Deep Learning Framework for Low-Observable Distribution System State Estimation with Multi-timescale Measurements

Xihai Zhang, *Student Member, IEEE,* Shaoyun Ge, Yue Zhou, *Member, IEEE,* Hong Liu *Member, IEEE*

***Abstract*— The deployment of micro-phasor measurement units, supervisory control and data acquisition systems, and smart meters has revolutionized distribution systems, moving toward more intelligence and facilitating state estimation. However, the asynchronous nature and varying sample rates of the measurements raise a significant challenge to state estimation in low-observable distribution systems. This paper proposes a novel data-driven framework for state estimation in low-observable distribution systems. The framework incorporates super-resolution imputation techniques that are capable of handling multi-time scale measurements. Specifically, the proposed framework utilizes Wasserstein divergence conditional adversarial networks to implement multivariate super-resolution imputation. This involves employing a gated recurrent unit generator with additive attention to impute super-resolution measurements, as well as a discriminator with self-attention to approximate the Wasserstein divergence between the imputed values and the ground truth. Furthermore, a tailor-made physical-guided bilinear neural network is employed to estimate the operational state of low-observable distribution systems by leveraging matrix factorization techniques to capture the low-rank characteristics. Simulation results illustrate the superiority of the proposed approach from the standpoint of multi-timescale measurements imputation and state estimation in low-observable distribution systems.**

***Index Terms*— Low-observable distribution system, Multi-timescale measurements, Conditional adversarial networks, Bilinear neural networks, Matrix factorization.**

I. Introduction

DISTRIBUTION systems are experiencing substantial transformations due to the increasing integration of distributed energy resources (DERs). The deployment of DERs promotes local energy trading but brings bidirectional power flow and poses risks of breaking distribution system security constraints [1]. Moreover, the DERs largely integrated within the distribution systems are operated 'behind-the-meter' and cannot be directly monitored by most utilities, which renders the systems unobservable/low-observable [2]. With the emergence of smart meters and information and communication technologies, situational awareness and orientation for distribution systems are improved which makes distribution system state estimation (DSSE) on the horizon.

In general, the challenges of state estimation in low-observable distribution systems have converged in how to reconcile multi-timescale measurements into the consistent temporal resolution and further facilitate DSSE via the limited availability of real-time measurements. In terms of imputation of multi-timescale measurement, it can be approached using numerical approaches or machine learning techniques. The numerical approaches attempt to generate unknown values based on extrapolation or interpolation. In [3], a comprehensive comparison of stepwise, linear extrapolation, and linear interpolation for measurement generation is evaluated. The machine learning-based approaches mainly focus on statistical learning theory to impute unmeasured points. Specifically, the extended Kalman filter (EKF) [4], multi-task Gaussian process [5], [6], and recursive Gaussian process [7] are proposed to align multi-timescale measurements. Nevertheless, the extrapolation/interpolation approaches are incompetent to recover the nature of the measurements since each meter is considered separately but the relationship of other measurements in distribution systems is ignored. The EKF has drawbacks in tackling highly non-linear systems due to the limited ability of the Jacobian matrix. Furthermore, the adoption of Gaussian noise is not realistic in real-life application scenarios. The Gaussian process and its variants suffer from cubic complexity to data size, which has its limitations in the era of big data.

It is worth noting that the above-mentioned approaches assume all measurements in a global synchronous condition but the realistic measurements have multiple timescales and rarely synchronize. Refs. [8], [9] introduce normally distributed out-of-date signals to model asynchronous smart meters for monitoring load variations. However, no study has been conducted to assess the statistical distribution of short-term load variation [8]. Ref. [10] presents a low-rank method to mitigate asynchronous errors in grid monitoring, but it relies on low-rank assumptions and is sensitive to outliers. Therefore, it is a challenging question to impute non-synchronized multi-timescale measurements.

Traditional DSSE is heavily explored via weighted least squares (WLS) and its variants [11]. However, the low-observability conditions render the vector of measurements

This work was supported by the Tianjin Science and Technology Program under Grant 22JCZDJC00820. *(Corresponding Author: Hong Liu)*

X. Zhang, S. Ge, and H. Liu are with the School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China. (e-mail:xihaizhang@tju.edu.cn; syge@tju.edu.cn; liuhong@tju.edu.cn)

Y. Zhou is with the Department of Electrical and Electronic Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, U.K (e-mail: ZhouY68@cardiff.ac.uk)

less than the quantities to estimate, resulting in the gain matrix not existing or not unique, which makes WLS *under-determined* and inapplicable to low-observable distribution systems. Although pseudo-measurement is the most intuitive means to address insufficient measurements [12], dramatically 'behind-the-meter' information makes the typical load profile vary from scenario to scenario and difficult to forecast. The existing low-observable DSSE is focused on using deep learning [13], [14], physical-guided deep learning [15]–[17], matrix completion (MC) techniques [18], [19] and signal processing approaches [20], [21]. The existing deep learning approaches require the construction of a comprehensive dataset pertaining to measurements and state variables, which poses a big challenge in low-observable distribution systems. Distinguished from the purely data-driven approaches, there is a growing interest in integrating physics-based prior knowledge with deep learning approaches. However, the DSSE approaches proposed in [15] and [16] require around $10\%\sim15\%$ of nodes to install phasor measuring units, raising questions about the feasibility of these costs. Ref. [17] addresses low-observable DSSE caused by cyber-attacks or physical failures unexpectedly leveraging historical fully observable data. However, the applicability of this approach in distribution systems with persistent low-observable conditions remains uncertain. It is worth noting that substantial physical-guided deep learning approaches are customarily designed for estimating rotor angle and frequency of generator [22] or observable DSSE [23]–[25]. However, these topics are beyond the scope of our current work. The MC involves singular value decomposition which requires a high computational complexity at each iteration. Additionally, the network losses and connection ways (Wye and Delta) in distribution systems are not considered in the existing works. It is worth noting that high R/X ratios in distribution systems make network losses become nonnegligible and the linearized three-phase power flow model leaves more approximate errors. The signal processing approach has specific requirements for measurement place and number in distribution systems since it only involves one type of measurement, which causes an additional investigation on optimal measurement placement. Moreover, the inappropriate measurement placement will limit the performance of DSSE. Therefore, it is an ongoing topic to facilitate state estimation in low-observable distribution systems considering network losses, computation complexity, and arbitrary measurement placement.

This work endeavors to address the aforementioned research gaps by developing a data-driven paradigm coupled with promising machine learning techniques to learn AC power flow, while accommodating multi-timescales distribution systems measurements. The proposed data-driven paradigm encompasses two primary components: super-resolution imputation of distribution system measurements via conditional adversarial networks (CANs), and the implementation of DSSE using physical-guided bilinear neural networks (BNNs). Specifically, a novel CAN is proposed to produce more convincing and robust results that align with distribution system measurements manifold by employing a customized generator and discriminator. It is worth noting that the proposed CAN is distinguished with condition generative adversarial network [26], as our generator focuses on learning deterministic outputs based on conditional information solely and without any noise inputs. This approach facilitates a better capture of the mapping relationships among measurements and has been successfully applied in image tasks [27], [28]. In addition, the low-observable DSSE is tackled using physical-guided BNNs in conjunction with matrix factorization techniques. Unlike existing deep learning approaches, the proposed physical-guided BNNs do not necessitate an extensive dataset of measurements and state variables across all distribution systems. Its salient features and advantages are as follows.

