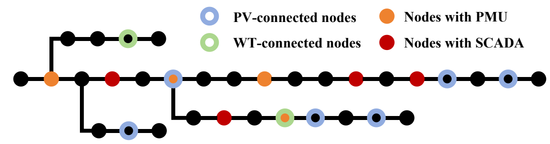

Fig. 2. Topology of the modified IEEE 33-bus test system.

3) **Temporal Smoothness Loss ($\mathcal{L}_{\text{temp}}$):** A simple L2 regularization on the first-order difference of the estimates to discourage unrealistic state jumps: $\mathcal{L}_{\text{temp}} = \mathbb{E}_t \left[\|\hat{x}_t - \hat{x}_{t-1}\|^2 \right]$.

This hybrid objective allows the model to learn from sparse, high-quality labels while leveraging the dense, unlabeled measurement data and physical priors for robust and plausible estimation.

III. Case Study

A. Experimental Setup

Numerical validation was performed on a modified IEEE 33-bus test system(FIG.2) integrated with photovoltaic (PV) and wind turbine (WT) generators. Time-series profiles were sourced from SimBench to create spatio-temporal windows. Each data sample includes nodal injection powers, branch power flows, and the ground-truth system voltage magnitudes ($|V|$) and phase angles (θ).

To emulate a realistic and challenging observability scenario, we established a hybrid measurement model during data generation, comprising three distinct types:

1) **Pseudo-measurements:** All nodal active (P) and reactive (Q) power injections were treated as pseudo-measurements, contaminated with Gaussian noise corresponding to a standard deviation of **20%** relative to their nominal values.
2) **SCADA measurements:** Bus voltage magnitudes ($|V|$) at selected locations and branch active (P) and reactive (Q) power flows were simulated as conventional SCADA measurements, contaminated with Gaussian noise corresponding to a standard deviation of **1%**.
3) **μPMU measurements:** Voltage phasors (both magnitude $|V|$ and angle θ) at sparsely chosen locations were simulated as μPMU measurements, featuring higher precision with Gaussian noise corresponding to a standard deviation of **0.1%**.

The measurement covariances R_t used in $\mathcal{L}_{\text{phys}}$ and the 'Reliability-Aware Channel Pooling' are populated with the ground-truth variances from this noise generation process.

The numerical experiments were conducted on an NVIDIA RTX 4090 GPU platform.

B. Accuracy Analysis of State Estimation

We benchmark the proposed TAGRU framework against four representative baselines: weighted least squares (WLS), a topology-aware graph neural network (GNN), a recurrent GRU model, and a Transformer-based sequence model. Table I summarizes the state-estimation accuracy in terms of RMSE

TABLE I
STATE-ESTIMATION ACCURACY COMPARISON ON THE MODIFIED IEEE 33-BUS SYSTEM.

Algorithm	θ-RMSE (°)	$\|V\|$-RMSE (p.u.)	θ-MAE (°)	$\|V\|$-MAE (p.u.)
WLS	15.30	0.054	7.54	0.014
GNN	2.603	0.101	2.456	0.071
GRU	1.12	0.081	0.721	0.057
Trans	1.414	0.027	1.057	0.023
Ours(TAGRU)	**0.990**	**0.028**	**0.694**	**0.019**

and MAE for both voltage phase angle (θ) and magnitude ($|V|$).

WLS yields the lowest accuracy (θ-RMSE 15.30°, $|V|$-RMSE 0.054 p.u.; θ-MAE 7.54°, $|V|$-MAE 0.014 p.u.), which is consistent with strong reliance on pseudo-measurements and sensitivity to measurement noise, missing data, and low-observability regimes that are common in practical distribution systems. Incorporating network topology with a GNN improves the phase-angle accuracy substantially (θ-RMSE 2.603°) but does not consistently improve voltage-magnitude estimation ($|V|$-RMSE 0.101 p.u.). The relatively large MAE values (θ-MAE 2.456°, $|V|$-MAE 0.071 p.u.) further indicate instability under heterogeneous sensor reliability.

Sequence-learning baselines provide additional improvements. The GRU model reduces the phase-angle error to θ-RMSE 1.12° but exhibits degraded voltage-magnitude accuracy ($|V|$-RMSE 0.081 p.u.). The Transformer achieves the best voltage-magnitude accuracy among the baselines ($|V|$-RMSE 0.027 p.u.; $|V|$-MAE 0.023 p.u.) but produces larger phase-angle errors (θ-RMSE 1.414°; θ-MAE 1.057°), which can be attributed to a weaker inductive bias for graph-structured physical systems. Overall, these results suggest that temporal modeling mitigates part of the noise and misalignment, but it remains insufficient when spatial–temporal interactions and measurement uncertainty are tightly coupled.

