

Explainable Multi-Fidelity Bayesian Neural Network for Distribution System State Estimation

X Chen

September 2025

Applied Energy

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Explainable Multi-Fidelity Bayesian Neural Network for Distribution System State Estimation

Jinxian Zhang, *Student Member, IEEE*, Junbo Zhao, *Senior Member, IEEE*, Gang Cheng, *Member, IEEE*, Alireza Rouhani, *Senior Member, IEEE*, Xiao Chen

Abstract—Distribution System State Estimation (DSSE) is frequently constrained by limited real-time measurements, the uncertainties introduced by distributed energy resources, and the presence of bad data. To address them, this paper proposes an enhanced Multi-Fidelity Bayesian Neural Network (MFBN) DSSE approach. A low-fidelity layer based on a Deep Neural Network (DNN) is first pre-trained on pseudo-measurement data to learn fundamental state features. Subsequently, a high-fidelity Bayesian Neural Network (BNN) layer leverages limited but high-quality real-time measurements to refine these features, thereby achieving accurate DSSE. In addition, the deep Shapley Additive Explanation (SHAP) is developed to quantify the influence of measurement data on DSSE through dual perspectives of global feature importance and local nodal contributions, establishing a hierarchical explainability framework for machine learning-based DSSE. Comparative studies conducted on the IEEE 13-bus system and a real-world 2,135-node system from Dominion Energy demonstrate that the proposed method excels in estimation accuracy, even under situations of high noise levels, bad data, and missing data. Further comparisons with Weighted Least Squares (WLS) and other machine learning-based DSSE approaches verify that the proposed framework offers higher accuracy, improved interpretability, and enhanced robustness.

Index Terms—Interpretability, distribution system state estimation, robustness, multi-fidelity Bayesian neural networks, data fusion, measurement noise, bad data.

I. INTRODUCTION

DISTRIBUTION System State Estimation (DSSE) is the process of inferring the values of state variables utilizing limited measurement data [1]. Over the past few decades, distribution systems have undergone extensive modernization facilitated by the integration of renewable energy, open energy markets, and communication-enabled advanced metering infrastructures [2]. Consequently, the accuracy and reliability of DSSE play a pivotal role in safeguarding the operational integrity of distribution systems [3].

State Estimation (SE) was first applied in transmission systems and has undergone significant development [4]. Var-

ious algorithms, such as Weighted Least Squares (WLS) and Kalman Filter-based methods [5], [6], [7], [8], are widely used. However, distribution systems typically have fewer measurement devices, resulting in lower observability [9]. To increase measurement redundancy, pseudo-measurements based on historical data or standard load profiles are often employed [10], [11], [12], but their low accuracy can significantly degrade DSSE performance [13].

DSSE can also suffer from bad or missing data caused by communication failures, congestion, or cyberattacks [14]. Several probabilistic and statistical methods have been proposed to enhance robustness [15], [16], [17], including approaches based on Projection Statistics (PS) [15] and M-estimators that combine WLS and Weighted Least Absolute Value (WLAV) [16]. While robust estimators can mitigate the impact of bad data, they often demand higher computational costs [17].

Moreover, conventional DSSE methods rely heavily on accurate network parameters and topology data. In practice, outdated or inaccurate information can reduce their effectiveness. For instance, a convex relaxation technique can recast WLS as a semidefinite program to handle bad data [18], but requires strict topology details. Similarly, methods like Least Median of Squares (LMS) [19], Least Absolute Value (LAV) [20], and Generalized Maximum-likelihood (GM) [15] show strong robustness but typically need high measurement redundancy and precise system information. Finally, the numerical stability of these optimization-based DSSE methods is sensitive to factors like initial values, leverage measurements, and inappropriate weighting of different measurement classes [21].

In recent years, data-driven methods have attracted considerable attention for solving DSSE problems, with efforts focused on both pseudo-measurement generation and SE surrogate models. For pseudo-measurement generation, machine learning algorithms such as Artificial Neural Network (ANN) [22], Convolutional Neural Network (CNN) [23], Support Vector Machine (SVM) [24], k-means cluster [25] and Generative Adversarial Network (GAN) [26] have been used to produce more accurate pseudo-measurements, which helps maintain observability and accuracy in DSSE. However, these approaches still fundamentally rely on conventional estimators for system state inference, which makes it difficult to eliminate the dependence on accurate network parameters and topology data. As a result, more research has shifted toward surrogate SE models. Unlike conventional DSSE methods, data-driven DSSE surrogate models can learn the relationships between measurements and state variables directly from historical data

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was supported in part by Dominion Energy and the U.S. Department of Energy Solar Energy Technology Office under award 10422.

Jinxian Zhang, Junbo Zhao and Gang Cheng are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA. (Email: jinxian.zhang@uconn.edu, junbo@uconn.edu, mag24005@uconn.edu).

A. Rouhani is with Dominion Energy, Richmond, VA, 23290, USA (Email: Alireza.Rouhani@dominionenergy.com).

X. Chen is Lawrence Livermore National Laboratory, Livermore, CA 94550, USA (Email: chen73@llnl.gov).

[27], [28]. By leveraging extensive datasets during training, these models can make significant progress in state estimation. Consequently, they depend less on the network parameters in distribution systems than traditional DSSE methods. Neural Networks (NNs) and their variants, such as Deep Neural Networks (DNN) [29], Graph Neural Networks (GNNs) [30], [31], Physics-informed GNN [32] and Physics-Aware Neural Networks (PAWNNs) [33] have demonstrated good performance in DSSE while exhibiting low dependency on distribution system parameter information. However, due to their purely data-driven nature, surrogate models are highly susceptible to inaccuracies when encountering bad or missing data, which can significantly impact their performance. By representing the estimated state as a probability distribution rather than a single estimated value, Bayesian state estimation effectively captures measurement uncertainties, thereby enhancing the robustness of DSSE. Consequently, several studies have proposed data-driven DSSE models based on Bayesian inference [34], [35], which enhance robustness by representing the estimated states as probability distributions rather than single-point estimates. For example, in [34], Bayesian inference is integrated to update the parameters of a DNN-based DSSE model, significantly improving its robustness. Similarly, in [35], a Bayesian probabilistic weighted averaging method is proposed to update the parameters of a GNN, enhancing the model's robustness to topology changes and bad data.

Although data-driven DSSE methods are model-free, they usually require large amounts of high-quality training data to ensure reliable DSSE performance. In addition to bad or missing data, data-driven DSSE approaches [36], [37] often involve multiple data sources such as pseudo-measurements [22], [38], advanced metering infrastructure (AMI), supervisory control and data acquisition (SCADA) systems, and distribution phasor measurement units (D-PMUs). These sources typically have different sampling frequencies, accuracy levels, and measurement types, which makes integrating multi-source heterogeneous data for DSSE challenging. A single deep learning model is often influenced by the portion of the dataset with the lowest accuracy, which reduces overall estimation performance. To address this issue, our previous work [37] proposed a deep multi-fidelity Bayesian approach that fuses limited SCADA or AMI data with pseudo-measurements and uses a deep Gaussian Process (GP) to estimate distribution system voltages, leading to a significant improvement in DSSE performance under complex scenarios. This method demonstrates the strong potential of multi-fidelity learning for DSSE with multi-source data. By applying multi-fidelity learning, a single fidelity deep learning model can be integrated into a hybrid framework designed for data with different noise levels, effectively handling the challenges posed by multi-source datasets.

Data-driven DSSE models are promising but often difficult to interpret, which is critical for safety-critical power systems. Explainable Artificial Intelligence (XAI) has been proposed and utilized to ensure reliable results that are understandable to human operators [39]. SHapley Additive exPlanation (SHAP), a widely used XAI technique in energy and power systems [40], has been applied in areas such as grid disturbance

event identification [41], emergency control [42], transient stability assessment [43], and reactive power optimization [44]. Explainable DSSE clarifies how each input facilitates the state estimation process and quantifies its impact on the output, thereby offering potential value for ensuring trust and transparency in future advanced grid applications, including real-time monitoring, fault diagnosis, and data quality assessment.

Although many approaches have been proposed to tackle the aforementioned challenges in DSSE, no generic methodology exists that can simultaneously address the issues of multi-source data integration, measurement uncertainty, and interpretability. To address the interpretability challenges of data-driven DSSE methods while ensuring strong robustness against bad data and adaptability to multi-source data in distribution systems, we propose and implement a novel, interpretable, multi-fidelity-based robust DSSE algorithm designed for unbalanced three-phase systems. The multi-fidelity structure enhances observability, BNN ensures robustness, and SHAP provides transparency, resulting in a unified framework tailored to real-world DSSE challenges. The main contributions of this work are as follows:

- 1) The proposed DSSE method leverages multi-fidelity learning, where pseudo-measurements or historical data are considered low-fidelity data, and accurate but limited SCADA measurements are considered high-fidelity data. Through multi-fidelity learning, the proposed DSSE method enables the integration of multi-source data with different accuracy classes.
- 2) The proposed DSSE method demonstrates robustness against bad or missing data by incorporating a Bayesian neural network (BNN) in the high-fidelity layer. This allows the state estimator to output a probability distribution, maintaining high performance even in the presence of data quality issues.
- 3) An interpretability framework based on SHAP is introduced to quantify the impact of each measurement on the estimation accuracy of individual nodes. This improves the transparency of the data-driven DSSE, making it more interpretable for practical applications.

The rest of this paper is organized as follows: Section II describes the DSSE problem statement. Section III illustrates the proposed method. Section IV evaluates the performance of the proposed method by comparative tests, and Section V concludes the paper.

II. PROBLEM STATEMENT

For a distribution system with M measurements and N state variables, the relationship between state variables and measurements can be formulated as:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}, \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ is the state variable vector; $\mathbf{z}$ is the measurement vector in vector form $[z_1, z_2, \dots, z_M]^T$; $\mathbf{h}(\cdot)$ denotes the measurement function; and $\mathbf{e}$ represents the measurement error vector, which is typically assumed to follow a zero-mean Gaussian distribution, i.e., $\mathbf{e} \sim \mathcal{N}(0, \mathbf{R})$,

with the covariance matrix $\mathbf{R} = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_j^2, \dots, \sigma_m^2]$, and σ_j^2 is the variance of the j -th measurement error.