1) A super-resolution imputation approach is proposed for distribution system multi-timescale measurements via CANs with temporal attention. It utilizes content loss and adversarial loss to favor imputation that resides on the manifold. Moreover, the asynchronous measurements are alleviated by proposing a gate recurrent unit (GRU) and attention mechanism to extract temporal features of aggregated adjacent timestamp measurements.
2) A physical-guided BNN is proposed for low-observable DSSE. It employs matrix factorization to characterize the low-rank natural distribution system state variables, which are further approximated via BNNs. Furthermore, the physical laws governing the distribution systems are exploited to regularize the proposed learning model.

The rest of this article is organized as follows. Section II briefly introduces existing distribution systems measurements and general methodology framework. Section III proposes Wasserstein divergence CANs with temporal attention and its key designs. Section IV illustrates the proposed physical-guided BNNs. The proposed data-driven low-observable DSSE with super-resolution imputation is validated in Section V, and Section VI concludes this paper.

II. Distribution System Measurements and General Methodology Framework

In distribution systems, three primary types of measurements are typically employed: micro-phasor measurement units (μPMUs), supervisory control and data acquisition (SCADA), and advanced metering infrastructure (AMI). The μPMUs are specialized devices that provide high-resolution voltage and ampere phasor measurements with timestamps. Nevertheless, μPMUs are typically installed only at selected strategic locations because it is a relatively costly technology. SCADA systems serve as the main source of measurements within distribution networks. They provide information such as node voltage amplitudes, branch currents, branch power flows, and node injection powers. It is worth noting that the SCADA measurements lack explicit time-stamp information, which makes it challenging to ensure global synchronization across the system. The AMI is a critical infrastructure of smart distribution systems, which consists of smart meters, communication networks, and intra-user networks. AMI can provide a wide range of electricity metering data with timestamps [29]. Additionally, certain AMI meters with high accuracy class offer superior measurement accuracy compared to SCADA meters. However, AMI suffers from limitations in real-time

TABLE I
THE COMPARISONS OF MEASUREMENTS IN DISTRIBUTION SYSTEMS.

	μPMU [30]	SCADA [31]	AMI [32]
Data	voltage phasors line current phasors	voltage/line current magnitude branch/injection power	voltage magnitude injection power
Accuracy class	Changes in time of interest, 0.06 - 0.5	Changes in types of meters, 0.15, 0.3, 0.6, 1.2	Changes in types of meters, 0.1, 0.2, 0.5
Timestamp	✓	×	✓
Sample Rate	sub-second	few seconds to minute	15min, 30min, 60min

performance since it primarily records electricity consumption information from end-users. Therefore, integrating and reconciling these measurements effectively is crucial for the analysis and control of distribution systems. The comparisons of these multi-timescale measurements are summarized in Table I.

A novel data-driven low-observable DSSE with super-resolution data imputation is proposed in this paper. Sampling the time alignment measurements between high and low temporal resolutions serves as a training set to impute and harmonize the low temporal resolution measurements to the same temporal resolution and further facilitate the low observable DSSE. The framework of the general methodology is illustrated in Fig. 1.

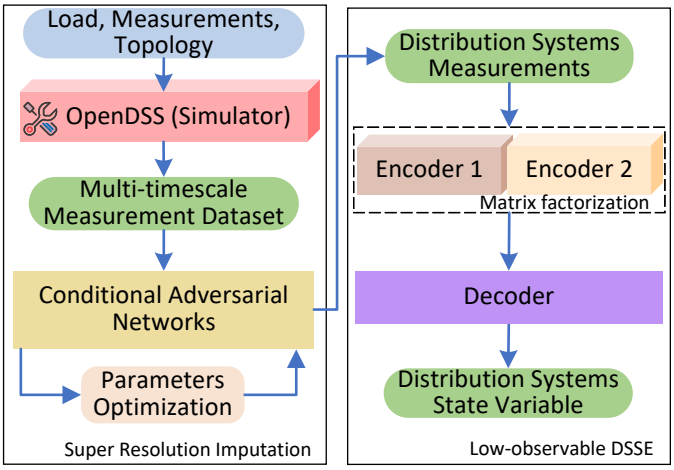


Fig. 1. General framework of the proposed super-resolution imputation enabled DSSE.

For the super-resolution imputation, a novel CAN is adopted to reconcile distribution systems measurements into a consistent temporal resolution. In cases where historical measurement recordings are insufficient, the OpenDSS software can be utilized to generate a comprehensive dataset based on the three-phase unbalanced distribution systems power flow model. To alleviate inevitable asynchronous measurement phenomena, snapshot measurements are incorporated with adjacent 2τ time slots (τ time slots before and after snapshot measurements) and modeled as sequential data with the shape of $(2\tau + 1) \times n$, where n is the number of measurements. During model training, the proposed CANs prioritize structural risk, which represents the risk associated with selecting a complex model that overfits the training data, resulting in poor generalization to new, unseen data. This approach aims to strike a balance between model complexity and the model's ability to generalize well beyond the training data [33], rather than solely focusing on empirical risk measured by a mean squared error on the training dataset, which quantifies the discrepancy between the prediction and the ground truth. In this condition, the adversarial loss and content loss are incorporated comprehensively as structural risks to resolve the super-resolution imputation of multi-timescale measurements effectively. Consequently, it has the potential to enhance the accuracy and robustness of aligning multi-timescale measurements into a consistent temporal resolution. By employing the CAN framework and incorporating structural risk, the proposed methodology has the potential to offer an improved solution for super-resolution imputation in distribution systems, enabling a more accurate representation of the underlying relationship among measurements.

In the context of low-observable DSSE, a physical-guided BNN is introduced to validate the state estimation. Generally speaking, there exist two types of correlations in distribution systems: spatial correlations among locations, and quantitative correlations between measurements. The spatial correlations reflect the relationship between loads at neighboring buses in distribution systems, which is a system-dependent property. However, the quantitative correlations are considered more general due to the approximate linear representation of distribution systems [34]. This approximate linear representation correlates certain measurements, implying that the measurements within system are not entirely independent but can be expressed in terms of a smaller set of underlying patterns. This results in a low-rank characterization within distribution systems. The proposed physical-guided BNNs consist of two encoders and a decoder. The encoders are inspired by matrix factorization techniques, which are designed to capture the low-rank property observed in distribution systems. Subsequently, the outputs are reshaped into matrix form and subjected to matrix multiplication before being fed into the decoder. To improve the performance of DSSE, the proposed methodology integrates the physical laws, which act as a regularization mechanism for constraining the learning process to generate more realistic and physically meaningful results. Therefore, the proposed methodology is capable of providing a more comprehensive and accurate estimation of state variables, leading to a better understanding of the power flow within the low-observable distribution system.

III. Wasserstein Divergence CANs with Temporal Attention for Super Resolution Imputation

In this section, Wasserstein divergence CANs with temporal attention are proposed to resolve super-resolution data imputation, which comprises a GRU generator with additive attention and a discriminator with self-attention. It aims to align low-resolution measurements (z) with high-resolution measurements (x) in the temporal domain. It is worth noting that the gate design, attention mechanism, and adversarial training paradigm are widespread in language processing tasks exhibiting similarities with super-resolution imputation tasks. Specifically, both tasks rely on leveraging contextual information to capture relevant features and facilitate corresponding tasks. Therefore, these mechanisms inspire the proposed approach and can be well-suited for super-resolution imputation tasks. The overall architecture is shown in Fig. 2.

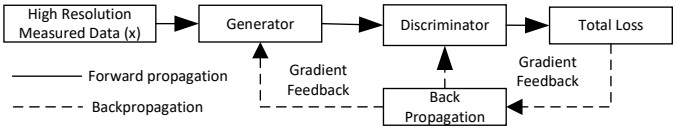


Fig. 2. Architecture of Wasserstein divergence CANs with temporal attention.