TAGRU achieves the best performance across all metrics, with θ-RMSE 0.990° and $|V|$-RMSE 0.028 p.u., and corresponding MAE values of 0.694° and 0.019 p.u.. Relative to the strongest baseline, which is the Transformer, TAGRU reduces θ-RMSE by 30.0% while maintaining comparable voltage-magnitude accuracy. Relative to the GNN, TAGRU reduces θ-RMSE by 62.0% and $|V|$-MAE by 73.2%. Relative to WLS, TAGRU reduces θ-RMSE by 93.5% and $|V|$-RMSE by 48.1%.

Figure 3 depicts the RMSE trajectories of voltage magnitude and phase angle over the final 95 timesteps of the test window. However, the WLS phase-angle curve is omitted because its errors are orders of magnitude larger than all learning-based models, making joint visualization impractical and visually misleading.

Across both voltage channels, the proposed model follows the true temporal evolution more faithfully than the baseline. It exhibits substantially reduced angular drift and diminished magnitude bias, aligning with the quantitative improvements reported in Table I.

To further evaluate the effectiveness of the proposed model, the experiments are conducted on the IEEE 123-bus system

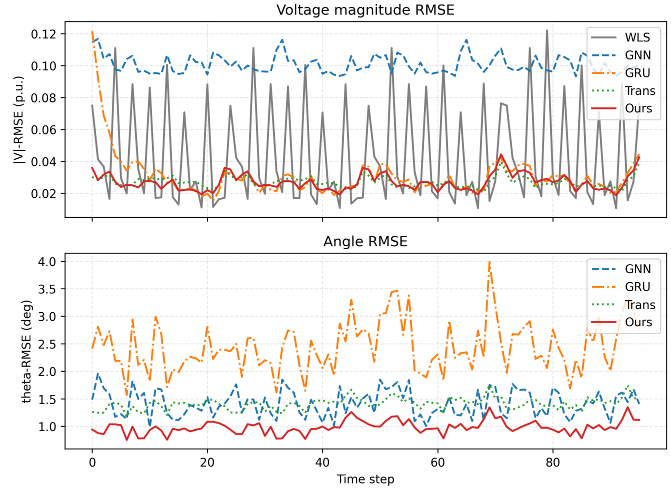

Fig. 3. RMSE of (a) voltage phase angle and (b) voltage magnitude over the last 95 test steps for WLS, GNN, GRU, Transformer, and TAGRU.

TABLE II
STATE-ESTIMATION ACCURACY COMPARISON ON THE IEEE 123-NODE SYSTEM (THREE-PHASE UNBALANCED, Y-NODE STATE).

Algorithm	θ-RMSE (°)	$\|V\|$-RMSE (p.u.)	θ-MAE (°)	$\|V\|$-MAE (p.u.)
GNN	1.56	0.060	1.47	0.043
GRU	0.52	0.036	0.33	0.026
Trans	0.66	0.012	0.48	0.010
Ours (TAGRU)	**0.45**	**0.013**	**0.31**	**0.009**

using the same parameter settings for the proposed method and the baseline models. The corresponding results are reported in the table and the figure.

For three-phase unbalanced systems, we adopt a Y-node (bus-phase) formulation where each bus-phase is treated as a node; accordingly, $n = |\mathcal{V}|$ denotes the number of Y-nodes and the state dimension is $2n$.

Compared with the modified IEEE 33-bus setup, the IEEE 123-node scenario is configured with denser and more informative measurements, which reduces the overall estimation uncertainty. Consequently, the absolute errors of all neural estimators decrease noticeably, with typical reductions on the order of 50–60% relative to IEEE 33-bus results. Moreover, the performance gaps among sequence models become smaller because the richer sensing alleviates the reliance on temporal dynamics to compensate for missing information. In contrast, the graph-only baseline tends to degrade relatively in the unbalanced three-phase setting, since it does not explicitly model inter-phase coupling in the Y-node formulation. Overall, our model remains the most accurate, benefiting from jointly leveraging temporal structure and the unbalanced-network representation.

These results further demonstrate that enforcing physics-consistency constraints allows the model to better capture the intricate temporal–topological correlations of distribution networks, yielding estimates that remain stable under long-horizon propagation and heterogeneous sensor conditions.

979-8-3315-8138-1

C. Robustness Evaluation under Asynchronous and Corrupted Measurements

We further evaluate TAGRU under realistic measurement degradations by transforming the synchronous test set into heterogeneous, asynchronous, and corrupted streams. The perturbations emulate three major sources of uncertainty: (i) heterogeneous sampling rates across PMU, SCADA, and pseudo-measurements, (ii) temporal misalignment induced by jitter and reporting delays, and (iii) missing or anomalous measurements caused by burst outages and outliers.

To control the difficulty of the robustness assessment in a systematic manner, the scenarios are grouped into three tiers, *L*, *M*, and *H*. The *L* tier introduces mild corruption that is representative of low-cost sensing deployments. The *M* tier increases the degree of misalignment, the length of burst-missing segments, and the magnitude of outliers. The *H* tier represents a worst-case setting with aggressive down-sampling, strong jitter, long burst gaps, and rare but extremely large outliers.