A. Conventional DSSE Method

The WLS method formulates DSSE as an optimization problem by minimizing the weighted sum of squared measurement residuals:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ [\mathbf{z} - \mathbf{h}(\mathbf{x})]^T \mathbf{W} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \right\}, \quad (2)$$

where the weight matrix of measurements $\mathbf{W}$ can be framed as $\mathbf{W} = \mathbf{R}^{-1}$. Traditionally, the Gauss-Newton method is employed to iteratively solve the WLS problem. However, when measurement data are tampered due to network intrusions or cyberattacks, bad data detection and correction algorithms often fail to identify compromised data, yielding significantly degraded estimation performance [45]. Furthermore, bad data can exacerbate the non-convexity of the objective function, potentially causing the Gauss-Newton method to converge to local extrema instead of the global optimum [1], [3].

B. Data-driven DSSE Method

Data-driven methods are used to directly approximate the relationship between measurements and system states, as represented by the following equation:

$$\hat{\mathbf{x}} = \mathcal{F}(\mathbf{z}), \quad (3)$$

where $\mathcal{F}$ is the mapping function. Similar to conventional DSSE, data-driven methods is dependent on the precision of the input data. However, in distribution systems, the widespread insufficiency of measurement devices often leads to reduced measurement redundancy, adversely affecting system observability and accuracy of the data driven-based DSSE methods. The presence of bad data can lead to significant estimation errors in traditional data-driven DSSE methods. Moreover, data-driven algorithms are often regarded as “black boxes” and are not fully understood in most cases. If users perceive a model as a “black box”, they may not trust its predictions and thus be reluctant to use it [40]. This is also a significant challenge faced by data-driven DSSE methods.

III. PROPOSED MULTI-FIDELITY NEURAL NETWORK ENABLED PROBABILISTIC DSSE

This paper proposes an interpretable and robust multi-fidelity learning DSSE framework that effectively integrates heterogeneous measurement sources and enhances estimation robustness under data quality issues. The proposed method consists of two main components: i) a Multi-Fidelity Bayesian Neural Network (MFBNN)-based DSSE model that fuses low-fidelity pseudo-measurements with high-fidelity SCADA/AMI data and leverages Bayesian inference to enhance robustness against bad or missing data, and ii) a SHAP-based explainability module that quantifies both global and local feature impacts, thereby enhancing the transparency and interpretability of the DSSE process. The overall structure of the proposed method is illustrated in Fig. 1. This unified framework is specifically designed to address three key challenges in modern

distribution systems: multi-source data fusion, robustness to measurement anomalies, and the lack of interpretability in deep learning models.

A. Multi-fidelity Bayesian Neural Network for DSSE

The proposed method consists of three key components: (i) a DNN designed to quickly capture coarse features of DSSE for each node, providing a foundation for fine-tuning with high-fidelity measurements; (ii) a BNN tailored to handle measurement uncertainties and integrate high-fidelity measurement data, ensuring robust and accurate state estimates; (iii) a multi-fidelity framework that can adaptively learn both linear and complex nonlinear correlations between low-fidelity and high-fidelity data.

1) *DNN for Low-Fidelity Model*: The performance of NNs is heavily dependent on the quality of training dataset. However, in real-world distribution systems, only a limited number of nodes are equipped with measurement devices capable of providing real-time data with good accuracy. By contrast, pseudo-measurements, usually predicted from historical data, are widely used to improve the system observability [17]. But, they generally lack accuracy.

In the proposed DSSE framework, the low-fidelity layer functions as a cost-effective approximation model, offering preliminary predictions to the high-fidelity layer. The abundant pseudo-measurements serve as input for the low-fidelity layer, enhancing the overall system observability by compensating for the limited availability of high-fidelity data.

DNNs, as a widely adopted model, are particularly well-suited to serve as an approximate model for the low-fidelity layer due to their ability to provide reliable predictions even from large and diverse datasets. Given the training dataset $\mathcal{D} = \{\mathbf{z}, \mathbf{x}\}$, where the input $\mathbf{z}$ has dimension $n_{\text{in}} \in \mathbb{N}$ and the output $\mathbf{x}$ has dimension $n_{\text{out}} \in \mathbb{N}$, a DNN can be designed with J hidden layers. This network architecture consists of interconnected neurons, each characterized by specific weights and activation functions, allowing for complex computational processing. Such networks are designed to approximate complex non-linear functions mapping state variables and measurements in DSSE. The architecture of a DNN consists of multiple layers of neurons, where each layer applies a non-linear transformation to its input. Mathematically, this can be expressed as:

$$\mathbf{H}^0 = \mathbf{z}, \quad (4)$$

$$\mathbf{H}^\ell = f_\ell(\mathbf{w}^\ell \mathbf{H}^{\ell-1} + \mathbf{b}^\ell), \quad \ell = 1, \dots, J, \quad (5)$$

$$\mathbf{x} = \mathbf{H}^J, \quad (6)$$

where $\mathbf{w}^\ell$ and $\mathbf{b}^\ell$ are weights and bias parameters of the ℓ -th layer, respectively; f_ℓ is the activation function in each layer and the *ReLU* function is used in this paper; and $\mathbf{H}^\ell$ is the output of the ℓ -th layer.

In this study, the Mean Squared Error (MSE) and the Maximum a Posteriori Probability (MAP) estimation are chosen as the loss function to evaluate prediction accuracy. The loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = \frac{1}{t} \sum_{i=1}^t \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 + \lambda \sum_{l=1}^J (\boldsymbol{\theta}^\ell)^2, \quad (7)$$

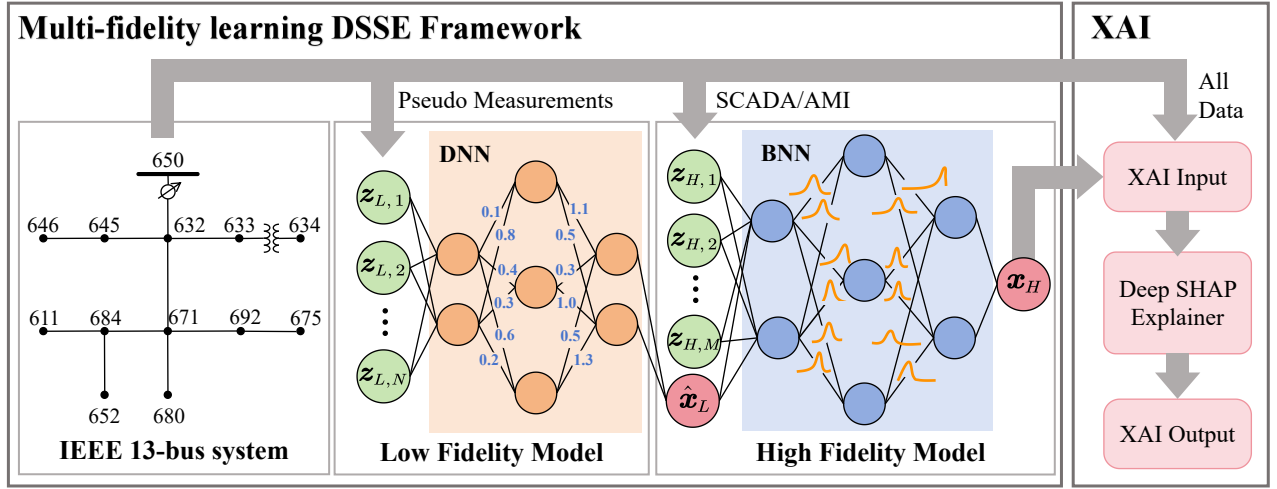


Fig. 1. Proposed multi-fidelity DSSE framework.

where x_i denotes the label values for the i -th sample, $\hat{x}_i$ represents the predicted output, θ^ℓ represent DNN's parameters weights w^ℓ and biases b^ℓ , and t is the total number of samples. This approach ensures that the parameters are optimized to minimize the error between predictions and ground truth. The parameters of the DNN model are updated iteratively via:

$$\theta \leftarrow \theta + \eta \nabla_{\theta} \mathcal{L}(\theta), \quad (8)$$

where η represent the learning rate for the DNN. The implementation of DNN training is shown in Algorithm 1.

Algorithm 1 Low-Fidelity DNN Layer Training

Input: Training dataset $\mathcal{D} = \{z, x\}$, learning rate η , number of epochs E , minibatch size B

Output: Learned parameters $\theta = \{(w, b)\}$

- 1: Initialization: Initialize weights w^ℓ and biases b^ℓ for each layer ℓ
- 2: **for** $epoch = 1 : E$ **do**
- 3: Shuffle $\mathcal{D}$ into minibatches $\{\mathcal{B}_1, \mathcal{B}_2, \dots\}$ of size B
- 4: **for** each minibatch $\mathcal{B}$ **do**
- 5: Forward propagation according to (4)–(6)
- 6: Compute the loss according to (7)
- 7: Update the parameters θ according to (8)
- 8: **end for**
- 9: **end for**

2) *BNN for High-Fidelity Model*: In the proposed DSSE framework, the low-fidelity layer uses pseudo-measurements as input, which may have a systematic bias due to low accuracy. By contrast, real-time measurements from SCADA and AMI are limited but with good accuracy. These high-fidelity measurements are used as input to the high-fidelity layer, allowing the framework to correct the biases introduced by the low-fidelity data. However, given that SCADA and AMI measurements may contain bad or missing data due to communication network failures, congestion, or cyber attacks, the performance of DNN-based DSSE can be significantly degraded. Considering that BNNs can effectively quantify measurement uncertainties, using BNNs as the approximation

model for the high-fidelity layer significantly enhances the robustness of DSSE in the presence of bad data.

A BNN can be defined as any DNN trained using Bayesian inference [46]. The training of BNN similarly follows in (4)–(6). Unlike the approach used for DNNs in (5), where weights w and biases b are treated as deterministic values optimized to minimize the loss function, BNNs model these parameters as probability distributions, thereby capturing uncertainty in their values. This is achieved by assigning stochastic weights to the network (shown in the blue BNN section of Fig. 1), enabling it to represent multiple possible models θ with their associated probability distribution $p(\theta)$.