A. GRU Generator with Additive Attention

The generator leverages high temporal resolution measurements as a guide to impute points between intervals of low temporal resolution measurements. Specifically, the generator learns the mapping function from the sequential high temporal resolution measurements to low temporal resolution measurements and generates the imputation of unsampled low temporal resolution measurements. Due to the temporal nature of the sequence data, extracting features by flattening the raw data becomes challenging. Therefore, we propose a GRU-based generator with additive attention mechanisms to encode and focus on task-related information, as illustrated in Fig. 3.

The GRU is a recurrent neural network (RNN) variant composed of interconnected single-cell neurons, designed to address the vanishing gradient problem while maintaining computational efficiency. The gate mechanisms within the GRU regulate the flow of information throughout the network, enabling it to capture dependencies over longer sequences more effectively [35]. It processes information through gate mechanisms and the output hidden state ($\boldsymbol{h}_t$) of a single cell neuron is shown in (1).

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}_r \boldsymbol{x}_t + \boldsymbol{U}_r \boldsymbol{h}_{t-1} + \boldsymbol{b}_r]) \tag{1a}$$
$$\boldsymbol{u}_t = \sigma(\boldsymbol{W}_u \boldsymbol{x}_t + \boldsymbol{U}_u \boldsymbol{h}_{t-1} + \boldsymbol{b}_u]) \tag{1b}$$
$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W}_h \boldsymbol{x}_t + \boldsymbol{U}_h(\boldsymbol{r}_t \odot \boldsymbol{h}_t) + \boldsymbol{b}_h) \tag{1c}$$
$$\boldsymbol{h}_t = (1 - \boldsymbol{u}_t) \odot \boldsymbol{h}_t + \boldsymbol{u}_t \odot \tilde{\boldsymbol{h}}_t \tag{1d}$$

where σ and $\tanh$ are logistic sigmoid and hyperbolic tangent function, respectively. $\boldsymbol{r}_t \in (0,1)^{(2\tau+1)\times e}$ and $\boldsymbol{u}_t \in (0,1)^{(2\tau+1)\times e}$ are the outputs of reset and update gates. $\boldsymbol{W}_{r,u,h} \in \mathbb{R}^{n\times e}$, $\boldsymbol{U}_{r,u,h} \in \mathbb{R}^{e\times e}$, and $\boldsymbol{b}_{r,u,h} \in \mathbb{R}^{e}$ are configurable parameters. $\boldsymbol{x}_t \in \mathbb{R}^{(2\tau+1)\times n}$ represents the input high temporal resolution data at time t. $\boldsymbol{h}_{t-1} \in \mathbb{R}^{(2\tau+1)\times e}$ and $\tilde{\boldsymbol{h}}_t \in \mathbb{R}^{(2\tau+1)\times e}$ are the hidden state at time $t-1$ and temporary hidden state at time t. e denotes the number of hidden features. $\odot$ represents Hadamard product.

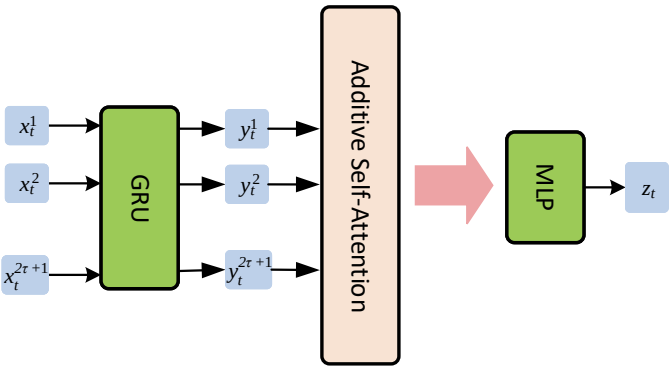


Fig. 3. The schematic of GRU generator with additive attention.

On the basis of GRU output, we further introduce additive attention mechanisms to improve cognitive ability by selecting some key information while ignoring unimportant information. The additive attention mechanism is for sequence-to-sequence tasks that involve attention scores based on a learned compatibility score derived from the input sequence to focus on relevant parts via a combination of linear transformations and a non-linear activation function [36]. Specifically, the scores are computed using a learned linear transformation followed by a non-linear activation function, and the token $\boldsymbol{c}_t$ represents summarized hidden states as shown in (2).

$$\boldsymbol{c}_t = \sum_{j=1}^{2\tau+1} \boldsymbol{h}_{t,j} \times softmax(\boldsymbol{W}_v \tanh(\boldsymbol{W}_q \boldsymbol{h}_{t,j} + \boldsymbol{W}_k \boldsymbol{h}_{t,j})) \tag{2}$$

where $\boldsymbol{W}_v \in \mathbb{R}^{d\times 1}$, $\boldsymbol{W}_q \in \mathbb{R}^{e\times d}$ and $\boldsymbol{W}_q \in \mathbb{R}^{e\times d}$ are learnable parameters and d represents the latent feature dimension in additive attention.

Ultimately, a multi-layer perceptron (MLP) is employed as a post-processing step following the token ($\boldsymbol{c}_t \in \mathbb{R}^d$) to generate the imputation of unmeasured low temporal resolution points.

B. Discriminator with Self Attention

The discriminator is proposed to approximate Wasserstein divergence between ground-truth measurements and generated measurements along with the given condition information. Since the condition information is in the form of sequential data, a self-attention mechanism is employed to select the relevant important condition information.

The intractable nature of Wasserstein divergence poses a challenge to measuring the quality of generated value. However, it can be tracked by the Kantorovich-Rubinstein dual [37], as shown in (3).

$$\inf_D \mathbb{E}[D(\boldsymbol{z}_t|\boldsymbol{x}_t)] - \mathbb{E}[D(\tilde{\boldsymbol{z}}_t|\boldsymbol{x}_t)] + k\mathbb{E}[||\nabla G(\hat{\boldsymbol{z}}_t|\boldsymbol{x}_t)||^p] \tag{3}$$

where k and p are the coefficient and power of Wasserstein divergence, respectively. $\tilde{\boldsymbol{z}}_t \in \mathbb{R}^m$ and $\boldsymbol{z}_t \in \mathbb{R}^m$ are generated

and ground-truth measurements. $\hat{\boldsymbol{z}}_t \in \mathbb{R}^m$ is a linear combination of generated and ground-truth measurements shown as (4).

$$\hat{\boldsymbol{z}}_t = \mu\tilde{\boldsymbol{z}}_t + (1-\mu)\boldsymbol{z}_t \tag{4}$$

where μ is a vector sampled for uniform distribution $U[0,1]$.

Therefore, the discriminator needs to receive conditional information and object (ground-truth or generated) data and compute Wasserstein divergence. To estimate Wasserstein divergence more effectively, we employ self-attention mechanisms in the discriminator. More specifically, it computes a weighted sum of the value vectors ($\boldsymbol{V} \in \mathbb{R}^{(2\tau+1)\times d_v}$) based on the similarity between query vectors ($\boldsymbol{Q} \in \mathbb{R}^{1\times d_k}$) and key vectors ($\boldsymbol{K} \in \mathbb{R}^{(2\tau+1)\times d_k}$). The formula is shown as (5).

$$\boldsymbol{e}_t = softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}})\boldsymbol{V} \tag{5}$$

where d_k is the dimension of queries and keys, d_v is the dimension of values. $\boldsymbol{Q}$ is the affine transformation of $\boldsymbol{z}_t$ with learnable parameters $\boldsymbol{W}^q \in \mathbb{R}^{e\times d_k}$. $\boldsymbol{K}$ and $\boldsymbol{V}$ are the affine transformation of $\boldsymbol{x}_t$ with learnable parameters $\boldsymbol{W}^k \in \mathbb{R}^{n\times d_k}$ and $\boldsymbol{W}^v \in \mathbb{R}^{n\times d_v}$, respectively. $\boldsymbol{e}_t \in \mathbb{R}^{d_v}$ is the embeddings of the self-attention mechanism.

