

Graph Neural Network-Based State Estimation In Simulated Distribution Grids

Vlad-Alexandru Sănduleac
Faculty of Energy Engineering
National University of Science and
Technology Politehnica
Bucharest, Romania
vlad.sanduleac@upb.ro

Lucian Toma
Faculty of Energy Engineering
National University of Science and
Technology Politehnica
Bucharest, Romania
luciantomaro@gmail.com

Mihaela Albu
Faculty of Electrical Engineering
National University of Science and
Technology Politehnica
Bucharest, Romania
albu.marilena@gmail.com

Ion Tristiu
Faculty of Energy Engineering
National University of Science and
Technology Politehnica
Bucharest, Romania
ion_tristiu@yahoo.com

Andreea Iantoc
Faculty of Energy Engineering
National University of Science and
Technology Politehnica
Bucharest, Romania
andreea.iantoc@upb.ro

Ion Andronic
Directorate of Electricity
Ministry of Energy
Chisinau, Republic of Moldova
ion.andronic@energie.gov.md

Abstract— Distribution System State Estimation (DSSE) is a need for the power management of smart grids, where measurements may be missing or incomplete. With the increase of electricity generated from renewable energy sources (RES), the need for DSSE is even more relevant. Solutions to this complex problem might be based on Graph Neural Networks. Current methods in this class assume unchanging network conditions during training. For a given topology, we propose a new method based on generating new network. OpenDSS is then used to simulate power flows which are then used to train a Graph Attention Network as a state estimator in a masked reconstruction problem. The resulting state estimator is demonstrated for the IEEE-33 topology. Finally, we analyze the energy efficiency of training and effect of varying masking rates on the estimation accuracy.

Keywords— distribution system, graph attention network, state estimation, partial observability, network reconfiguration

I. INTRODUCTION

The proliferation of smart grid technologies, together with the uptake of RES-based electricity by distribution grids, has increased the complexity of the power system. Due to these developments, DSSE becomes an increasingly important task in the management of smart grids, particularly as power flows may become reversed. DSSE is useful in tasks such as demand response and distribution automation [1].

Recent advancements have allowed for far better time resolution data to be collected from smart metering infrastructure through SCADA. However, measurements may be missing or incomplete due to communication or equipment failures. In distribution networks with limited redundancy, such missing data can significantly degrade estimation performance.

However, one drawback of classical solutions, even when the loss of data is minimized, is that the information has to be correlated with data provided with low and very low reporting rates, still in use due to the low capabilities of power system operators to address the realm of big data. One can address this issue by fostering the high reporting rates accurate measurement information as provided by the new generation of smart meters [2].

To this end, for the DSSE scope, neural networks (NN) change the paradigm from “Model-based” to “Data-Driven”. Some of the backdrops of classical State Estimators (SE) comparing with neural networks are: 1) at higher level of missing data (imputation), their performance is decreasing, while NN are more robust and can better cope with missing or inaccurate data [3] or [4], the latest using Wasserstein generative adversarial network (WGAN); 2) R and X parameters of the network are necessary and sometimes difficult to be given, while NN uses a “model-free” approach, as these values are not necessarily needed or can be estimated [5], as NN can learn the relation between voltages and active powers in the grid without knowing them; 3) changes in network topology represent challenges in classic SE, while NN are better adapted through technologies such as Graph Attention Network (GAT) and can bring good results to accommodate the operational characteristics of active distribution networks [6]; 4) the speed of solving a state estimation problem is less appropriate for real-time in comparison with NN which have high computational efficiency [7] and potential to make estimation even for high dynamic processes such as inertia in modern power systems [8].

As pointed before, traditional DSSE methods suffer more heavily from large amounts of missing data compared to deep learning methods, which can learn structural correlations directly from data.

In this work, we will investigate a graph neural network (GNN)-based approach for DSSE in the case of partial observability, treated as a masked measurement reconstruction problem. Additionally, we will study the effect of varied observability during training and testing, as well as make a comparison of accuracy against energy consumption and training time.