A prior reflects the belief about the parameters before observing any data. It can be updated based on the observed data. The posterior weights $p(\theta | x, z)$ are obtained using Bayesian inference:

$$p(\theta | x, z) = \frac{p(x | \theta, z) p(\theta | z)}{p(x | z)}, \quad (9)$$

where $p(x | \theta, z)$ is the likelihood, $p(\theta | z)$ is the prior, and $p(x | z)$ is the marginal likelihood, which represents the probability of observing the output x given the input z , marginalized over the entire distribution of the weights:

$$p(x | z) = \int p(x | z, \theta) p(\theta) d\theta. \quad (10)$$

To reduce computational complexity while maintaining the core principles of Bayesian inference, the mean-field variational inference (VI) method is utilized. The goal of variational inference is to approximate the complex probability distribution $p(x | \theta, z) = p(\mathcal{D} | \theta)$ with a simpler distribution $q_\phi(\theta)$, parameterized by $\phi = (\mu, \rho)$. As proposed in [47], the parameters ϕ are optimized by minimizing the Kullback-Leibler (KL) divergence between the approximating distribution, i.e., $q_\phi(\theta)$, and the true posterior, i.e., $p(\theta | x, z)$. Minimizing the KL divergence is equivalent to maximizing the Evidence

Lower Bound (ELBO). Therefore, the loss function is defined as the negative ELBO:

$$\begin{aligned}\mathcal{L}(\phi) &= -\text{ELBO} = \text{KL}[q_\phi(\theta) \| p(\theta)] - \mathbb{E}_{q_\phi(\theta)}[\log p(\mathcal{D} | \theta)] \\ &\approx \frac{1}{K} \sum_{i=1}^K \left[\log q_\phi(\theta^{(i)}) - \log p(\theta^{(i)}) - \log p(\mathcal{D} | \theta^{(i)}) \right],\end{aligned}\quad (11)$$

where K represents the number of samples. This formulation ensures efficient approximation of the posterior distribution by leveraging the simpler variational distribution. From a Bayesian perspective for variational layer, a prior is assigned to the weights as follows:

$$\theta \sim \mathcal{N}(\mu, \sigma) = \mu + \sigma \circ \epsilon, \quad (12)$$

where the prior is commonly modeled as a Gaussian distribution with a zero mean $\mu = \mathbf{0}$ and a diagonal covariance matrix defined by σ . The term $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ represents a random variable drawn from a Gaussian distribution, and $\circ$ denotes element-wise multiplication. To ensure that σ remains non-negative, which is a requirement for a valid standard deviation, it is parameterized using a softplus transformation as:

$$\sigma = \log(1 + \exp(\rho)), \quad (13)$$

where ρ is an unconstrained parameter. Let $f(\theta, \phi) = \log q_\phi(\theta) - \log p(\theta) - \log p(\mathcal{D} | \theta)$. The optimization of parameters in variational inference can be expressed as:

$$\Delta_\mu = \frac{\partial f(\theta, \phi)}{\partial \theta} + \frac{\partial f(\theta, \phi)}{\partial \mu}, \quad (14)$$

$$\Delta_\rho = \frac{\partial f(\theta, \phi)}{\partial \theta} \frac{\epsilon}{1 + \exp(-\rho)} + \frac{\partial f(\theta, \phi)}{\partial \rho}, \quad (15)$$

$$\mu \leftarrow \mu - \alpha \Delta_\mu, \quad (16)$$

$$\rho \leftarrow \rho - \alpha \Delta_\rho, \quad (17)$$

where α represents the learning rate for the BNN. The implementation of BNN training is shown in Algorithm 2.

3) *Multi-Fidelity Framework for Robust DSSE*: As shown in Fig. 1, the MFBNN comprises two different neural networks: the first is a DNN used to approximate low-fidelity data, and the second is a BNN [49] designed to learn the correlation between the low- and high-fidelity data and to handle bad data [50].

For a multi-fidelity model with T fidelity levels, represented by approximate models $\{\mathcal{F}^1, \dots, \mathcal{F}^T\}$, a generalized autoregressive scheme is employed to capture nonlinear correlations in multi-fidelity learning [51]. This approach is formulated as:

$$\mathbf{x}^n = \mathcal{F}^n(\mathbf{z}^n, \hat{\mathbf{x}}^{n-1}), \quad n = 2, \dots, T \quad (18)$$

Thus, a multi-fidelity learning structure is constructed as:

$$\mathbf{x} = \mathcal{F}(\mathbf{z}) = \mathcal{F}^T \circ \mathcal{F}^{T-1} \circ \dots \circ \mathcal{F}^1. \quad (19)$$

In practical distribution systems, sensors are always limited, and their data often comes from various sources such as SCADA, AMI, and pseudo-measurements. These data sources differ in sampling frequency and measurement accuracy. Multi-fidelity models are particularly well-suited for such

Algorithm 2 High-Fidelity BNN Layer Training via Variational Inference [48]

Input: Training dataset $\mathcal{D} = \{\mathbf{z}, \mathbf{x}\}$, learning rate α , number of epochs E , minibatch size B , number of samples K

Output: Learned variational parameters $\phi = \{\mu, \rho\}$

```

1: Initialization: Initialize  $\mu$  and  $\rho$ 
2: for  $epoch = 1 : E$  do
3:   Shuffle  $\mathcal{D}$  into minibatches  $\{\mathcal{B}_1, \mathcal{B}_2, \dots\}$  of size  $B$ .
4:   for each minibatch  $\mathcal{B}$  do
5:      $\mathcal{L}(\phi) \leftarrow 0$ 
6:     for  $s = 1$  to  $K$  do
7:       Sample noise  $\epsilon^{(s)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
8:       Compute standard deviation  $\sigma$  according to (13)
9:       Reparameterize  $\theta^{(s)}$  according to (12)
10:      Compute the loss for this sample according to (11)
11:    end for
12:     $\mathcal{L}(\phi) \leftarrow \frac{1}{K} \mathcal{L}(\phi)$ 
13:    Update parameters  $\phi$  using gradient descent according to (16)–(17)
14:  end for
15: end for
```

nonlinear systems with multiple data sources. Specifically, low-fidelity data are usually unlimited but noisy, and thus are typically used to capture the overall behavior and trends of the distribution system. In contrast, high-fidelity data are limited but accurate, and are therefore used to fine-tune the DSSE results more precisely. The autoregressive NN model can recursively adapt and correct low-fidelity data and update the prediction model. This paper involves multi-fidelity modeling with two fidelity models, the historical pseudo-measurement data are considered low-fidelity data, while SCADA/AMI data are considered high-fidelity data. The input-output relationships for both the low-fidelity neural network $\mathcal{F}_L$ and high-fidelity neural network $\mathcal{F}_H$ can be expressed as:

$$\mathcal{F}_L : \mathbf{z}_L \rightarrow \mathbf{x}_L, \quad \mathcal{F}_H : \{\mathbf{z}_H, \hat{\mathbf{x}}_L\} \rightarrow \mathbf{x}_H, \quad (20)$$

where $\mathbf{z}_L$ represents the low-fidelity input, consisting of pseudo-measurements, and $\mathbf{z}_H$ represents the high-fidelity input, consisting of SCADA/AMI measurements. Similarly, $\mathbf{x}_L$ and $\mathbf{x}_H$ represent the low-fidelity and high-fidelity outputs, respectively. The implementation of the proposed DSSE method is shown in Algorithm 3.

B. Data-Driven DSSE Model Interpretation

Interpretability implies that a specific task can be summarized clearly by an algorithm and linked to principles defined within human cognition [39]. For DSSE, it is crucial to identify the influence of each node feature on state estimates. However, data-driven DSSE surrogate models function as “black boxes”, making it challenging to uncover the hidden mechanisms between measurements and the estimated states. XAI techniques provide a method to analyze data-driven DSSE models in a comprehensible manner, ensuring reliable results that can be understood by human operators. The DSSE XAI framework proposed in this paper is based on Deep SHAP.

Algorithm 3 Multi-Fidelity DSSE Inference

Input: Trained DNN parameters $\theta = (\mathbf{W}, \mathbf{b})$, Trained BNN variational parameters $\phi = \{\mu, \rho\}$, Low-fidelity input $\mathbf{z}_L$, High-fidelity input $\mathbf{z}_H$, number of samples C

Output: Estimated high-fidelity output $\hat{\mathbf{x}}_H$

- 1: Step 1: DNN Inference (Low Fidelity)
- 2: Load the parameters of the DNN model
- 3: Perform a forward propagation of the DNN with input $\mathbf{z}_L$ using θ , obtaining $\hat{\mathbf{x}}_L = \mathbf{H}^L$
- 4: Step 2: BNN Inference (High Fidelity)
- 5: Initialize an empty list
- 6: Load the parameters of BNN model
- 7: **for** $s = 1 : C$ **do**
- 8: Perform a forward propagation of the BNN with input $(\mathbf{z}_H, \hat{\mathbf{x}}_L)$ using ϕ , obtaining $\hat{\mathbf{x}}_H^{(s)}$
- 9: Append $\hat{\mathbf{x}}_H^{(s)}$ to Outputs
- 10: **end for**

Generally, Deep SHAP combines Shapley values and Deep Learning Important Features (DeepLIFT) to provide a specialized interpretability tool for attributing feature importance in deep learning models [52]. It is an efficient and reliable method with significant advantages in improving the transparency and interpretability of deep learning models.

The goal of the Shapley value-based explainable framework is to construct a post-hoc explainable model, i.e., $g(\cdot)$ that quantifies the contribution of features to the decision-making process of the original model $\mathcal{F}(\cdot)$ using Shapley values [52]. This framework ensures consistency between the predictions of the interpretable model and those of the original model.

Specifically, a single sample is denoted as $\mathbf{z} = \{z_1, z_2, \dots, z_M\}$, M is the number of input features. $\mathcal{F}(\mathbf{z})$ is the predicted value of the original black box model, and $g(\mathbf{z})$ is the predicted value of the post-hoc explainable model:

$$g(\mathbf{z}) = \phi_0 + \sum_{i=1}^M \phi_i z_i, \quad (19)$$

where ϕ_i represents the Shapley value of the i -th feature, and ϕ_0 is the expected prediction. The Shapley value quantifies the contribution of a feature to the prediction by calculating a weighted sum over all possible combinations of feature values.

$$\phi_i = \sum_S \frac{(N - |S| - 1)! |S|!}{N!} [\mathcal{F}(S \cup \{z_i\}) - \mathcal{F}(S)], \quad (20)$$

where the subset $S \subseteq \{1, 2, \dots, N\} \setminus \{i\}$ represents a collection of features utilized within the model, $\mathcal{F}(S \cup \{z_i\})$ is the outcome from S without feature z_i , and $\mathcal{F}(S)$ is the model outcome from S .