By contacting the embeddings and query vectors into an MLP, the Wasserstein divergence is estimated straightforwardly to use the proposed discriminator.

C. Loss function and training

The total loss function of proposed CANs includes content loss, adversarial loss, and Wasserstein divergence penalty regularization. Discriminator aims to maximize Wasserstein divergence for encouraging perceptually superior solutions residing in the manifold of ground-truth measurements, it can be smoothly derived as shown in (6).

$$\mathcal{J}_D = D(\tilde{\boldsymbol{z}}|\boldsymbol{x}) - D(\boldsymbol{z}|\boldsymbol{x}) - k||\nabla D(\hat{\boldsymbol{z}}|\boldsymbol{x})||^p \tag{6}$$

The generator loss includes content loss and Wasserstein divergence adversarial loss. The content loss evaluates the error between the imputation results and the ground truth, which is instrumental in guiding the generator to impute more realistic values, as shown in (7).

$$\mathcal{L}_{content} = \frac{1}{N}\sum_{i=1}^{N}||\tilde{\boldsymbol{z}}_i - \boldsymbol{z}_i||_l^l \tag{7}$$

where N, $\tilde{\boldsymbol{z}}_i$, and $\boldsymbol{z}_i$ denote the number of estimated low temporal resolution measure sensors, the generated measurements, and the ground-truth measurements. l means the order of loss, $l=1$ means the mean absolute error (MAE) while $l=2$ means the mean square error (MSE). It is worth noting that both MAE and MSE are intuitive metrics in power systems to impute multi-timescale measurements [5]–[7].

The Wasserstein divergence adversarial loss encourages the generator to favor solutions that reside on the manifold of ground-truth measurements by trying to fool the discriminator. Therefore, the resulting final objective loss function for the generator is to minimize the sum of the content loss and generation loss, as shown in (8).

$$\mathcal{J}_G = \mathcal{L}_{content} + \lambda_{gen}D(\tilde{\boldsymbol{z}}|\boldsymbol{x}) \tag{8}$$

where λ_{gen} is a hyper-parameter that controls the proportion between the content loss and the discriminative loss.

The overall procedure of the proposed CANs is demonstrated in Algorithm 1.

Algorithm 1: Wasserstein Divergence CANs with Temporal Attention

Input: Batch size b, coefficient k and λ, power p, learning rate α_G and α_D, decay rate d.

Initial generator $\boldsymbol{w}_G$ and discriminator $\boldsymbol{w}_D$ parameters;
for epoch = 1 to M **do**
  Sample sequential high-resolution data $x_1, x_2, ..., x_b$ from P_r ;
  Sample low-resolution data $z_1, z_2, ..., z_b$ from P_r ;
  Update discriminator network (w_D) by ascending: $w_D \leftarrow w_D + \alpha_D \nabla \mathcal{J}_D$;
  if time step $\mod d$ **then**
    Update generator network (w_G) by descending: $w_G \leftarrow w_G - \alpha_G \nabla \mathcal{J}_G$;
  end
end

IV. Physical-guided BNNs for Low-observable DSSE

In this section, we present a low-observability DSSE algorithm, which employs novel physical-guided BNNs to estimate the system states from limited observations. BNNs is a novel neural network architecture that incorporates bilinear operations to model interactions between pairs of input variables. The bilinear operations involve computing the element-wise product of features from different sources, facilitating the network to capture complex relationships and more natural regularization through rank restriction [38]. Given that the system states in distribution systems exhibit a low-rank property, the matrix $\boldsymbol{X}$ is structured such that each row represents a phase and columns represent the quantity relevant to that bus. Specifically, the row of matrix $\boldsymbol{X} \in \mathbb{R}^{u\times v}$ is formulated as:

$$[v_{b,re}, v_{b,im}, |v_b|, P_b, Q_b] \tag{9}$$

where $v_{b,re}$, $v_{b,im}$, and $|v_b|$ represent the real, imaginary, and amplitude parts of voltage phasors at each phase of non-slack buses, respectively. P_b and Q_b are the active and reactive power injections.

To address the DSSE problem, it is formulated as a rank-minimization problem, where the goal is to estimate the low-rank structure of $\boldsymbol{X}$. The minimization of rank directly is a non-convexity and NP-hardness problem, thus the nuclear norm is used as a surrogate for rank. However, its recast is computationally inefficient for large matrices and extensive parameter tuning. Inspired by low-rank matrix factorization, it is assumed that $\boldsymbol{X}$ can be decomposed into two matrices:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^T \tag{10}$$

where $\boldsymbol{U} \in \mathbb{R}^{u\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{v\times r}$, such that $\mathrm{rank}(\boldsymbol{X}) \leq \min(\mathrm{rank}(\boldsymbol{U}), \mathrm{rank}(\boldsymbol{V}))$ and $r < \min(u, v)$, which achieve the low-rank property of matrix $\boldsymbol{X}$. To estimate $\boldsymbol{U}$ and $\boldsymbol{V}$, we develop two encoders with parameterized $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ to generate matrices $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$, which is denoted as:

$$\hat{\boldsymbol{u}} = f_1(\boldsymbol{x}; \boldsymbol{\theta}_1) \qquad \hat{\boldsymbol{v}} = f_2(\boldsymbol{x}; \boldsymbol{\theta}_2) \tag{11}$$

where $\hat{\boldsymbol{u}} \in \mathbb{R}^{1\times ur}$ and $\hat{\boldsymbol{v}} \in \mathbb{R}^{1\times vr}$ are the extension of matrix $\boldsymbol{U}$ and $\boldsymbol{V}$ into the vector form, respectively.

By reshaping the vector $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ into matrix $\hat{\boldsymbol{U}} \in \mathbb{R}^{u\times r}$ and $\hat{\boldsymbol{V}} \in \mathbb{R}^{v\times r}$, then implementing $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{V}}^T$ and reshape into a vector form $\hat{\boldsymbol{y}} \in \mathbb{R}^{1\times uv}$ subsequently. Finally, we feed it to a decoder with parameterized $\boldsymbol{\psi}$ to obtain the final output:

$$\hat{\boldsymbol{z}} = g(\hat{\boldsymbol{y}}; \boldsymbol{\psi}) \tag{12}$$

where $\hat{\boldsymbol{z}} \in \mathbb{R}^{1\times uv}$ is reshaped into a matrix with the shape of $\boldsymbol{Z} \in \mathbb{R}^{u\times v}$. It is worth emphasizing that the decoder can be viewed as $\boldsymbol{Z} = \boldsymbol{W}[\hat{\boldsymbol{Y}}; \boldsymbol{1}]^T$ in mathematical. Therefore, the maximum rank of $\boldsymbol{Z}$ is $r + 1$ due to the properties of matrix products.