II. STATE OF THE ART

A. Weighted Least Squares

Weighted Least Squares (WLS) estimators are the most common method of power system state estimation. WLS is performed by minimizing the difference between estimated and real measurements, being solved iteratively through the Gauss-Newton method, due to the nonlinear relation between

active and reactive powers to the voltage measurements. In transmission system state estimation, where measurements are often highly redundant and observable, the WLS method is well established [9].

However, WLS-based methods face significant challenges in distribution networks due to still limited measurement redundancy and higher R/X ratios. A common mitigation for distribution system state estimation is to introduce pseudo-measurements to compensate for missing data and ensure convergence. However, these methods introduce additional tuning complexity and make these methods less robust [10].

WLS methods differ depending on the chosen state-describing variables, such as node-voltage and branch-current formulations [10]. Estimators using the branch-current formulation are considered more computationally efficient in radial networks but remain sensitive to measurement weighting and ill-conditioning issues.

Despite various techniques to mitigate the issue, the performance of WLS methods degrades significantly when faced with high levels of missing data or poorly tuned pseudo-measurements [10].

B. Deep Learning Models

An example of a neural network-based method for state estimation utilizes a Bayesian inference method to determine a minimum mean squared error (MMSE). The Monte Carlo method is then used to train a multi-layer perceptron to estimate the aforementioned MMSE. This method additionally employed detection of bad data, which was found to significantly improve the quality of predictions [11].

More recently, Transformer architectures have been applied in the domain of DSSE. Transformer models utilize self-attention mechanisms to learn a model of the internal dependencies between data, additionally allowing for easier interpretation of model decisions. An example of such a model applied in literature is the Transformer-Bi-LSTM model which combines self-attention layers with LSTM recurrent neural units to capture the spatial and temporal characteristics of measurement data. This method has been demonstrated to have significantly improved robustness in the case of missing measurements, performing best compared to other deep learning methods such as non-hybridized LSTM, GRU or Transformer models. Additionally, as masking rates were increased to 50%, the increase in error was found to be low [12].

C. Graph Neural Networks

Another architecture for deep learning which has been explored recently for DSSE is the graph neural network (GNN). GNNs utilize a message-passing architecture to update the states of nodes in a graph, allowing a model to innately leverage the graph structure of a task such as in distribution networks. One such approach, termed GraphSAGE, maps the buses and lines of a distribution network to the edges and nodes, respectively, of the graph used in the GNN. Learned node embeddings are generated through neighborhood aggregation to create an internal representation of power measurements [13].

GNN-based state estimators remain effective in the case of lower observability without requiring pseudo-measurements [13].

One implementation of a GNN, specifically a graph convolutional neural network (GCNN), has used Matpower to generate steady-state regime power flows on IEEE 39, IEEE 57 and IEEE 118 topologies, while in this case the ability of the state estimator to correct distorted measurements from cyber-attacks and noise was also studied. In this work, edge features were incorporated, and the graph description was split into a real and imaginary part, and the GCNN was found to perform substantially better than WLS, in particular by its ability to better isolate the effects of cyber-attacks around the nodes attacked [14].

A method tested on IEEE 33, 37 and a 559 node distribution system used a similar GNN method for network topology reconfiguration, studying a wider range of masking rates between 15% and 35% on a temporal graph convolutional neural network (TGNN). The method was capable of classifying the status of switches in the topology with an accuracy of 92.1%, while the MAPE of voltage amplitudes in the case of 35% masking was as low as 0.2% [15]. Notably, pseudo-measurements associated with the power lines (also called edges in the convention of mathematical graphs) were not estimated.

D. Graph Attention Networks

Another extension of GNNs utilizes attention layers to improve the message passing step, resulting in an architecture named Graph Attention Networks. This architecture computes attention coefficients between a node and its neighbors, allowing the model to determine the importance of various edges to the estimation of a node's state [16]. This stands in contrast with other GNN methods, such as GCNN, where aggregation is typically done with a fixed weight.