When the number of features is large, the number of possible coalitions grows exponentially. Therefore, the Shapley framework incorporates the DeepLIFT method, which relies on a backpropagation strategy to improve computational efficiency. The core idea behind DeepLIFT involves computing each neuron's contribution by propagating information backward from the output layer to the input layer [53].

Formally, the DeepLIFT multiplier $m_{\Delta \mathbf{z} \Delta \mathbf{x}}$ is defined as the ratio of the contribution $C_{\Delta \mathbf{z} \Delta \mathbf{x}}$ to the difference $\Delta \mathbf{z}$:

$$m_{\Delta \mathbf{z} \Delta \mathbf{x}} = \frac{C_{\Delta \mathbf{z} \Delta \mathbf{x}}}{\Delta \mathbf{z}}, \quad (21)$$

where $\Delta \mathbf{x} = \mathcal{F}(\mathbf{z}) - \mathcal{F}(\mathbf{r})$ and $\Delta \mathbf{z} = \mathbf{z} - \mathbf{r}$, $\mathbf{r}$ is the reference input, which usually represents a neutral or baseline input such as the mean or median of the training data. If we let $\phi_i = C_{\Delta \mathbf{z} \Delta \mathbf{x}}$, set $\phi_0 = \mathcal{F}(\mathbf{r})$, and choose $\mathbf{r} = E[\mathbf{z}]$, then the resulting explanation model matches Eq. (19).

Deep SHAP uses the backward propagation approach of DeepLIFT to efficiently compute approximate Shapley values for a neural network. By integrating these approximations over background samples, the entire model obtains approximate Shapley values with

$$\sum \phi_i = \mathcal{F}(\mathbf{z}) - \mathcal{F}(\mathbf{r}). \quad (22)$$

Deep SHAP assigns each input feature a numerical value representing its marginal contribution to the prediction, enabling fine-grained sensitivity analysis while preserving both global interpretability (feature importance across the dataset) and local interpretability (feature impact for individual instances). Hence, in the multi-fidelity BNN context, Deep SHAP helps system operators diagnose the significance of each measurement source and understand how the model refines low-fidelity pseudo-measurements with more accurate high-fidelity data. By comparing aggregated SHAP values across different operating conditions, operators can also detect measurement anomalies or data shifts, thereby enhancing the reliability of DSSE decisions. These properties ensure that deep learning models can be analyzed as "open black boxes," allowing the statistical and individual contributions of each input feature to the model output to be quantified. By quantifying the marginal contribution of each measurement to the DSSE output, Deep SHAP reveals how the MFBNN model derives its estimates, thereby enhancing transparency and clarifying which measurements exert the greatest influence on each node's state.

IV. SIMULATION RESULTS

In this section, comparative tests are conducted on the IEEE 13-bus system and on a realistic 2135-node distribution system in the Dominion Energy service territory. First, the simulation methodology and evaluation metrics are introduced. Next, the accuracy and robustness of the proposed DSSE method are evaluated. Then, the interpretability of the proposed DSSE method is analyzed. Finally, the computational efficiency of each method is presented.

A. Simulation Setup

1) *IEEE 13-bus System Setting*: For the IEEE 13-bus system, its topology is shown in Fig. 1. Table I provides the detailed configuration of all measurement devices, including the types of devices, noise levels, measurement categories, and installation locations. Particularly, different measurement devices have different noise levels, the measurement noise is assumed to follow a zero-mean Gaussian distribution, i.e.,

TABLE I
MEASUREMENT SETTING OF IEEE 13-BUS SYSTEM

Measurement Device	Noise Level	Measurement Type	Location	Sampling Frequency
SCADA	1%	Voltage Magnitude, Active Power, Reactive Power	650, 692	2 s to 15 s
AMI	5%	Active Power, Reactive Power	670, 671, 646, 675	5 to 15 mins
Pseudo Measurement	30%	Voltage Magnitude, Active Power, Reactive Power	634, 611, 652	10 to 30 mins
Zero Injection	0%	Active Power, Reactive Power	RG60, 633, 632, 680, 645, 684	None

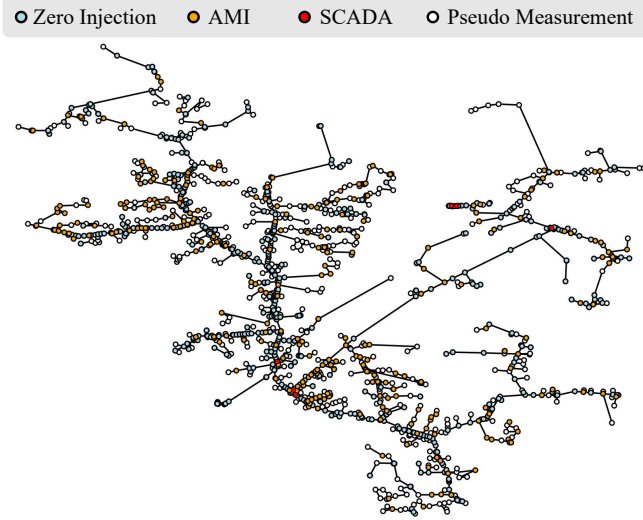


Fig. 2. The single line diagram of the Dominion 2135-node system.

$e_i \sim \mathcal{N}(0, \sigma_i)$. The relationship between the precision level and the noise standard deviation σ_i is $\sigma_i = |z_i| pr_i / 3$, where pr_i is the precision of the i -th measurement device. To better reflect real-world conditions in our dataset, we added noise not only to the input measurements but also to the output states: the low-fidelity outputs are assigned a noise level of 0.1%, while the high-fidelity outputs have a noise level of 0.01%. The dataset consists of samples collected over one year, totaling 8760 samples. Among them, 80% are used as the training set, while 20% are used as the testing set.

2) *Dominion 2135-node Distribution System Setting*: For the realistic Dominion 2135-node distribution system, the method is implemented using OpenDSS, where the original Synergi model has been converted to OpenDSS. The system topology, shown in Fig. 2, represents a three-phase unbalanced feeder with 2135 nodes, 1469 branches, 2 front-of-meter photovoltaic systems, 3 capacitors, 10 step-down transformers, and 6 reclosers. The dataset is derived from one year of data from the Dominion 2135-node system. Fig. 2 also illustrates the types of measurement devices at each node. The specific noise settings for different devices are the same as those in the IEEE 13-bus system configuration. The dataset is configured similarly to the IEEE 13-bus system, consisting of 8760 samples. Among them, 80% are allocated as the training set, and 20% are allocated as the testing set.

3) *Benchmarks and Evaluation Metrics*: To evaluate the superiority of the developed MFBNN method, the conventional WLS, DNN, BNN, and MFNN methodologies are chosen as benchmarks. The Mean Absolute Error (MAE) is used to

quantify the absolute difference between the actual values of state variables and their estimates. It is defined as:

$$\text{MAE} = \frac{1}{t_{\text{test}}} \sum_{i=1}^{t_{\text{test}}} |x_i - \hat{x}_i| \quad (23)$$

where x_i and $\hat{x}_i$ denote the actual value and the estimated value of the state variable for the i -th sample, respectively; t_{test} represents the total number of samples in the test set.

In addition to MAE, the maximum (MAX) deviation between the actual value and the estimated value is employed as an additional metric to evaluate the accuracy of point estimations for different DSSE methods.

4) *DSSE Algorithms Setting*: The DNN model consists of three dense layers with 32, 64, and 128 neurons, respectively. The model is trained using the Adam optimizer with a learning rate of 0.001. The BNN shares the same hidden layer parameters and optimizer as the DNN. However, unlike the DNN, the BNN includes two additional parameters: prior mean $\mu = 0$ and the initial posterior $\rho = -2$. The MFNN model comprises a backbone network, a low-fidelity network, a high-fidelity network, and a shortcut layer. The low-fidelity network includes two dense layers with 32 and 64 neurons; while the high-fidelity network consists of three dense layers with 32, 64, and 128 neurons. Training is conducted using the Adam optimizer with a learning rate of 0.001. The proposed method maintains the same low-fidelity hidden layer parameters and optimizer settings as the MFNN. For the high-fidelity layer, the Adam optimizer is used with the learning rate set to 0.001. Unlike the MFNN, which employs DNN, the DNN is replaced with an equivalent BNN.

5) *Bad Data Setting*: To comprehensively evaluate the robustness of the proposed method, measurement uncertainties are emulated, which consists of two main components: bad data and missing data. Specifically, bad data refers to data points that deviate from the actual value by 10σ to 15σ and accounts for 80% of measurement uncertainty points. Missing data refers to absent data points, represented as zero, and constitutes 20% of measurement uncertainty points. In each sample, bad data is uniformly and randomly distributed across all nodes, with the proportion of bad data ranging from 3% to 4% of the total number of nodes per sample. Additionally, for each node in different samples, a certain number of bad data points are introduced. We ensure that each node contains at least 5 bad data points.

B. Effectiveness on IEEE 13-Bus System

To validate the accuracy and robustness of the proposed MFBNN method, three different scenarios are designed. In

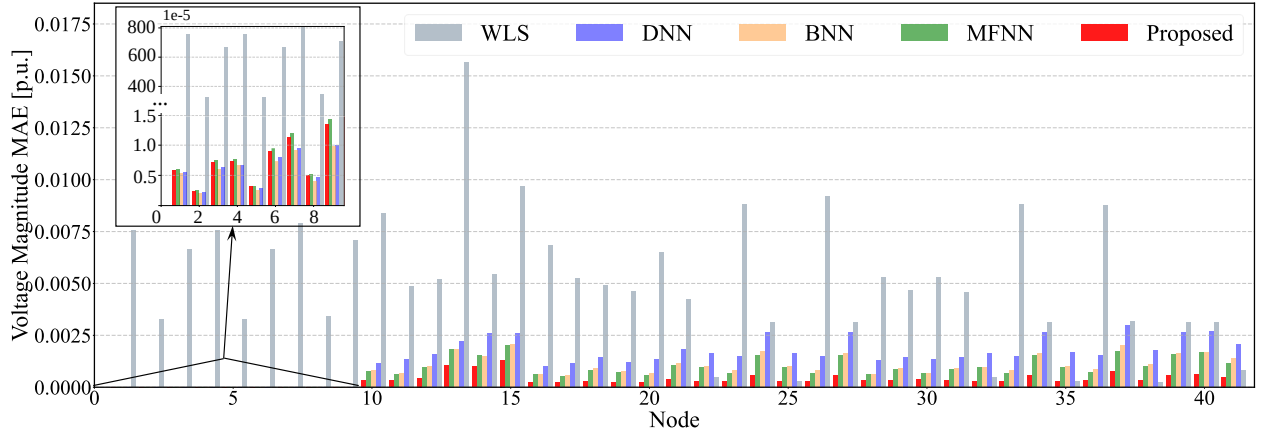


Fig. 3. MAEs of voltage magnitude estimates of IEEE 13-bus system under bad data situation in Scenario 1.