In contrast to the electrical model-agnostic data-driven approach, the multidimensional Ohm's law is investigated and given by (13) based on the Z-Bus method [39], which is preferred over Newton-Raphson methods due to the predominant high R/X ratios in distribution networks.

$$\begin{bmatrix} \boldsymbol{i} \\ \boldsymbol{i}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{Y}_{NS} \\ \boldsymbol{Y}_{SN} & \boldsymbol{Y}_{SS} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v}_s \end{bmatrix} \tag{13}$$

where $\boldsymbol{i}_s$ is the complex current injection of the slack bus. Matrices $\boldsymbol{Y}$, $\boldsymbol{Y}_{NS}$, $\boldsymbol{Y}_{SN}$, and $\boldsymbol{Y}_{SS}$ are formed by concatenating the admittance matrices and solved by Picard's iteration for the nonlinear power flow model. Besides that, if the prior knowledge about line parameters is unavailable, the data-driven surrogate model can be developed for mapping the relationship between injection power and power flow, even with learning power flow in unobservable distribution system topology and parameter [40].

The total loss function of the proposed physical-guided BNNs is composed of matrix completion error, power flow reconstructed error, and the regularization to $\boldsymbol{Z}$. The matrix completion error measures the difference between the observed entries of $\boldsymbol{X}$ and the corresponding entries in $\boldsymbol{Z}$ at the known positions P_Ω. The power flow reconstruction error quantifies the discrepancy between the estimated system voltage $\boldsymbol{v}$ obtained from the physical-guided BNNs and the result $\hat{\boldsymbol{v}}$ derived from the power flow model. The regularization term on $\boldsymbol{Z}$ encourages a simpler representation and helps prevent overfitting. It is worth noting that the matrix completion error is a soft constraint, and the data imputation result can be further corrected via the integration of power flow models. The total loss function is shown as (14).

$$\mathcal{J} = ||P_\Omega \odot (\boldsymbol{Z} - \boldsymbol{X})||_F^2 + ||\boldsymbol{v} - \hat{\boldsymbol{v}}||_F^2 + \lambda||\boldsymbol{Z}||_F^2 \tag{14}$$

where $\boldsymbol{v}$ represents the estimation of the physical-guided BNNs, while $\hat{\boldsymbol{v}}$ corresponds to the result obtained from the power flow model. P_Ω denote the known elements index in X. The Frobenius norm $||\cdot||_F$ is used to measure the matrix completion error and the regularization term. The regularization parameter λ controls the trade-off between fitting the data and promoting a simpler solution.

V. Simulation Result

The proposed approach is verified on the three-phase unbalanced IEEE 37 bus system, recommended by the Test Feeder Working Group of the Distribution System Analysis Subcommittee for evaluating the performance of state estimation algorithms [41]. All networks are programmed in Pytorch and implemented on a PC with an NVIDIA GeForce RTX 3080Ti GPU and Intel(R) Core(TM) i9-10900X CPU. The aggregated 24-hour load profile at bus nodes is obtained from [42], and the load distribution for other days is generated by adding a sine wave of random amplitude and random noise terms. Load reactive power is defined proportionately to the load profile connected to the same bus with a lagging power factor between $[0.93, 0.95]$. Moreover, the bus which has its own PV and battery system assumes a three-phase load, while others suppose a single-phase load. The aggregated AMI meters are sampled over 15 minutes for active and reactive power injections over a bus node. For Delta connection load, the AMI meter is installed phase-wise. SCADA measures node voltage magnitudes at 1-minute intervals and the μPMUs are sampled at every 0.1s for obtaining node voltage phasors. The baseline of numerical errors for μPMUs, SCADA, and AMI is modeled as a zero-mean Gaussian distribution, which is truncated to simulate maximum errors of 0.05%, 0.5%, and 0.1% [32], [43]. The formulation of numerical error is shown in (15).

$$\hat{z}_i = z_i(1 + e) \tag{15}$$

where $e \sim \mathcal{N}(0, \varsigma)$ and ς is equal to one-third of corresponding maximum errors specified for μPMUs, SCADA, and AMI. It is worth noting that the error can be modeled in various forms, and we leverage the Gaussian distribution to exemplify the effectiveness of the proposed approach.

Moreover, the asynchronous error is modeled using a Bernoulli Distribution, where there is a 1% possibility of recording measurements within ± 1 step, a 20% possibility within ± 10 steps, and a 10% possibility within ± 5 steps for μPMUs, SCADA, and AMI measurements [30], [44], respectively. The measurements have both numerical error and asynchronous error is noted as the hybrid error. The locations of these meters are indicated in Fig. 4 and distribution systems power flow analysis is implemented by running the OpenDSS software.

To evaluate the proposed approach, peak signal-to-noise ratio (PSNR) and mean absolute error (MAE) are adopted to evaluate the performance of super-resolution imputation and DSSE, which are defined as (16a) and (16b), respectively.

$$\mathrm{PSNR} = 10\log_{10}\frac{N\max\{z_i^2\}_{i=1}^N}{\sum_{i=1}^N(\tilde{z}_i - z_i)^2} \tag{16a}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^N|\tilde{z}_i - z_i| \tag{16b}$$

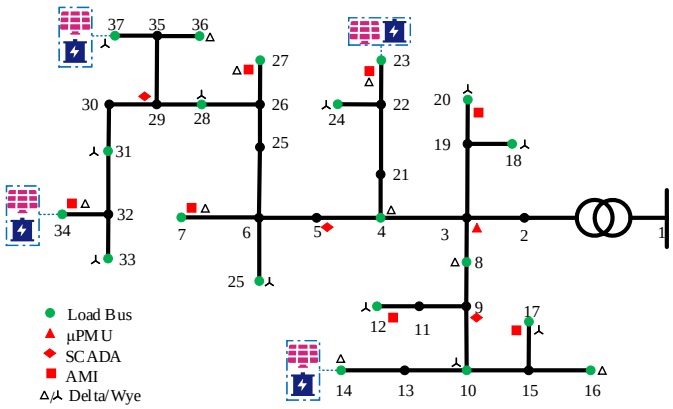


Fig. 4. Measurements configuration for IEEE 37-bus feeder system.

where z_i and $\tilde{z}_i$ are the ground truth and imputated values, respectively.

The degree of observability in distribution systems is quantified via fraction of available data (FAD) as shown in (17).

$$\text{FAD} = \frac{N_m}{N_{total}} \tag{17}$$

where N_m is the number of measurements within the distribution system. N_{total} represents the total physical quantities at the non-slack buses, including the voltage phasors and complex power phasors mentioned in (9). It is worth noting that when the FAD is less than 50%, it will be considered as a low-observability system [19].

A. Super resolution imputation

The comparative tests are implemented among Multi-Task GP (MT-GP) [6], Conditional Variational Auto Encoder (CVAE) [45], Wasserstein Generative Adversarial Networks (WGANs) [46], Super-Resolution CANs (SRCANs) [47], Super-Resolution WCANs (SRWCANs), and SRCANs with Temporal Attention (SRCANTA). It is worth noting the comparison of SRCANs, SRWCANs, and SRCANTA serve as ablation studies aimed at evaluating the performance improvement attributed to different components. The measurements are from distribution system power flow results among 8640000 instances for 10 days. The super-resolution imputation result of AMI and SCADA is shown in Table II. The hyperparameter of the proposed algorithm is formulated by trial and error and summarized as follows. The generator is a single hidden layer GRU with 64 hidden features and 32 hidden features for additive attention. The self-attention mechanism employs 64 hidden features and 32 hidden features in the subsequent neural network within the discriminator. The learning rate of the generator/discriminator is 1e-4/2e-4 and the activation function is set as exponential linear units. The batch size is set as 1024 and λ_{gen} is 1e-3. The coefficient and power of Wasserstein divergence are 2 and 6, respectively. The ablation studies aim to evaluate performance improvements across various aspects:

- SRCANs: Investigating the effectiveness of Wasserstein divergence and attention mechanisms;
- SRWCANs: Focusing specifically on the utilization of attention mechanisms;
- SRCANTA: Exploring the effectiveness of Wasserstein divergence.