In [17] the method performed state estimation using GATs over the IEEE 33-bus distribution network topology. The topology was reconfigured using set daily load and solar power curves, together with different combinations of settings for section and interconnection switches, allowing for both radial and looped topologies. The MATPOWER toolbox was then used to simulate steady flow data for these configurations, and the results were masked by up to 20% missing data to create the state estimation task. The GAT model was found to have estimated voltage amplitude with a MAPE of 0.2% [17]. Only node measurements, such as injection power, were included in the estimation.

E. Edge Features

Traditional GNN architectures assume attributes which are assigned to nodes, but not to edges. However, in distribution networks, lines have meaningful attributes such as impedance, power flow and current, which would be attributed to the edges of the graph representation. The incorporation and estimation of these attributes into state estimation models can improve their effectiveness. To this end, several techniques that extend GNNs and more specifically GATs exist by modifying their message passing method. Edge features may be incorporated by concatenation to node embeddings or by adjusting attention coefficients. Edge embeddings may also be treated in a separate step from node embeddings, allowing the model to treat these attributes explicitly.

III. EXPERIMENTAL SETUP

A. Network Generation

The study is conducted on a modified IEEE-33 distribution network operating at 10 kV, which can be seen in Fig. 1. For each experimental configuration, the network is reinitialized with a new set of elements at each bus, including consumer loads, photovoltaic arrays, and battery storage systems, sampled within predefined limits.

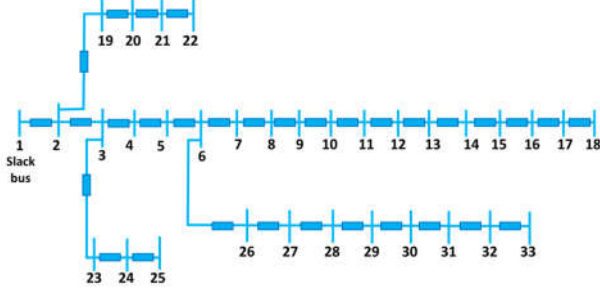


Fig. 1. The IEEE-33 distribution network, as used in this paper

The first step in the generation of a configuration is the initialization of the topology. The IEEE-33 network load-flow is calculated with the electric power distribution system simulator OpenDSS [18] using its default line parameters (R , X , B , and line length), adapted for the 10kV rated voltage, and randomly disturbed by up to 10% to introduce structural variability between configurations. The slack bus is assigned with a voltage magnitude sampled uniformly in the range $[1.0, 1.05]$ p.u., and a fixed short-circuit rating of 9000 MVA.

For each configuration, network elements are allocated by sampling, independently for each bus, the number of loads (0–2), the connection of photovoltaic (PV) installations (30% probability and uniform distribution over the scenarios), and storage units (10% probability). Each element is assigned a maximum capacity drawn from an uniform distribution in the predefined ranges, forming the structural configuration of the network in that configuration.

After element allocation, a target net system power at full loading is selected in the range $[-2000 \text{ kW}, 12000 \text{ kW}]$, constraining the configuration within a realistic operational envelope. A naïve / interim net power is computed as the algebraic sum of maximum load consumption and maximum PV production, assuming no losses and inactive storage dispatch. The maximum capacities of PV units and loads are then iteratively scaled until the net system power matches the selected target within a tolerance of 1 W. This ensures global power consistency prior to simulation.

The full-loading configuration is subsequently solved using OpenDSS. If the AC power flow does not converge, the configuration is rejected and regenerated from the element allocation stage, ensuring that only physically feasible base cases are retained.

For each accepted configuration, 10 operating scenarios are generated by scaling load and PV output between 0.3 and 1.0 of their maximum capacity. Each modified scenario is independently solved using OpenDSS to produce physically consistent measurements. Scenarios that fail to converge are regenerated; however, if more than 50% of scenarios in a configuration fail to converge, the entire scenario is discarded and regenerated from the element allocation stage.