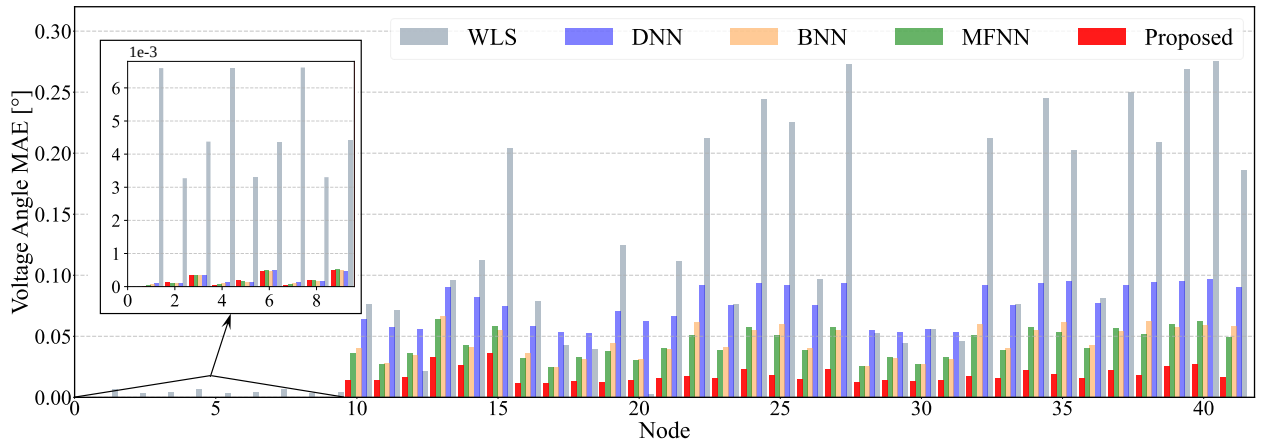


Fig. 4. MAEs of voltage phase angle estimates of IEEE 13-bus system under bad data situation in Scenario 1.

Scenario 1, DSSE tests are conducted under a normal measurement environment and a situation containing bad data to evaluate the feasibility and robustness of the proposed method. In Scenario 2, different noise levels are applied to AMI measurements to investigate the impact of varying measurement quality on DSSE performance in datasets containing bad data. In Scenario 3, different PV penetrations are applied to investigate the impact of varying PV penetration on DSSE performance in datasets that contain bad data.

1) Scenario 1: MAEs of voltage magnitude and phase angle estimates of different methods under the bad data situation are illustrated in Figs. 3 and 4, respectively. It is worth noting that both Figs. 3 and 4 present the MAE of state estimates for each individual node, which is calculated as the average over 1752 samples from the test dataset. In Fig. 3, the conventional WLS approach exhibits significantly higher error levels, particularly at certain nodes where it exceeds 0.01 p.u.. The DSSE accuracy of MFNN and BNN methods is roughly comparable, and both outperform the DNN method. This indicates that both multi-fidelity approaches and Bayesian inference provide a certain level of robustness against bad data. By contrast, the proposed MFBNN method combines the advantages of BNN and multi-fidelity approaches, thereby achieving the lowest MAE across all nodes. Similarly, Fig. 4

shows that while the WLS method maintains MAEs comparable to the DNN method at nodes equipped with measurement devices, it exhibits significant discrepancies between estimated and actual values at nodes lacking measurement devices. By contrast, the proposed method maintains consistently small MAEs across all nodes, demonstrating its robustness and accuracy in the presence of bad data. Moreover, the results for the first nine nodes differ significantly from those of the remaining nodes. For data-driven DSSE methods, the first nine nodes, corresponding to the Sourcebus, Bus 650, and RG60, exhibit smaller MAEs because these nodes belong to either the slack bus or are directly connected to the slack bus with no load. This leads to minimal variations for state variables, resulting in consistently low MAE values. By contrast, the WLS method, due to its iterative nature and sensitivity to bad data, produces relatively higher MAEs at these nodes.

Tables II and III present a comparative analysis of various DSSE methods on the IEEE 13-bus system in terms of voltage magnitude and phase angle estimates, respectively, both with and without bad data. In all situations, the proposed MFBNN approach consistently achieves the lowest MAE and MAX, outperforming WLS, DNN, BNN, and MFNN methods.

In the *without bad data* situation, the proposed method delivers significant improvements in both voltage magnitude

TABLE II
MAE AND MAX OF VOLTAGE MAGNITUDE ESTIMATES OF IEEE 13-BUS SYSTEM (UNITS: $[10^{-3} p.u.]$)

Situation	Phase	WLS		DNN		BNN		MFNN		Proposed	
		MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX
Without Bad Data	A	1.9097	4.3484	0.6034	1.4676	0.4148	1.0297	0.2941	0.5688	0.1206	0.3510
	B	1.5486	3.6326	0.5704	1.4042	0.3148	0.8517	0.2069	0.5047	0.0905	0.2486
	C	1.2244	3.7678	1.0997	2.0977	0.6017	1.2431	0.4252	0.7796	0.1505	0.2822
	Total	1.5444	4.3484	0.7745	2.0977	0.4515	1.2431	0.3144	0.7796	0.1220	0.3510
With Bad Data	A	6.0400	10.8128	0.8894	1.9066	0.8087	1.8277	0.7657	1.8475	0.3006	1.0583
	B	4.5298	6.3985	0.8718	1.9132	0.6608	1.4921	0.5979	1.5377	0.2831	0.9899
	C	4.8812	6.7547	1.4357	2.4503	1.1225	2.0523	1.0515	2.0037	0.4604	1.3218
	Total	5.1372	10.8128	1.0837	2.4503	0.8766	2.0523	0.8171	2.0037	0.3535	1.3218

TABLE III
MAE AND MAX OF VOLTAGE PHASE ANGLE ESTIMATES OF IEEE 13-BUS SYSTEM (UNITS: $[^\circ]$)

Situation	Phase	WLS		DNN		BNN		MFNN		Proposed	
		MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX
Without Bad Data	A	0.0303	0.0854	0.0388	0.0604	0.0216	0.0352	0.0104	0.0174	0.0055	0.0117
	B	0.0596	0.1609	0.0238	0.0376	0.0124	0.0216	0.0111	0.0178	0.0050	0.0087
	C	0.0190	0.0331	0.0337	0.0554	0.0178	0.0301	0.0156	0.0268	0.0064	0.0104
	Total	0.0355	0.1609	0.0322	0.0604	0.0173	0.0352	0.0125	0.0268	0.0056	0.0117
With Bad Data	A	0.0769	0.1466	0.0686	0.1022	0.0425	0.0662	0.0367	0.0643	0.0137	0.0329
	B	0.0849	0.2558	0.0372	0.0572	0.0272	0.0428	0.0267	0.0424	0.0120	0.0264
	C	0.1496	0.2736	0.0471	0.0746	0.0367	0.0592	0.0382	0.0625	0.0168	0.0365
	Total	0.1061	0.2736	0.0458	0.0788	0.0355	0.0662	0.0340	0.0643	0.0143	0.0365

and phase angle estimates. For the total MAE of voltage magnitude estimates, it achieves reductions of approximately 92%, 84%, 73%, and 61% compared to WLS, DNN, BNN, and MFNN methods, respectively. Similarly, for the total MAE of voltage phase angle estimates, its accuracy provides relative improvements of 84%, 83%, 68%, and 55%, respectively. In the *with bad data* situation, the proposed approach maintains its superior performance. It achieves average reductions in total MAE of voltage magnitude estimates of approximately 93%, 67%, 60%, and 57% compared to WLS, DNN, BNN, and MFNN methods, respectively. For the total MAE of voltage phase angle estimates, it outperforms these benchmarks by 87%, 69%, 60%, and 58%, respectively. Notably, the proposed method demonstrates robustness under bad data situations. While the WLS method experiences significant accuracy degradation, the proposed approach reduces the total MAE of phase angle estimates to 0.0143° , markedly outperforming both traditional WLS and competing data-driven methods.

In addition to MAEs, Tables II and III also show the MAX error for each method under different situations in Scenario 1. The comparative analysis of maximum error values reveals critical insights into worst-case performance. In both *without bad data* and *with bad data* situations, the proposed DSSE method demonstrates consistent superiority in estimation for both voltage magnitudes and phase angles.

2) *Scenario 2*: Table IV shows how increasing AMI noise levels affect MAEs of voltage magnitude estimates of the IEEE 13-bus system under the bad data situation. As the noise percentage rises from 1% to 9%, the proposed method consistently achieves the lowest MAE compared to WLS, DNN, BNN, and MFNN methods. Table V presents the MAE

TABLE IV
MAEs OF VOLTAGE MAGNITUDE ESTIMATES UNDER DIFFERENT AMI NOISE LEVELS (UNITS: $[10^{-3} p.u.]$)

AMI Noise	WLS	DNN	BNN	MFNN	Proposed
1%	2.2383	1.0155	0.8820	0.8093	0.3094
3%	3.5226	1.0294	0.9397	0.8091	0.3114
5%	5.1372	1.0837	0.8766	0.8171	0.3535
7%	6.7096	1.1226	0.9474	0.8245	0.3759
9%	8.9700	1.1414	0.9530	0.8491	0.4085

TABLE V
MAEs OF VOLTAGE PHASE ANGLE ESTIMATES UNDER DIFFERENT AMI NOISE LEVELS (UNITS: $[^\circ]$)

AMI Noise	WLS	DNN	BNN	MFNN	Proposed
1%	0.0597	0.0457	0.0357	0.0338	0.0131
3%	0.0811	0.0429	0.0388	0.0339	0.0133
5%	0.1061	0.0458	0.0355	0.0341	0.0143
7%	0.1327	0.0474	0.0385	0.0343	0.0151
9%	0.1691	0.0477	0.0390	0.0352	0.0161

of voltage phase angle estimates of the IEEE 13-bus system under the bad data situation. Notably, the proposed method maintains high accuracy even in the presence of bad data, achieving significantly lower MAEs for the total voltage phase angle compared to other DSSE methods. This demonstrates that, with sufficient training, the proposed method exhibits superior robustness and effectiveness in handling DSSE under more challenging noise situations.