It is observed that the proposed approach achieves a competitive performance among other competitors. Comparisons with Multi-Task GP demonstrate that adversarial loss effectively enhances generalization ability. The performance of CVAE and WGANs underscores the critical importance of incorporating conditional information for multi-timescale measurement imputation. Furthermore, Wasserstein divergence has shown improvements in reducing numerical error, while attention mechanisms have proven effective in addressing asynchronous error. Therefore, combining these techniques in the proposed approach is expected to yield significant performance improvements. Furthermore, it is worth noting that once the proposed CAN model is trained and the parameters are saved, the approach can perform measurement imputation in real time. This capability ensures the efficient application of the model in real-life data processing and analysis scenarios.

In scenarios where phasor measurements are unavailable, the SCADA measurements are used directly to reconcile AMI into a consistent temporal resolution with SCADA. It is worth recognizing that the maximum imputation resolution will be significantly lower compared to scenarios with phasor measurements, primarily due to the restricted sampling rate of SCADA. The simulation results are presented in Table III. It is observed that the proposed algorithm outperforms MT-GP, CVAE, and WGANs competitors. The ablation studies involving SRCANs, SRWCANs, and SRCANTA demonstrate improvements related to Wasserstein divergence for numerical error and temporal attention mechanisms for asynchronous error. Furthermore, it is worth noting that the imputation accuracy in terms of MAE and PSNR is lower compared to scenarios with phasor measurement. This highlights the significance of integrating μPMU into low-observable distribution systems to enhance accuracy and overall performance. Besides that, the simulation result shown in Table II and III illustrates that the super-resolution imputation of multi-timescale measurement in distribution systems is not sensitive to asynchronous error but sensitive to numerical error.

To further evaluate the super-resolution imputation performance, the baseline noise levels were amplified stepwise up to 10 times. The simulation results are shown in Fig. 5. It is evident from the results that the imputation outcomes display insensitivity to asynchronous errors, owing to the incorporation of the attention mechanism. However, it is important to note that numerical errors exert a greater impact on super-resolution imputation, as they tend to disrupt the training process of the proposed CANs. Moreover, the hybrid error, which encompasses both asynchronous and numerical errors, leads to a more substantial error in the imputation process. Last but not least, the SCADA imputation exhibits superior capabilities compared to AMI imputation, primarily due to the higher precision of μPMUs measurements. In addition, it is worth noting that in scenarios with significant asynchronous errors or serious data quality issues, it is advisable to pair meters and develop sampling algorithms to address these challenges.

TABLE II
THE SUPER-RESOLUTION IMPUTATION ERROR OF AMI/SCADA IN IEEE 37-BUS FEEDER SYSTEM.

	None Error		Numerical Error		Async. Error		Hybrid Error	
	PSNR(dB)	MAE	PSNR(dB)	MAE	PSNR(dB)	MAE	PSNR(dB)	MAE
MT-GP	44.55/77.44	0.34/0.27	30.07/72.11	1.42/0.54	44.62/77.49	0.34/0.27	30.00/72.14	1.44/0.53
CVAE	43.35/76.56	0.40/0.31	29.20/71.68	1.65/0.56	43.36/76.54	0.39/0.32	29.21/71.74	1.63/0.56
WGANs	35.97/72.80	0.84/0.47	29.94/68.79	1.97/0.79	32.28/72.99	1.29/0.45	27.07/67.70	1.95/0.99
SRCANs	44.13/79.80	0.29/0.22	29.68/72.62	1.49/0.49	46.12/79.27	0.29/0.22	29.63/72.68	1.51/0.59
SRWCANs	44.85/78.13	0.33/0.22	30.35/72.72	1.39/0.49	44.79/77.90	0.33/0.26	30.04/72.68	1.38/0.49
SRCANTA	44.88/79.85	0.29/0.19	26.63/75.81	1.30/0.28	46.17/**79.94**	0.29/**0.18**	31.10/75.37	1.25/0.30
Proposed	**46.08/80.26**	**0.28/0.16**	**31.36/76.55**	**1.03/0.26**	**46.27**/79.78	**0.27**/0.19	**32.32/76.45**	**1.04/0.26**

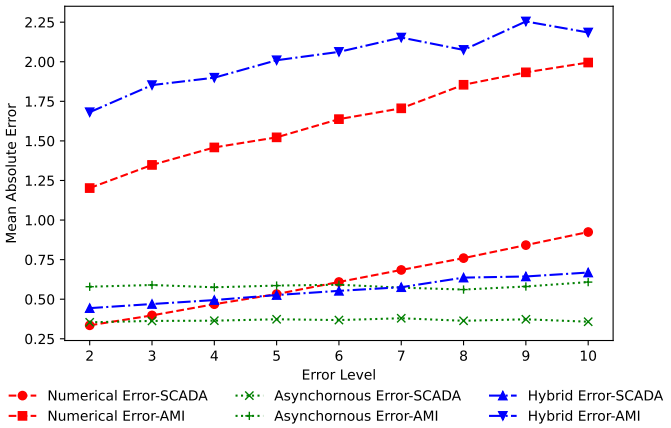
TABLE III
THE SUPER-RESOLUTION IMPUTATION ERROR OF AMI WITH NO PHASOR MEASUREMENTS IN TERMS OF PSNR/MAE.

	None Error	Numerical Error	Async. Error	Hybrid Error
MT-GP	40.23/0.52	22.12/2.82	39.92/0.54	23.61/2.73
CVAE	37.86/0.68	23.17/2.50	37.91/0.68	25.17/2.51
WGANs	31.87/1.40	22.59/3.34	31.59/1.39	22.53/3.30
SRCANs	41.70/0.43	21.85/3.63	41.57/0.45	21.91/3.63
SRWCANs	41.92/0.39	23.55/2.84	41.82/0.39	23.78/2.92
SRCANTA	42.51/0.37	23.53/2.98	42.52/0.37	22.94/2.94
Proposed	**42.63/0.33**	**25.09/2.33**	**43.04/0.34**	**24.61/2.50**

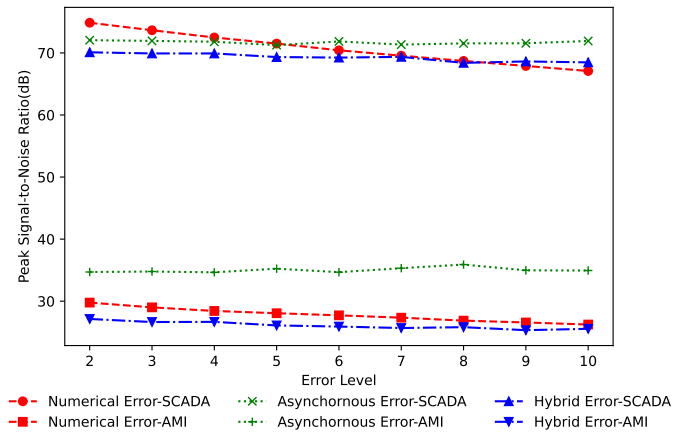
B. Low-observable DSSE

The matrix (9) in the proposed IEEE 37 distribution systems shown in Fig. 4 has 540 elements at non-slack buses. However, only 39 measurements are distributed across distribution systems, encompassing 6 voltage vectors, 9 voltage magnitudes, and 24 power injection readings. Thus, the FAD is 39/540 = 7.59%, significantly below the 50% threshold for an observable system and indicates a notably low observability level. The hyperparameters employed in the proposed physical-guided BNNs comprise two hidden layers within the encoders. The encoder is composed of hidden numbers of 1024, 512, and 512, 256 via trial and error, respectively. On the other hand, the decoder is a single hidden neural network with identical inputs and outputs. The DSSE performance is evaluated based on the absolute error of voltage phasors, considering both magnitude and angle components among the testing set. Fig. 6 demonstrates the simulation results and reveals that the mean absolute error of nominal voltage magnitude across all buses remains below 1. Additionally, the voltage angle exhibits a loss of approximately 1e-4 radians.