Using the above method, a dataset of 3000 configurations is generated with a 70%/10%/20% train-validation-test split. A set of parameters for the generation of this dataset is available in Table I.

TABLE I. NETWORK ELEMENTS

Element type	Placement Rule	Capacity Range	Other
Load	0-2 per bus	50-500kW	$0.85 < \text{Power factor} < 0.95$
PV	30% prob. per bus	10-2000kW	$< 60\%$ total load
Storage	10% prob. per bus	10-500kW	Energy capacity considered sufficient

The dataset is generated using a fixed seed to ensure reproducibility across experiments.

B. Masking Method

From the OpenDSS simulations, measurements are extracted on power lines (P_{line} , Q_{line} , I_{line}) and buses (U , U_{angle} , P_{bus} , Q_{bus} , I_{bus}), as well as topology parameters such as line parameters (R , X) and installed loads ($P_{maxload}$, $P_{installedPV}$, $P_{maxstorage}$).

During learning, a set percentage M of line and bus measurements are stochastically hidden in the input of the model, being replaced (masked) with learned mask embeddings. The masked inputs vary with each epoch.

C. Model Architecture

The proposed model is a modified, dual-stream, attention-based graph neural network. Node and edge attributes – line, bus and topology parameters – are embedded using multi-layer perceptrons into an embedding space with a hidden dimension of 512 for nodes and 256 for edges respectively. Node and edge embeddings are performed separately, and at this stage learned masked embeddings are inserted.

The attention-based graph neural network backbone consists of 8 layers of dual-stream graph attention blocks. Each block maintains two independent message-passing streams, one stream performing spatial attention over physical line edges, and the other stream performing inter-scenario attention over abstract edges connecting the same bus across different scenarios, each producing distinct embeddings.

The output embeddings of these two message-passing streams are concatenated and passed through multi-layer perceptron block to generate the result embedding. For the output of the model, two decoders are used. A node head maps the final node embeddings to reconstruct bus measurements, while an edge head maps the concatenated edge embedding and source and destination node embeddings to reconstruct power line measurements. The total number of learnable parameters of the model as described is 12.7 million.

IV. RESULTS

The model was trained for 600 epochs on a 128GB unified RAM DGX Spark machine [19]. A cosine annealing with warm restarts scheduler was used to control the learning rate, with a maximum learning rate of 10^{-4} and a minimum learning rate of 10^{-7} . This type of scheduler reduces the learning rate until the minimum is reached, then resets it, doubling the time until the minimum is reached every time. This scheduler was

chosen to avoid local minima during training. For the majority of the tables presented, unless noted otherwise, the model tested was at the point of the best/lowest validation loss.

The training process, shown by the training loss in Fig. 2, matches this expectation. Until epoch 50, training loss decreases significantly while validation loss remains in a high plateau, indicating a regime of significant overfitting. After this epoch, the training rate is restarted by the scheduler, upon which validation loss drops and training loss jumps to a significantly higher value. Training loss continues to diminish from this value, although remaining higher than during the first 50 epochs, reinforcing the idea that the model was overfitting in this period. Validation loss steadily diminishes during this period as well, eventually dropping below training loss.

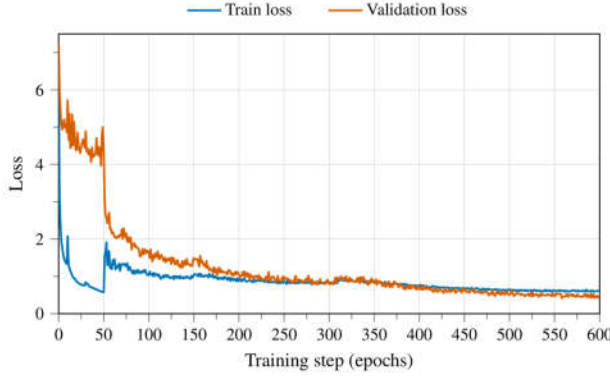


Fig. 2. Training and validation loss chart for the case of 50% masking rate

Total training time varied between 24 and 36 hours. During this time, GPU power draw was recorded. For the entire duration of training, the 50% masking rate model has run on the DGX Spark machine using 0.498 kWh of electric energy.