Specifically, for voltage magnitudes, the proposed method achieves an MAE reduction of 93.38% over WLS, 67.39% over DNN, 61.75% over BNN, and 57.19% over MFNN.

For voltage phase angles, the proposed method achieves an MAE reduction of 86.87% over WLS, 68.63% over DNN, 61.60% over BNN, and 58.02% over MFNN. The proposed method demonstrates superior performance in both voltage magnitude and phase angle estimation under varying levels of AMI noises. This highlights its robustness and effectiveness in addressing DSSE challenges under more stringent noise situations.

3) *Scenario 3*: Table VI shows how increasing PV penetration affects MAEs of voltage magnitude estimates of the IEEE 13-bus system under the bad data situation. As PV penetration increases from 10% to 80%, the conventional WLS method exhibits a significant MAE rise, while DNN, BNN, and MFNN show moderate increases. In contrast, the proposed method demonstrates remarkable stability, with MAE fluctuating minimally between $0.3653 \times 10^{-3} p.u.$ and $0.3925 \times 10^{-3} p.u.$. Table VII presents the MAE of voltage phase angle estimates of the IEEE 13-bus system under the bad data situation. Table VII reveals similar trends for voltage phase angle estimation: WLS degrades sharply, while DNN, BNN, and MFNN exhibit minor variations, and the proposed method maintains near-constant precision. This divergence highlights the proposed method's unique capability to mitigate performance degradation caused by increasing renewable intermittency.

Specifically, for voltage magnitude estimation at 80% PV penetration, the proposed method achieves MAE reductions of 95.27% compared to WLS, 66.31% over DNN, 60.39% over BNN, and 54.25% over MFNN. For voltage phase angles, the reductions reach 95.59% (WLS), 68.05% (DNN), 61.21% (BNN), and 56.50% (MFNN). These results indicate that the proposed method effectively models power fluctuation characteristics, significantly enhancing estimation robustness as system uncertainty intensifies with increasing PV penetration.

TABLE VI
MAES OF VOLTAGE MAGNITUDE ESTIMATES UNDER DIFFERENT PV PENETRATION (UNITS: $[10^{-3} p.u.]$)

PV Penetration	WLS	DNN	BNN	MFNN	Proposed
10%	5.3954	1.0960	0.9141	0.7990	0.3653
20%	5.4105	1.1061	0.9303	0.8209	0.3735
40%	6.2197	1.1018	0.9614	0.8553	0.3709
80%	8.2930	1.1655	0.9909	0.8584	0.3925

TABLE VII
MAES OF VOLTAGE PHASE ANGLE ESTIMATES UNDER DIFFERENT PV PENETRATION (UNITS: $[^\circ]$)

PV Penetration	WLS	DNN	BNN	MFNN	Proposed
10%	0.1105	0.0457	0.0380	0.0337	0.0146
20%	0.1597	0.0448	0.0369	0.0338	0.0147
40%	0.1882	0.0458	0.0386	0.0353	0.0147
80%	0.3495	0.0482	0.0397	0.0354	0.0154

C. Explaining the Solution from Proposed DSSE Method

Once the proposed DSSE method is trained, its results are interpreted by Deep SHAP. In this section, the Deep SHAP

algorithm provides both local and global explanations for the outputs of the proposed DSSE method.

1) *Global explanations*: The Deep SHAP algorithm offers a comprehensive global explanation for the DSSE method, enabling the quantification and visualization of the influence of all input features on the DSSE outputs. This provides an overarching understanding of the key features that significantly impact the performance of DEES method.

Fig. 5 illustrates the top 15 input features that have the greatest impact on the DSSE outputs. Each input feature z_i represents its average impact on the DSSE output x . A higher value indicates a greater overall influence of the feature on the DSSE output. Different colors correspond to DSSE outputs of different nodes and phases. Each bar consists of multiple colored segments, where each segment represents the SHAP value of the input feature for a specific output. This visualization not only highlights feature importance but also illustrates its contribution to different outputs.

For example, Q675-A represents the active power at node 675 in phase A. The most significant impact of this feature is indicated in red, which primarily corresponds to the voltage magnitude of node 675 in phases A, B, and C. This suggests that the voltage estimate of DSSE is highly influenced by active power measurements. Additionally, the green segment is also prominent, which mainly represents the voltage phase angle of nodes 671 and 670 across all three phases. This indicates that the active power of phase A at node 675 significantly affects the phase angles at adjacent measurement points, such as nodes 670 and 671.

2) *Local explanations*: The Deep SHAP algorithm provides a local explanation for the DSSE method, enabling the quantification and visualization of the impact of all input features on specific DSSE outputs. In Fig. 6, the 7 most influential input features for the voltage estimation at node 670 phase B can be observed. Red represents a positive contribution, and blue represents a negative contribution. Since node 670 is equipped with an AMI measurement device, it provides higher accuracy and greater reliability. As a result, the active and reactive power of phase B at node 670 dominate the estimation. Additionally, neighboring AMI measurement nodes, such as 671 and 675, contribute positively to the DSSE at node 670. By contrast, the pseudo-measurement at node 634 and the zero-injection node 645 have the most significant negative contributions. This observation aligns with the weight assignment principles in WLS.

Similarly, Figs. 7 and 8 present the influence and distribution of the top 9 input features with the highest Shapley values on the DSSE estimates for the voltage magnitude and phase angle at node 634, respectively. The horizontal axis represents the quantitative impact of each feature on the model output. Each point represents an individual data sample. The distribution of points, along with their colors and positions, provides a comprehensive visualization of the inference process of the model. This facilitates a deeper understanding of the mechanism of the DSSE model. For example, as shown in Fig. 7, for the DSSE of the voltage magnitude at pseudo-measurement node 634 phase C, several key features play a crucial role. In addition to the active and reactive power

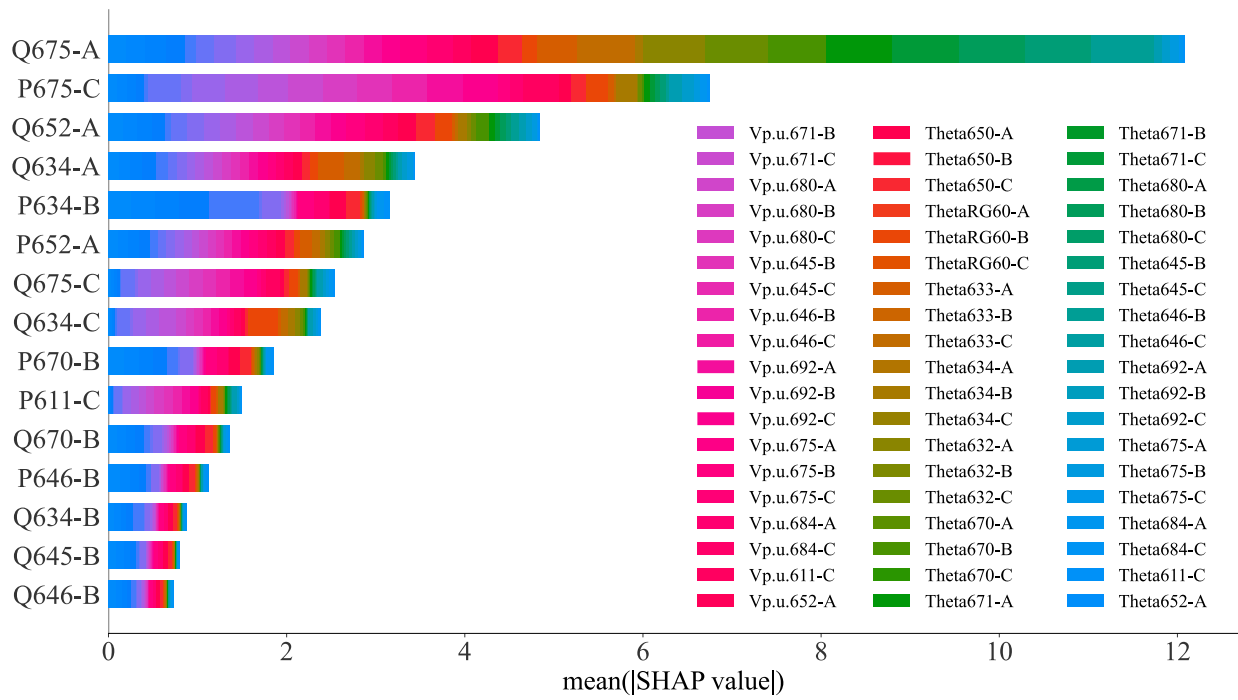


Fig. 5. The top 15 features that greatly influence the proposed DSSE method.

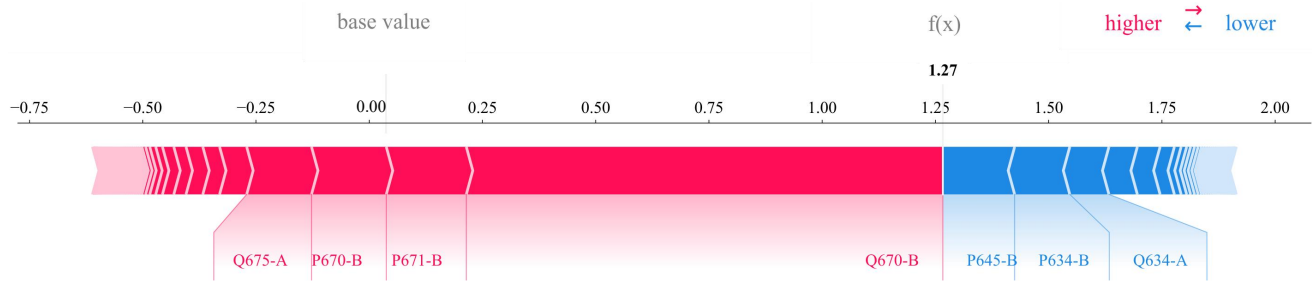


Fig. 6. Influence of features for phase-B voltage estimation at node 670.