In order to showcase the effectiveness of the proposed data-driven DSSE algorithm, we compare it with other data-driven approaches and model-based approaches, that is, physical-guided deep learning (PGDL), BNNs without physical guidance, and low-rank matrix completion (LRMC) [6] as the baseline to evaluate the performance of low-observable DSSE. These algorithms are evaluated to assess the performance of DSSE in low-observable scenarios. The BNNs used in the comparison have the same neural structure as the physical-guided BNNs. The physical-guided deep learning approaches also employ 1024 and 512 hidden features. On the other hand, the LRMC method is solved using the Gurobi software. The comparison of DSSE in terms of error level and computational complexity is shown in Table IV.



(a) The MAE for super-resolution imputation for AMI/SCADA.



(b) The PSNR for super-resolution imputation for AMI/SCADA.

Fig. 5. The MAE/PSNR for super-resolution imputation with the different error levels.

TABLE IV
THE COMPARISON OF DSSE IN TERMS OF ERROR LEVEL AND COMPUTATIONAL COMPLEXITY FOR IEEE 37 BUS SYSTEM.

Algorithm	MAE	PSNR(dB)	Computational time(s)
PGDL	0.44	65.04	8.44e-6
BNNs	2.17	53.50	8.92e-6
LRMC	2.69	52.63	2.40
Proposed	**0.35**	**66.47**	8.87e-6

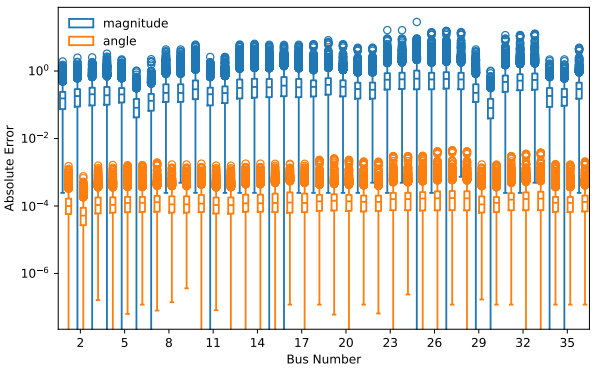


Fig. 6. Representative voltage phasor estimates of the IEEE 37-bus feeder system.

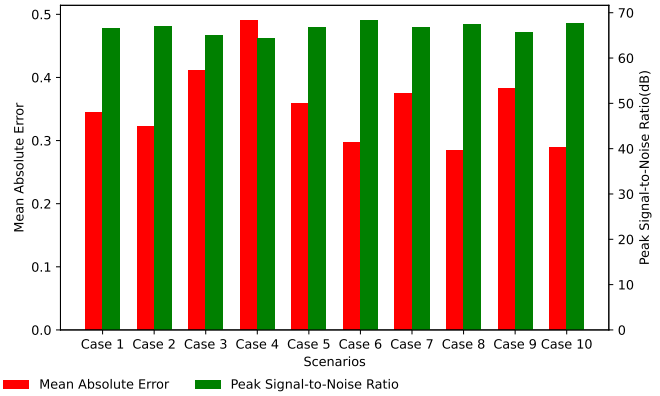


Fig. 8. MAE and PSNR metrics for DSSE at different sensor locations.

To further evaluate the performance of low-observable DSSE, we reduce the measurements stepwise to evaluate the proposed approach on the lower FAD. Specifically, the number of measurements varies between 23 and 38, and these measurements are randomly placed within the distribution system. The random placement of measurements is designed to test the sensitivity of the proposed algorithm to measurement placement under very low observability conditions. The simulation result is shown in Fig. 7, which demonstrates the proposed approach can still exhibit superior performance on 4.26% FAD.

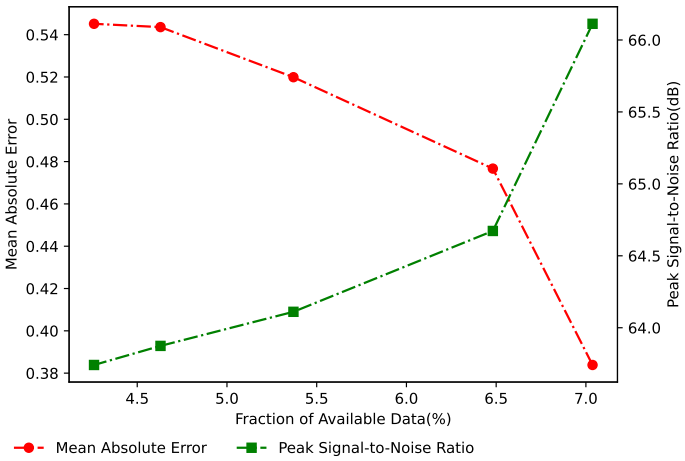


Fig. 7. MAE and PSNR metrics for DSSE at different FADs.

The performance of DSSE under arbitrary measurement locations is shown in Fig. 8. Without loss of generality, some sensors are located in the zero-load bus in order to mimic that the distribution systems are unobservable to the operator. The simulation result illustrates the proposed approach is robust to measuring location changes since the PSNR metrics almost keep constant.

To validate the scalability of the proposed approach for the large distribution network, the proposed approach is implemented on the IEEE 123 bus system, which comprises single-phase, two-phase, and three-phase overhead and underground lines. The measurements were placed randomly to further evaluate the robustness of the pro-posed method under different measurement configuration schemes, adhering to the constraint that there are 85 available measurements out of 1220 elements, representing a very low-observable condition with a FAD rate of 6.97%. The DSSE algorithms are executed 20 times and convergence metrics are illustrated in Table V. The simulation result indicates that the proposed algorithm is robust against variations in measurement placements and consistently outperforms other competing approaches.

TABLE V
THE COMPARISON OF DSSE IN TERMS OF ERROR LEVEL AND COMPUTATIONAL COMPLEXITY FOR IEEE 123 BUS SYSTEM.

Algorithm	MAE	PSNR(dB)	Computational time(s)
PGDL	1.02±0.02	62.08±0.14	9.74e-6±0.01e-5
BNNs	3.59±0.05	51.45±0.21	1.95e-5±0.02e-5
LRMC	5.54±0.54	38.05±0.59	3.04±0.03
Proposed	**0.37**±0.01	**70.14**±0.21	2.02e-5±0.02e-5

VI. CONCLUSION

We present a novel data-driven framework that combines super-resolution imputation with low-observability DSSE. The proposed approach employs two innovative machine learning techniques, that is, Wasserstein divergence CANs with temporal attention and physical-guided BNNs. These methods enable accurate imputation of multi-scale measurements and subsequent estimation of voltage phasors, particularly in scenarios with limited observability. The CANs demonstrate impressive performance, achieving a PSNR of 80.26 dB and 46.08 dB for SCADA and AMI imputation, respectively. Moreover, the physical-guided BNNs exhibit the ability to adapt to various measurement configurations and data sources in real-time, yielding a PSNR of 66.47 dB for DSSE. It is worth noting that the proposed DSSE approach still works even when the FAD is 4.25%.

In the future, we will incorporate transfer learning techniques or graph neural networks into the proposed framework to accommodate the unknown topology changes or

reconfigurations that occur in distribution systems. Besides that, integrating physical prior knowledge into super-resolution imputation for multi-timescale measurements in distribution systems is also an interesting topic.