For evaluation, we adopted a range-based, normalized mean absolute error (NMAE), where errors are reported relative to the largest absolute value for the given parameter in the dataset.

$$NMAE = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\max_j |y_j|} \right| \quad (1)$$

Additionally, a normalized variance, termed σ has been calculated, which is similarly calculated relative to the largest absolute given value for a given parameter:

$$\sigma = \frac{100}{\max_j |y_j|} \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2} \quad (2)$$

$$e_i = y_i - \hat{y}_i \quad (3)$$

The test results, measured as error percentage of the maximum absolute value of the given parameter, were adequate, indicating that the model has learned a good representation of power flow in the IEEE-33 network. Table II shows that the best performing parameter was voltage angle, with a range-normalized mean absolute error of 0.16%. We also observe good performance on reactive power parameter Q_{line} , with a mean absolute error of 0.43%. The worst performing parameters were bus power P_{bus} and current I_{bus} .

TABLE II. TEST RESULTS (CASE OF 50% MASKING RATE)

Parameter	$NMAE_{50}$	σ_{50}
U	0.79%	1.08%
U_{angle}	0.16%	0.21%
P_{bus}	1.29%	2.54%
P_{line}	0.95%	1.47%
Q_{bus}	0.60%	0.92%
Q_{line}	0.43%	0.66%
I_{bus}	1.52%	6.16%
I_{line}	0.98%	1.53%

We also observe that certain bus parameters (notably P_{bus} , I_{bus}) tend to perform worse than line parameters. Particularly in the case of I_{bus} the ratio between the standard deviation and the mean absolute error is high, indicating that there are more high residual outliers, but very low errors on average. We consider that this is due to the GNN placing a higher importance on topology parameters (line parameters) than local ones.

As seen in Table III, the mean average error of most variables remains low even with a higher masking percentage, indicating that even with lower observability there is sufficient redundancy to reconstruct the state without significant degradation. Voltage estimation remains stable even in the 70% observability scenario. Exceptions to this are U_{angle} and Q_{line} , where NMAE remains stable at the higher masking rate of 70%, suggesting a regularization effect from the increased masking rate may be counteracting the effect of lower observability, similar to the effect of dropout.

TABLE III. TEST RESULTS (70% MASKING)

Parameter	$NMAE_{70}$	σ_{70}
U	0.95%	1.43%
U_{angle}	0.16%	0.24%
P_{bus}	1.88%	3.61%
P_{line}	1.21%	1.87%
Q_{bus}	0.70%	1.12%
Q_{line}	0.48%	0.76%
I_{bus}	2.17%	10.23%
I_{line}	1.41%	2.13%

The greatest improvement is found in NMAE of bus and line currents, with a roughly 30% drop in mean absolute error, while the lower drop in standard deviation suggests that the mean absolute error of outliers improves less. Conversely, we observe that the standard deviation of voltage angle and line power errors improves more than the average error, suggesting that additional observations contribute more to correcting outliers.

The highest improvements, in terms of reducing percentual mean errors, were to be found in voltage angles and bus power mean absolute error. In Table IV, we have analyzed the effectiveness of longer training times in the 50% masking experiment by saving checkpoints every 100 epochs and evaluating them against the test set. We observe that, as

expected, the error rate diminishes monotonically, with the exception of voltages which will be discussed later. We observe that P_{line} exhibits the greatest relative decrease in mean errors from epoch 100 to 600, with mean absolute error decreasing by 56% and thus benefitting the most from the extended training, while Q_{bus} benefits the least, with mean absolute error decreasing by only 38%.