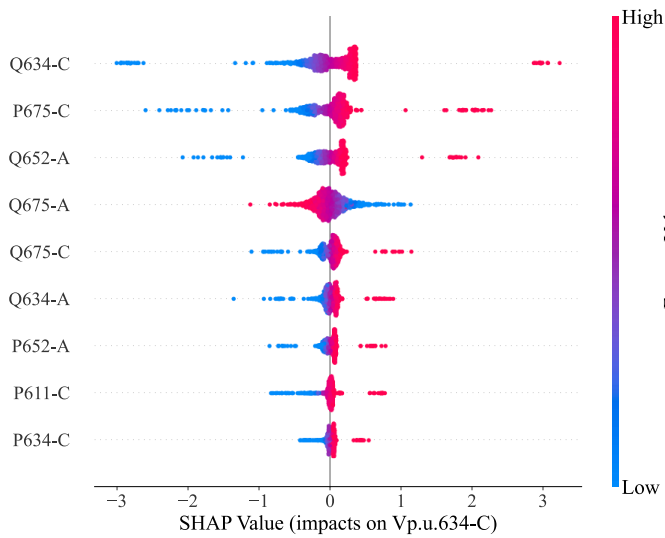


Fig. 7. Bee Swarm plot of SHAP value magnitude for 634-C voltage magnitude estimates.

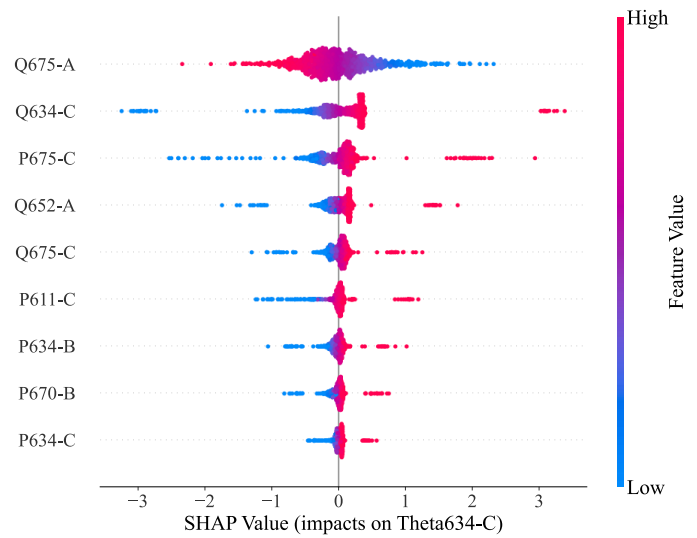


Fig. 8. Bee Swarm plot of SHAP value magnitude for 634-C voltage phase angle estimates.

TABLE VIII
MAE AND MAX OF VOLTAGE MAGNITUDE AND PHASE ANGLE ESTIMATES OF 2135-NODE DISTRIBUTION SYSTEM UNDER BAD DATA SITUATION
(UNITS: $[10^{-3} p.u.]$ OR $[^\circ]$)

	Phase	DNN		BNN		MFNN		Proposed	
		MAE	MAX	MAE	MAX	MAE	MAX	MAE	MAX
Voltage	A	0.64610	3.1991	0.21536	0.98217	0.32406	1.6632	0.082968	0.65029
	B	0.44848	3.1812	0.15945	0.97466	0.23709	1.5750	0.059729	0.36287
	C	0.40720	3.0718	0.14175	0.93390	0.22882	1.4101	0.054234	0.60109
	Total	0.50417	3.1991	0.17365	0.98217	0.26416	1.6632	0.066112	0.65029
Theta	A	0.091753	0.45962	0.029068	0.13769	0.044709	0.21386	0.011072	0.13088
	B	0.066961	0.46275	0.022373	0.13802	0.034836	0.21493	0.008772	0.077384
	C	0.085483	0.46458	0.027259	0.14240	0.042652	0.24793	0.010708	0.083994
	Total	0.080076	0.46458	0.025885	0.14240	0.040183	0.24793	0.010044	0.13088

at node 634 itself, nearby measurement nodes, such as node 675 and 671, also contribute significantly. Furthermore, since DSSE operates on the entire system, terminal nodes such as 611 and 652 also influence the voltage magnitude estimation at 634. Thus, Deep SHAP could provide both global and local views of DSSE for system operators, facilitating the identification of key input features that have the greatest impact on DSSE performance.

D. Scalability to 2135-Node Distribution System

To verify the scalability of the proposed DSSE method, the realistic Dominion 2135-nodes distribution system is used for further validation. Since the bad data ratio is set to about 4% in the 2135-node system, the WLS algorithm fails to converge effectively. Therefore, in the subsequent comparative experiments, only DNN, BNN, and MFNN are included as benchmark methods.

Table VIII presents a comparative analysis of voltage magnitude and phase angle estimation errors of a 2135-node distribution system under a bad data situation. Under the bad data situation, the proposed DSSE approach achieves substantial improvements over DNN, BNN, and MFNN methods in both voltage and phase angle estimates. For voltage magnitude estimates, the proposed method outperforms DNN, BNN, and MFNN methods by approximately 87%, 62%, and 75%, respectively. In terms of phase angle estimates, the proposed approach yields around 87%, 61%, and 75% improvements over these baselines, underscoring its robustness and accuracy even in the presence of corrupted measurement data. Consistent performance advantages are observed across individual phases and aggregate measurements, demonstrating the robustness of the proposed method in handling measurement anomalies and system complexity. Moreover, the magnitude of MAX is often positively correlated with the ability to handle bad data. The proposed method consistently achieves the lowest error in MAX, reflecting an improvement in stability under adverse data situations. A similar pattern is observed in phase angle estimates, where the proposed method demonstrates significant improvement in suppressing peak errors. This highlights its reliability in handling bad data.

Fig. 9 and Fig. 10 present the MAE statistics for all 2135 nodes in histogram form, illustrating the distribution of MAEs for voltage magnitudes and voltage angles across different

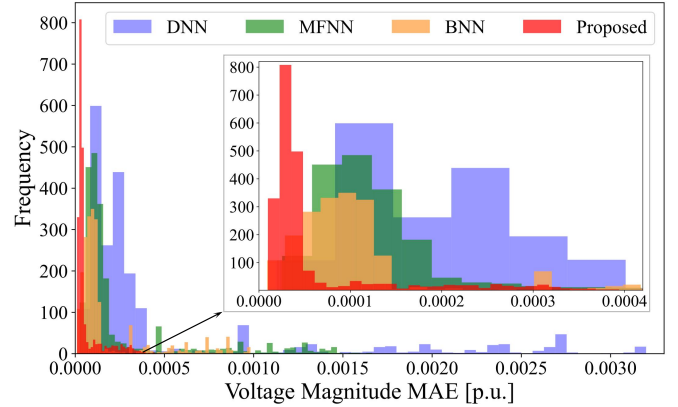


Fig. 9. Histogram of MAEs of voltage magnitude estimates of all nodes.

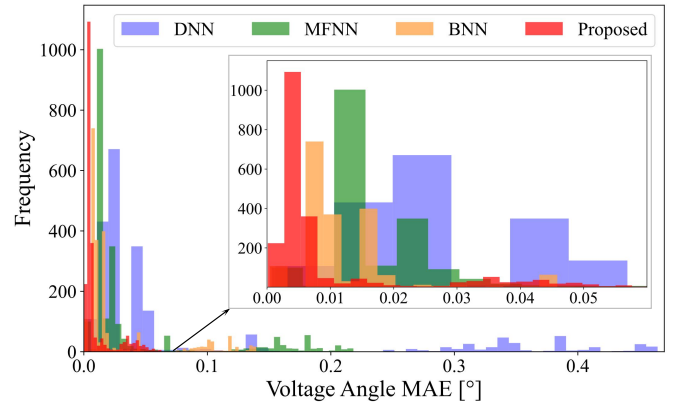


Fig. 10. Histogram of MAEs of voltage phase angle estimates of all nodes.

DSSE methods. The varying bar widths in each algorithm's histogram reflect differences between maximum value and minimum value, leading to inconsistent sampling intervals; thus, narrower bars indicate smaller MAE values. Notably, for both voltage magnitude and phase angle, the statistical results of the proposed MFBNN method are concentrated in the lower MAE range. This indicates that its estimates are more accurate in the presence of bad data. By contrast, other methods exhibit wider tails, especially the DNN method, which reaches a maximum error of $0.0033 p.u./0.44^\circ$. These distributions illustrate that the proposed method not only reduces overall estimation errors but also demonstrates enhanced robustness

against bad data in a realistic large-scale distribution system.

E. Computational Efficiency Analysis

To assess the computational costs of different DSSE methods, the time required for training and testing each method is summarized in Table IX. All computations are performed on a computer with a 3.2-GHz Intel Core i9-12900KF CPU, NVIDIA GeForce RTX 3090 GPU, and 128 GB RAM. Since the WLS algorithm is not a machine learning algorithm, it does not require training time. Although the training time of MFBNN is longer than that of single-fidelity models such as DNN and BNN, this process is conducted offline and does not impact online DSSE. For real-time applications, testing time is more critical. The proposed MFBNN requires only 91.3 *ms* for the IEEE 13-bus system and 97.0 *ms* for the 2135-node system. It is significantly smaller than the fastest sampling rate of SCADA measurements (i.e., 2 seconds) and the model-based WLS DSSE method, indicating that the developed MFBNN model can efficiently support online DSSE in large-scale systems. While the testing time of MFBNN is comparable to other data-driven DSSE methods, it consistently achieves higher accuracy and better robustness against bad data. Considering both performance and efficiency, MFBNN offers a more reliable solution within similar computational costs. Meanwhile, SHAP is employed as a post-hoc tool for interpretability and does not affect real-time performance. In summary, the testing times of the developed MFBNN DSSE on both the IEEE 13-bus and the 2135-node systems are significantly smaller than the sampling rate of measurements, as detailed in Table I, demonstrating its scalability and practicality for implementing online DSSE in large distribution systems.