Table III reports the estimation accuracy across the three tiers. Several observations follow. First, the classical WLS estimator degrades substantially as the corruption severity increases. The $|V|$-RMSE increases from 0.390 $\mathrm{p.u.}$ in the *L* tier to 41.187 $\mathrm{p.u.}$ in the *H* tier, which indicates high sensitivity to asynchronous and missing measurements due to reliance on synchronous pseudo-measurements and sensitivity to poor observability. In the most challenging *H*-tier scenarios, WLS occasionally diverges under severe outliers and long missing segments, which results in extremely large RMSE values. These failure cases are retained in the reported statistics to reflect the limited robustness of WLS. Second, the GNN and sequence-based baselines (GRU and Transformer) exhibit improved robustness compared with WLS, but their performance still deteriorates as temporal corruption increases. For example, the θ-RMSE of GRU increases from 1.051° in the *L* tier to 1.758° in the *H* tier. This trend indicates that purely data-driven models remain sensitive to severe jitter and misalignment when temporal dependencies cannot be reconstructed reliably.

Most importantly, the proposed TAGRU maintains consistently strong performance across all tiers. The θ-RMSE increases only mildly from 0.993° in the *L* tier to 1.051° in the *H* tier, and the $|V|$-MAE remains below 0.025 $\mathrm{p.u.}$ even in the *H* tier. These results indicate that the asynchronous alignment module and physics-guided graph refinement mitigate the impact of heterogeneous sampling and missing bursts effectively. By leveraging network topology and physical constraints explicitly, TAGRU reconstructs coherent temporal features even when the raw measurement streams deviate substantially from ideal synchronous conditions.

Overall, the results confirm that TAGRU outperforms both classical and learning-based baselines under realistic robustness scenarios, while remaining accurate and stable even under extreme measurement corruption.

TABLE III
STATE-ESTIMATION ACCURACY COMPARISON BY TIER (L, M, H) ON THE MODIFIED IEEE 33-BUS SYSTEM.

Algorithm	θ-RMSE (°)	$\|V\|$-RMSE (p.u.)	θ-MAE (°)	$\|V\|$-MAE (p.u.)
Tier L				
WLS	26.274	0.390	16.929	0.069
GNN	2.712	0.106	2.530	0.075
GRU	1.051	0.034	0.712	0.024
Trans	1.455	0.028	1.104	0.024
TAGRU	**0.993**	**0.030**	**0.697**	**0.019**
Tier M				
WLS	27.024	1.631	17.071	0.090
GNN	2.938	0.112	2.725	0.081
GRU	1.353	0.039	0.880	0.029
Trans	1.855	0.033	1.390	0.029
TAGRU	**1.002**	**0.037**	**0.710**	**0.019**
Tier H				
WLS	31.265	41.187	18.691	0.766
GNN	3.150	0.118	2.915	0.086
GRU	1.758	0.045	1.050	0.055
Trans	2.355	0.039	1.680	0.076
TAGRU	**1.051**	**0.046**	**0.818**	**0.024**

IV. CONCLUSION

The proposed TAGRU integrates reliability-aware pooling, topology-aligned spatio–temporal modeling, and physics-consistency regularization under weak supervision. Experiments on the IEEE 33-bus and IEEE 123-node systems verify improved accuracy and robustness under heterogeneous and asynchronous measurements. Future work will focus on larger real-world feeders and online adaptation to topology and sensing changes.

ACKNOWLEDGMENT

This work is supported by Smart Grid-National Science and Technology Major Project 2024ZD0802200.

REFERENCES

[1] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[2] F. Schmidt and M. de Almeida, "A theoretical framework for qualitative problems in power system state estimation," *Electric Power Systems Research*, vol. 154, pp. 528–537, 2018.

[3] B. Azimian, S. Moshtagh, A. Pal, and S. Ma, "Analytical verification of performance of deep neural network based time-synchronized distribution system state estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 4, pp. 1126–1134, 2023.

[4] M. Huang, Z. Wei, G. Sun, and H. Zang, "Hybrid state estimation for distribution systems with ami and scada measurements," *IEEE Access*, vol. 7, pp. 120 350–120 359, 2019.

[5] X. Zhang, S. Ge, Y. Zhou, and H. Liu, "Deep learning framework for low-observable distribution system state estimation with multitimescale measurements," *IEEE Transactions on Industrial Informatics*, 2024.

[6] B. Habib, E. Isufi, W. van 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, 2023.

[7] J. Hu, W. Hu, D. Cao, S. Li, J. Chen, Y. Huang, Z. Chen, and F. Blaabjerg, "Robust multiarea distribution system state estimation based on structure-informed graphic network and multitask gaussian process," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10 599–10 612, 2024.

[8] H. Wu, Z. Xu, and M. Wang, "Unrolled spatiotemporal graph convolutional network for distribution system state estimation and forecasting," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 297–308, 2022.