TABLE IV. NMAE₅₀ ACROSS EPOCHS (CASE OF 50% MASKING RATE)

Epoch	U	P_{line}	P_{bus}	Q_{line}	Q_{bus}
100	1.63%	2.12%	2.85%	0.79%	0.98%
200	1.21%	1.58%	2.26%	0.55%	0.75%
300	0.95%	1.41%	1.80%	0.50%	0.66%
400	0.97%	1.17%	1.65%	0.52%	0.66%
500	0.83%	0.94%	1.43%	0.44%	0.63%
600	0.78%	0.93%	1.30%	0.43%	0.61%

Similarly, in Table V we find that in the case of the 70% masking experiment, testing errors decrease monotonically. The impact of extended training on the various parameters is more uniform, with the mean absolute error of the parameter most affected, in this case U , decreasing by 55% while the mean error of the least affected, Q_{bus} , decreases by 15% from epoch 100 to epoch 600.

TABLE V. NMAE₇₀ ACROSS EPOCHS (CASE OF 70% MASKING RATE)

Epoch	U	P_{line}	P_{bus}	Q_{line}	Q_{bus}
100	2.01%	2.28%	2.88%	0.80%	0.84%
200	1.63%	2.00%	2.47%	0.65%	0.84%
300	1.10%	1.61%	2.16%	0.53%	0.75%
400	1.10%	1.51%	2.35%	0.55%	0.78%
500	0.97%	1.27%	2.02%	0.49%	0.73%
600	0.90%	1.21%	1.91%	0.48%	0.71%

Observing the testing error for voltage during the training of the 50% and 70% masking cases we observe that mean errors continue to decrease substantially by the final epoch, indicating that the model would benefit from a longer training regime.

We have analyzed the energy efficiency of extended training by evaluating the absolute decrease in normalized mean absolute error against estimated energy consumed by the given epoch, given as energy required to decrease NMAE by 1% after epoch 100. In Table VI, we find that training for 100 epochs yields the most efficient improvement in performance in terms of energy expenditure, where energy efficiency is defined by the formula:

$$\eta_x = \frac{E_{100 \rightarrow x}}{100 \cdot |NMAE_U(100) - NMAE_U(x)|} \quad (4)$$

In Fig. 3 we observe the evolution of the standard deviation of the voltage reconstruction error on the validation set during training. We observe a rapid initial decrease, the period of initial training until around epoch 100, followed by a short plateau until epoch 150, indicating that this is the period where most major dependencies are learned. After this point, the error continues to decline slowly.

TABLE VI. ENERGY EFFICIENCY COMPARED TO EPOCH 100 (kWh/1% IMPROVEMENT) FOR 50% MASKING RATE

Epoch	U	P_{line}	P_{bus}	Q_{line}	Q_{bus}
100	—	—	—	—	—
200	0.029	0.022	0.02	0.051	0.052
300	0.036	0.034	0.023	0.082	0.075
400	0.055	0.038	0.03	0.133	0.113
500	0.061	0.041	0.034	0.139	0.139
600	0.071	0.051	0.039	0.168	0.164

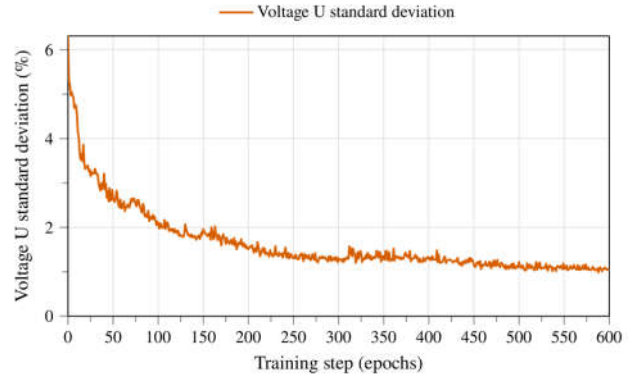


Fig. 3. Standard deviation of voltage error during training using the validation set

Similarly, in Fig. 4 we observe that mean absolute error exhibits a similarly rapid decline in the early periods of training. Improvements following epoch 300 are slower, and the stability in the final epochs of training indicates that the model has a robust internal representation of the problem of state estimation.