TABLE IX
TIME REQUIRED FOR TRAINING AND TESTING OF EACH ALGORITHM
(UNITS: [*ms* OR *s*])

Algorithm	IEEE 13 Bus System		2135 Node System	
	Training	Testing	Training	Testing
DNN	59.9 <i>s</i>	62.8 <i>ms</i>	305.1 <i>s</i>	45.7 <i>ms</i>
BNN	66.5 <i>s</i>	62.8 <i>ms</i>	258.4 <i>s</i>	62.8 <i>ms</i>
MFNN	193.8 <i>s</i>	91.3 <i>ms</i>	423.6 <i>s</i>	85.6 <i>ms</i>
MFBNN	188.7 <i>s</i>	91.3 <i>ms</i>	469.1 <i>s</i>	97.0 <i>ms</i>
WLS	None	430.1 <i>ms</i>	None	824241.6 <i>ms</i>

V. CONCLUSION

This paper presents an explainable and robust MFBNN framework for DSSE in three-phase unbalanced distribution networks. The proposed DSSE method effectively integrates low-fidelity pseudo-measurements with high-fidelity SCADA and AMI data to achieve enhanced estimation accuracy and robustness, even under challenging situations such as high noise levels, bad data, and missing data. Comprehensive validation on both the IEEE 13-bus system and a real-world 2135-node Dominion Energy distribution system demonstrates its superiority over conventional WLS and advanced data-driven methods (DNN, BNN, MFNN), with significant reductions

in both mean and maximum estimation errors for voltage magnitudes and phase angles. The integration of SHAP-based explainability quantitatively reveals feature contributions, enabling interpretable decision-making for grid operators. Moreover, the online inference phase remains highly efficient regardless of the power system size, as it only involves a fast feed-forward pass through the trained model. Future work may explore the integration of physical information constraints into this framework to further enhance its transparency and applicability.

REFERENCES

- [1] Y. Raghuvamsi and K. Teeparthi, "A review on distribution system state estimation uncertainty issues using deep learning approaches," *Renewable and Sustainable Energy Reviews*, vol. 187, p. 113752, 2023.
- [2] P. Bhattar, N. Pindoriya, and A. Sharma, "A combined survey on distribution system state estimation and false data injection in cyber-physical power distribution networks," *IET Cyber-Physical Systems: Theory Applications*, vol. 6, 03 2021.
- [3] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, 2018.
- [4] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part i: Exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120–125, 1970.
- [5] F. Schweppe and E. Handschin, "Static state estimation in electric power systems," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 972–982, 1974.
- [6] M. B. Do Coutto Filho and J. C. Stacchini de Souza, "Forecasting-aided state estimation—part i: Panorama," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1667–1677, 2009.
- [7] X. Bian, X. R. Li, H. Chen, D. Gan, and J. Qiu, "Joint estimation of state and parameter with synchrophasors—part i: State tracking," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1196–1208, 2011.
- [8] C. Carquex, C. Rosenberg, and K. Bhattacharya, "State estimation in power distribution systems based on ensemble kalman filtering," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6600–6610, 2018.
- [9] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing observability in distribution grids using smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5953–5961, 2018.
- [10] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, "Effects of measurements and pseudomeasurements correlation in distribution system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2813–2823, 2014.
- [11] S. Prasad and D. V. Kumar, "Trade-offs in pmu and ied deployment for active distribution state estimation using multi-objective evolutionary algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1298–1307, 2018.
- [12] R. Singh, B. Pal, and R. Jabr, "Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement," *IET generation, transmission & distribution*, vol. 4, no. 1, pp. 50–59, 2010.
- [13] Z. Cao, Y. Wang, C.-C. Chu, and R. Gadh, "Robust pseudo-measurement modeling for three-phase distribution systems state estimation," *Electric Power Systems Research*, vol. 180, p. 106138, 2020.
- [14] G. Cheng, Y. Lin, A. Abur, A. Gómez-Expósito, and W. Wu, "A survey of power system state estimation using multiple data sources: Pmus, scada, ami, and beyond," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1129–1151, 2023.
- [15] J. Zhao, G. Zhang, M. La Scala, and Z. Wang, "Enhanced robustness of state estimator to bad data processing through multi-innovation analysis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1610–1619, 2016.
- [16] O. Chilard and S. Grenard, "Detection of measurements errors with a distribution network state estimation function," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*. IET, 2013, pp. 1–4.
- [17] 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, 2016.
- [18] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1193–1205, 2017.

- [19] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 349–358, 1994.
- [20] M. Göl and A. Abur, "Lav based robust state estimation for systems measured by pmus," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1808–1814, 2014.
- [21] H. G. Abood, V. Sreeram, and Y. Mishra, "A new algorithm for improving the numerical stability of power system state estimation," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 14, no. 3, pp. 358–365, 2019.
- [22] 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.
- [23] Q. Qiang, S. Guoqiang, X. Wei, Y. Minghui, W. Zhinong, and Z. Haixiang, "Distribution system state estimation based on pseudo measurement modeling using convolutional neural network," in *2018 China International Conference on Electricity Distribution (CICED)*. IEEE, 2018, pp. 2416–2420.
- [24] G. Hong and Y.-S. Kim, "Supervised learning approach for state estimation of unmeasured points of distribution network," *IEEE Access*, vol. 8, pp. 113 918–113 931, 2020.
- [25] A. Al-Wakeel, J. Wu, and N. Jenkins, "K-means based load estimation of domestic smart meter measurements," *Applied energy*, vol. 194, pp. 333–342, 2017.
- [26] C. Ren and Y. Xu, "A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5044–5052, 2019.
- [27] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, 2019.
- [28] B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarathy, "State and topology estimation for unobservable distribution systems using deep neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.
- [29] M. Huang, Z. Wei, and Y. Lin, "Forecasting-aided state estimation based on deep learning for hybrid ac/dc distribution systems," *Applied Energy*, vol. 306, p. 118119, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261921013982>
- [30] A. Winter, M. Igel, and P. Schegner, "Supervised learning approach for state estimation in distribution systems with missing input data," in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. IEEE, 2021, pp. 1–5.
- [31] S. Tian, F. Zhu, J. Shen, X. Yang, Y. Fu, Y. Mi, and P. Ling, "Distributed state estimation of active distribution network considering mixed-frequency measurement data hierarchical encryption," *Applied Energy*, vol. 388, p. 125661, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261925003915>
- [32] Q.-H. Ngo, B. L. Nguyen, T. V. Vu, J. Zhang, and T. Ngo, "Physics-informed graphical neural network for power system state estimation," *Applied Energy*, vol. 358, p. 122602, 2024.
- [33] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4347–4356, 2020.
- [34] K. R. 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, 2019.
- [35] D. Cao, J. Zhao, W. Hu, N. Yu, J. Hu, and Z. Chen, "Physics-informed graphical learning and bayesian averaging for robust distribution state estimation," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 2879–2892, 2023.
- [36] J. A. Massignan, J. B. London, M. Bessani, C. D. Maciel, R. Z. Fannucchi, and V. Miranda, "Bayesian inference approach for information fusion in distribution system state estimation," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 526–540, 2021.
- [37] J. Zhang, J. Zhao, J. Yang, and J. Zhao, "Deep multi-fidelity bayesian data fusion for probabilistic distribution system voltage estimation with high penetration of pvs," *IEEE Transactions on Power Systems*, 2023.
- [38] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 548–555, 2005.
- [39] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52 138–52 160, 2018.
- [40] R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron, "Explainable artificial intelligence (xai) techniques for energy and power systems: Review, challenges and opportunities," *Energy and AI*, vol. 9, p. 100169, 2022.
- [41] O. L. Santos, D. Dotta, M. Wang, J. H. Chow, and I. C. Decker, "Performance analysis of a dnn classifier for power system events using an interpretability method," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107594, 2022.
- [42] K. Zhang, J. Zhang, P.-D. Xu, T. Gao, and D. W. Gao, "Explainable ai in deep reinforcement learning models for power system emergency control," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 2, pp. 419–427, 2021.
- [43] B. Tan, J. Zhao, T. Su, Q. Huang, Y. Zhang, and H. Zhang, "Explainable bayesian neural network for probabilistic transient stability analysis considering wind energy," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2022, pp. 1–5.
- [44] W. Liao, B. Schäfer, D. Qin, G. Zhang, Z. Wang, and Z. Yang, "An explainable framework for machine learning-based reactive power optimization of distribution network," *arXiv preprint arXiv:2311.03863*, 2023.
- [45] J. Zhao, G. Zhang, M. L. Scala, and Z. Wang, "Enhanced robustness of state estimator to bad data processing through multi-innovation analysis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1610–1619, 2017.
- [46] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, "Hands-on bayesian neural networks—a tutorial for deep learning users," *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 29–48, 2022.
- [47] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 1613–1622. [Online]. Available: <https://proceedings.mlr.press/v37/blundell15.html>
- [48] J.-L. Lin, R. Krishnan, K. R. Ranipa, M. Subedar, V. Sanghavi, M. Arunachalam, O. Tickoo, R. Iyer, and M. T. Kandemir, "Quantization for bayesian deep learning: Low-precision characterization and robustness," in *2023 IEEE International Symposium on Workload Characterization (IISWC)*. IEEE, 2023, pp. 180–192.
- [49] R. Krishnan, P. Esposito, and M. Subedar, "Bayesian-torch: Bayesian neural network layers for uncertainty estimation," Jan. 2022. [Online]. Available: <https://github.com/IntelLabs/bayesian-torch>
- [50] X. Meng, H. Babae, and G. E. Karniadakis, "Multi-fidelity bayesian neural networks: Algorithms and applications," *Journal of Computational Physics*, vol. 438, p. 110361, 2021.
- [51] X. Meng and G. E. Karniadakis, "A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse pde problems," *Journal of Computational Physics*, vol. 401, p. 109020, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021999119307260>
- [52] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.
- [53] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International conference on machine learning*. PMIR, 2017, pp. 3145–3153.