REFERENCES

[1] Y. Zhou, J. Wu, C. Long, and W. Ming, "State-of-the-art analysis and perspectives for peer-to-peer energy trading," *Engineering*, vol. 6, no. 7, pp. 739–753, 2020.

[2] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312–2322, 2019.

[3] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 421–430, 2015.

[4] A. M. Stanković, V. Švenda, A. T. Sarić, and M. K. Transtrum, "Hybrid power system state estimation with irregular sampling," in *2017 IEEE Power & Energy Society General Meeting*, 2017, pp. 1–5.

[5] S. Dahale and B. Natarajan, "Multi time-scale imputation aided state estimation in distribution system," in *2021 IEEE Power & Energy Society General Meeting (PESGM)*, 2021, pp. 1–5.

[6] D. Shweta and B. Natarajan, "Bayesian framework for multi-timescale state estimation in low-observable distribution systems," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4340–4351, 2022.

[7] S. Dahale and B. Natarajan, "Recursive gaussian process over graphs for integrating multi-timescale measurements in low-observable distribution systems," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3464–3475, 2023.

[8] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2919–2928, 2015.

[9] C. S. Kumar, K. Rajawat, S. Chakrabarti, and B. C. Pal, "Robust distribution system state estimation with hybrid measurements," *IET Generation, Transmission & Distribution*, vol. 14, pp. 3250–3259, 2020.

[10] Y. Yuan, K. Dehghanpour, and Z. Wang, "Mitigating smart meter asynchrony error via multi-objective low rank matrix recovery," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4308–4317, 2021.

[11] 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, 2018.

[12] J. A. D. Massignan, J. B. A. 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, 2022.

[13] 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.

[14] Y. Yuan, K. Dehghanpour, Z. Wang, and F. Bu, "A joint distribution system state estimation framework via deep actor-critic learning method," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 796–806, 2023.

[15] 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.

[16] T. Wu, I. L. Carreño, A. Scaglione, and D. Arnold, "Spatio-temporal graph convolutional neural networks for physics-aware grid learning algorithms," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 4086–4099, 2023.

[17] O. Jonatan, B. Konstantin, B. Andrey, and Z. Gil, "Physics-informed deep neural network method for limited observability state estimation," in *Proc. Workshop Auton. Energy Syst. NREL Golden CO 8/19-20/2020*, 2020, pp. 1–6.

[18] A. Sagan, Y. Liu, and A. Bernstein, "Decentralized low-rank state estimation for power distribution systems," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3097–3106, 2021.

[19] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2520–2530, 2020.

[20] S. S. Saha, A. Scaglione, R. Ramakrishna, and N. G. Johnson, "Distribution systems ac state estimation via sparse AMI data using graph signal processing," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3636–3649, 2022.

[21] A. Akrami, S. Asif, and H. Mohsenian-Rad, "Sparse tracking state estimation for low-observable power distribution systems using D-PMUs," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 551–564, 2022.

[22] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-informed neural networks for power systems," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[23] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep unrolled neural networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069–4077, 2019.

[24] Q. Yang, A. Sadeghi, and G. Wang, "Data-driven priors for robust psse via gauss-newton unrolled neural networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 12, no. 1, pp. 172–181, 2022.

[25] L. Wang, Q. Zhou, and S. Jin, "Physics-guided deep learning for power system state estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 607–615, 2020.

[26] M. Mirza and S. Osindero, "Conditional generative adversarial nets," in *Proceedings of advances in Neural Information Processing Systems*, 2014, pp. 1–6.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[28] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[29] R. Rashed Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473–484, 2014.

[30] A. von Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision micro-synchrophasors for distribution systems: A summary of applications," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, 2017.

[31] K. Sayed and H. Gabbar, "Chapter 18 - SCADA and smart energy grid control automation," in *Smart Energy Grid Engineering*. Academic Press, 2017, pp. 481–514.

[32] American National Standards Institute *ANSI C12.20-2015— Electricity Meters—0.1, 0.2, and 0.5 Accuracy Classes*. [Online]. Available: https://blog.ansi.org/ansi-c12-20-2015-electric-meters-accuracy-classes/

[33] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer New York, NY, 1999.

[34] A. Bernstein and E. Dall'Anese, "Linear power-flow models in multiphase distribution networks," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2017, pp. 1–6.

[35] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks," in *International Conference on Machine Learning*, 2015, p. 2067–2075.

[36] G. Zheng, S. Mukherjee, X. L. Dong, and F. Li, "Opentag: Open attribute value extraction from product profiles," in *International Conference on Knowledge Discovery & Data Mining*, 2018, p. 1049–1058.

[37] J. Wu, Z. Huang, J. Thoma, D. Acharya, and L. Van Gool, "Wasserstein divergence for GANs," in *European Conference on Computer Vision*, 2018, pp. 673–688.

[38] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear convolutional neural networks for fine-grained visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1309–1322, 2018.

[39] M. Bazrafshan and N. Gatsis, "Convergence of the z-bus method for three-phase distribution load-flow with zip loads," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 153–165, 2018.

[40] J. Yuan and Y. Weng, "Support matrix regression for learning power flow in distribution grid with unobservability," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1151–1161, 2022.

[41] K. P. Schneider, B. A. Mather, B. C. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and design basis for the ieee distribution test feeders," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3181–3188, 2018.

[42] X. Zhang, S. Ge, H. Liu, Y. Zhou, X. He, and Z. Xu, "Distributionally robust optimization for peer-to-peer energy trading considering data-driven ambiguity sets," *Applied Energy*, vol. 331, p. 120436, 2023.

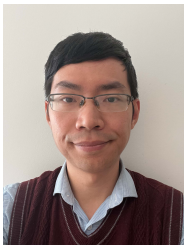
[43] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 5–20, 2013.

[44] J. F. Martins, A. G. Pronto, V. Delgado-Gomes, and M. Sanduleac, "Chapter 4 - smart meters and advanced metering infrastructure," in *Pathways to a Smarter Power System*. Academic Press, 2019, pp. 89–114.
[45] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *International Conference on Neural Information Processing Systems*, vol. 28, 2015, p. 3483–3491.
[46] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 214–223.
[47] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.

Xihai Zhang (Student Member, IEEE) is currently pursuing his Ph.D. degree in electrical engineering at Tianjin University, Tianjin, China. His research interest includes peer-to-peer energy trading, distribution system state estimation, and reinforcement learning.

Shaoyun Ge received his M.S. degree from the School of Electrical Engineering and Automation, Tianjin University, Tianjin, China, in 1991, and the Ph.D. degree from Hong Kong Polytechnic University, Hung Hom, Hong Kong, in 1998. He is currently a Professor at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include distribution system planning, electric vehicle charging facility planning, and smart grid.

Dr. Ge was a recipient of the State Science and Technology Progress Second Class Award in 2005 and 2010.



Yue Zhou (Member, IEEE) received his B.Eng. and Ph.D. degrees in Electrical Engineering from Tianjin University, China, in 2011 and 2016, respectively. He is the Lecturer in Cyber Physical Systems at the School of Engineering of Cardiff University, Wales, UK. His research interests include demand response, peer-to-peer energy trading, and cyber-physical systems.

Hong Liu (Member, IEEE) received his M.S. and Ph.D. degrees from the School of Electrical Automation Engineering in Tianjin University, Tianjin, China, in 2005 and 2009, respectively. He is currently a Professor at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include the planning and operation of distribution systems and integrated energy systems.

Dr. Liu was a recipient of the State Science and Technology Progress Second Class Award in 2010 and 2016.