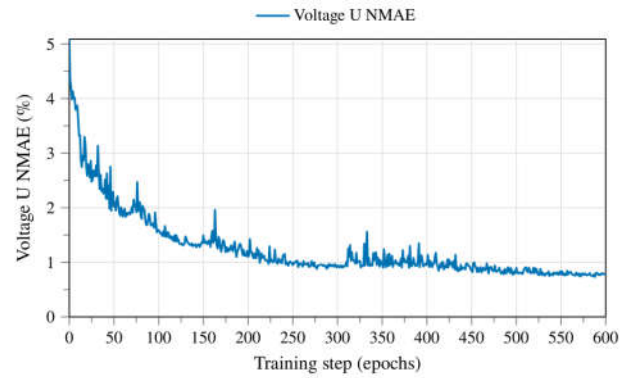


Fig. 4. NMAE₅₀ of voltage during training using the validation set

These results correspond to the data in literature. Although comparison is difficult due to different dataset construction strategies, rated generated power, and different network topologies, a comparison of power estimation accuracy with Transformer-Bi-LSTM appears to favor our model and approach.

V. CONCLUSIONS

In this work, we demonstrate that an attention-based graph attention network is capable of learning the necessary

representation for state estimation and reconstruction in a simulated IEEE-33 medium voltage network with renewable energy resources. Additionally, we have analyzed the relationship between training time, energy consumption and state estimator accuracy, allowing for a more energy conscious approach to training future models.

The analysis shows good estimation results for both 50% and 70% missing / masked data, thus demonstrating the robust behavior of GNN approach. In both cases the estimation of node voltages is below 1.0%, while reactive powers on lines have the second-best estimation, of less than 0.5% and the active powers estimation is also under an average of 2.0% after being checked all tested network scenarios.

There is only a small degradation of estimations for most variables when depreciating missing data from 50% to 70%, which also is a sign of promising behavior for DSSE implemented with GNN.

The proposed approach focuses on fully stochastic masking of collected measurements on a single, static, network architecture. Future work could explore correlated masking of measurements in various simulated scenarios, such as lacking meters capable of recording those measurements or cyber-attacks. Future implementations will consider more options for the implementation of the DSSE, e.g. the use of Physics Informed Neural Networks (PINN) and other techniques.

ACKNOWLEDGMENT

This work has received funding from the European Union's Directorate-General for Communications Networks, Content and Technology program, under grant agreement No 101158494 - ECLIPSE (Energy Consumption Reduction based on Open-source Reference framework). This text reflects only the author's views, and the Commission is not liable for any use that may be made of the information contained therein.

REFERENCES

- [1] F. Ahmad, A. Rasool, E. Ozsoy, R. Sekar, A. Sabanovic, and M. Elitaş, "Distribution system state estimation-A step towards smart grid," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2659–2671, Jan. 2018, doi: <https://doi.org/10.1016/j.rser.2017.06.071>.
- [2] Sanduleac, Mihai, Lucas Pons, Giampaolo Fiorentino, Rares Pop, and Mihaela Albu. "The unbundled smart meter concept in a synchro-SCADA framework." In *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1-5. IEEE, 2016.
- [3] Sidhartheneey Nayak; Divyanshi Dwivedi; K. Victor Sam Moses Babu; Chandrashekhhar Narayan Bhende; Pradeep Kumar Yemula; Pratyush Chakraborty, Data Imputation Using Self Attention Based Model for Enhancing Distribution Grid Monitoring and Protection Systems, *IEEE Transactions on Instrumentation and Measurement* (Volume: 73), 01 March 2024
- [4] Zhaohui Sun; Yuxin Duan; Lei Su; Li Xin; Runyu Cai; Liangjin Hua, Dynamic State Estimation for Distribution Networks Considering Large-Scale Missing Measurement Data Reconstruction, 2025 2nd International Symposium on New Energy Technologies and Power Systems (NETPS), 23-25 May 2025, Hangzhou, China
- [5] Guoxin Li; Sirui Cao; Tianyu Xu; Haozhe Ma; Jiarun Li, Data-Driven Topology Identification and Parameter Estimation for Distribution Networks via Multi-Scale Feature Enhancement, 2025 5th International Conference on Electrical Engineering and Control Science (IC2ECS), 05-07 December 2025, Beijing, China
- [6] Lei Song; Chunguang Lu; Kailiang Ge; Chenkai Xu; Xianbo Wang, An Active Distribution Network Topology Identification Method Based on GAIN-GAT, 2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2), 15-18 December 2023, Hangzhou, China
- [7] Weidong Zhong; Weiliang Liu; Yunfei Yao; Wei Yu, State Estimation of New Distribution Network Based on Deep Neural Network, 2023 International Conference on Power System Technology (PowerCon), 21-22 September 2023, Jinan, China
- [8] Faisal Albeladi; Kamal Basulaiman; Masoud Barati, Beyond GCN: Accurate Real-time Inertia Estimation with Multi-Head Spatial-Temporal Graph Attention Network Coupled with LSTM, *IEEE Transactions on Industry Applications* (Early Access), pages 1-10, 07 October 2025, IEEE
- [9] L. Holten, A. Gjelsvik, Sverre Aam, F. F. Wu, and W. Liu, "Comparison of different methods for state estimation," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1798–1806, Nov. 1988, doi: <https://doi.org/10.1109/59.192998>.
- [10] A. Primadianto and C.-N. Lu, "A Review on Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, Sept. 2017, doi: [10.1109/TPWRS.2016.2632156](https://doi.org/10.1109/TPWRS.2016.2632156).
- [11] Kursat Rasim Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian State Estimation for Unobservable Distribution Systems via Deep Learning," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910–4920, Nov. 2019, doi: <https://doi.org/10.1109/tpwrs.2019.2919157>.
- [12] Y. Raghuvamsi and K. Teeparthi, "Distribution system state estimation with Transformer-Bi-LSTM-based imputation model for missing measurements," *Neural Computing and Applications*, vol. 36, no. 3, pp. 1295–1312, Oct. 2023, doi: <https://doi.org/10.1007/s00521-023-09097-5>.
- [13] B. Zhai, D. Yang, B. Zhou, and G. Li, "Distribution System State Estimation Based on Power Flow-Guided GraphSAGE," *Energies*, vol. 17, no. 17, p. 4317, Aug. 2024, doi: <https://doi.org/10.3390/en17174317>.
- [14] Park, Sangwoo, Fernando Gama, Javad Lavaci, and Somayeh Sojoudi. "Distributed power system state estimation using graph convolutional neural networks." In *Proceedings of the Hawaii International Conference on System Sciences*. 2023.
- [15] Rahul Madbhavi, B. Natarajan, and B. Srinivasan, "Graph Neural Network-Based Distribution System State Estimators," *IEEE Transactions on Industrial Informatics*, pp. 1–10, Jan. 2023, doi: <https://doi.org/10.1109/tii.2023.3248082>.
- [16] Petar Veličković, G. Cucurull, A. Casanova, A. Romero, Pietro Liò, and Y. Bengio, "Graph Attention Networks," *arXiv (Cornell University)*, Oct. 2017, doi: <https://doi.org/10.48550/arxiv.1710.10903>.
- [17] Z. Ren, X. Chu, and H. Ye, "State Estimation for Reconfigurable Distribution Systems Using Graph Attention Network," *IEEE Access*, vol. 11, pp. 107237–107250, Jan. 2023, doi: <https://doi.org/10.1109/access.2023.3316017>.
- [18] OpenDSS - Electric power Distribution System Simulator (DSS), <https://www.epri.com/pages/sa/opensdss>, accessed on 15.03.2026
- [19] DGX Spark, <https://www.nvidia.com/en-us/products/workstations/dgx-spark/>, accessed on 15.03.2026