

Forecasting-aided State Estimation Framework for Distribution Networks with Multi-source Measurement Data Fusion

Yuanshi Zhang, *Member, IEEE*, Qirui Chen, Haizhou Liu, *Member, IEEE*, Qinran Hu, *Senior Member, IEEE*, and Bingxu Zhai

Abstract—Real-time state estimation underpins advanced applications in distribution networks by revealing system operating conditions with high fidelity. Intelligent measurement terminals increasingly combine heterogeneous measurements, including micro-phasor measurement units (μ PMUs), remote terminal units (RTUs), and data transmission units (DTUs). Their complementary strengths can overcome traditional limits in time resolution and accuracy. However, heterogeneous formats, asynchronous updates, and complex temporal dynamics challenge conventional estimators. This paper proposes a unified real-time forecasting-aided state estimation (FASE) framework for distribution networks with multi-source measurement data fusion that fuses μ PMU, RTU, and DTU data. A long short-term memory (LSTM)-based data imputation method is designed to effectively align delayed RTU and DTU updates with μ PMU sampling, markedly reducing imputation errors versus linear and historical-average baselines. Unified measurement equations are formulated for all devices. An improved cubature Kalman filter (CKF) with adaptive robust weighting is adopted to enhance numerical stability and outlier resilience. To capture multimodal operating regimes driven by variable distributed energy resources, a Gaussian mixture model (GMM)-based approach is integrated into the distribution network state transition. Validated on the IEEE 33-bus and 118-bus systems, the proposed FASE framework achieves higher estimation accuracy and computational efficiency than extended Kalman filter (EKF), unscented Kalman filter (UKF), and traditional CKF while meeting real-time constraints.

Index Terms—Distribution network, state estimation, long short-term memory (LSTM), Gaussian mixture model, cubature Kalman filter (CKF), multi-source measurement data fusion.

I. INTRODUCTION

WITH the increasing complexity of modern power systems, accurate state estimation has become crucial

for efficient energy management and system reliability in distribution networks [1], [2]. The growing penetration of distributed energy resources creates new operational challenges. Similarly, the dynamic nature of electricity demands further complicates system operation. These evolving conditions necessitate advanced monitoring and control strategies.

Accordingly, state estimation provides essential situational awareness for distribution network operation and control, state estimation enables essential supporting functions such as line parameter identification [3], system performance monitoring [4], and efficient power flow management [5]. These capabilities form the foundation for advanced distribution network applications.

Traditional state estimation approaches, while effective in transmission networks, face unique limitations in distribution networks. State estimation techniques using micro-phasor measurement units (μ PMUs) and supervisory control and data acquisition (SCADA) systems are well-established in transmission networks [6]. However, distribution networks present distinct challenges, including radial structures, shorter line segments, and frequent topology changes that complicate conventional state estimation approaches [7].

Current measurement infrastructure in distribution networks requires significant enhancement to meet modern operational demands. Existing measurement systems include RTUs, which collect real-time operational data [8], and DTUs, which facilitate communication between field devices and control systems [9]. However, these devices often provide insufficient temporal resolution for accurate state estimation. This limitation has driven the development of μ PMUs, which measure voltage and current phasors at microsecond timescales [10], [11]. These devices provide high-precision data more cost-effectively than traditional PMUs.

Achieving accurate distribution network state estimation requires integrating multiple measurement technologies. Relying on single data source poses several limitations. RTUs and DTUs provide low temporal resolution. μ PMUs alone cannot capture all necessary system information. Additionally, the asynchronous nature and heterogeneous update intervals of these devices complicate real-time estimation. This situation necessitates advanced data fusion techniques that can handle measurement heterogeneity effectively.

Traditional data imputation methods such as linear interpo-

Manuscript received: April 7, 2025; revised: July 25, 2025; accepted: November 25, 2025. Date of CrossCheck: November 25, 2025. Date of online publication: February 11, 2026.

This work was supported by the National Natural Science Foundation of China (No. U24B2081).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Y. Zhang, Q. Chen, H. Liu, and Q. Hu (corresponding author) are with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: yuanshizhang@seu.edu.cn; 220232967@seu.edu.cn; haizhou501@seu.edu.cn; qhu@seu.edu.cn).

B. Zhai is with the State Grid Jibei Electric Power Company, Beijing 100052, China (e-mail: zhaibingxu@126.com).

DOI: 10.35833/MPCE.2025.000259



lation (LI) [12], historical average (HA) [13], and optimally weighted averaging (OWA) provide baseline solutions but cannot accurately track fast system changes or nonlinear power flow behavior. More expressive sequence models, especially long short-term memory (LSTM) networks, have therefore been introduced to learn temporal correlations in smart grid measurements [14]-[17]. On the dynamic side, the state transition equation must represent multimodal operating regimes caused by intermittent renewable injections and time-varying loads [18]-[21]. Simple linear or quasi-linear models [22], [23] are insufficient for this purpose, whereas Gaussian mixture models (GMMs) [24] offer a compact way to represent multiple probabilistic modes within a unified framework.

These modeling choices require a numerically stable nonlinear filter. Extended Kalman filters (EKFs) and unscented Kalman filters (UKFs) [25]-[28] either rely on Jacobian linearization or incur high computational cost in large-scale systems. The cubature Kalman filter (CKF) [29] avoids explicit Jacobians and exhibits better numerical stability, and its integration with forecasting-aided state estimation (FASE) [30], [31] further strengthens predictive capability by exploiting historical and real-time measurements. Under the above condition, this paper presents a unified real-time FASE framework for distribution networks using an improved CKF, integrating multi-source data from DTU, RTU, and μ PMU. To handle asynchronous measurements, we comprehensively evaluate multiple imputation methods, including LI, HA, OWA, and LSTM-based method. Our extensive comparative analysis demonstrates that LSTM-based method is particularly effective in handling delays and capturing nonlinear dynamics in distribution network measurements. This paper also introduces the GMM to enhance the accuracy and robustness of state transition models.

The main contributions of this paper are as follows.

1) We develop a unified real-time FASE framework for distribution networks that integrates heterogeneous μ PMU, RTU, and DTU measurements through a hybrid multi-source measurement model. By combining forecasting information with multi-rate data fusion, the framework enables real-time state estimation under asynchronous and noisy measurements.

2) We design an LSTM-based data imputation method to effectively align various types of measurement data, resulting in improved accuracy and reliability in state estimation.

3) We propose a GMM-based approach for distribution network state transition. This model provides a more accurate and robust representation of network operational states by capturing complex system dynamics, and it estimates an auxiliary state vector that aggregates voltages and currents at both measured and unmeasured nodes.

4) We improve the CKF by incorporating adaptive robust functions within the proposed FASE framework. This improved algorithm enables faster state estimation while preventing error accumulation, leading to more reliable network monitoring.

The remainder of this paper is organized as follows. Section II introduces the asynchronous data processing mechanism and the LSTM-based imputation method. Section III es-

tablishes the measurement equations and the GMM-based state transition model. Section IV details the improved CKF with adaptive robust functions. Section V integrates these components into the unified FASE framework and presents the overall implementation steps. Section VI validates the proposed framework and compares the performance of different imputation methods and estimation algorithms through case studies on IEEE 33-bus and 118-bus systems. Finally, Section VII concludes this paper.

II. ASYNCHRONOUS DATA PROCESSING MECHANISM AND LSTM-BASED IMPUTATION METHOD

The deployment of heterogeneous measurement devices in the distribution network leads to data asynchrony, which affects the accuracy and real-time performance of state estimation. This section discusses the multi-source data synchronization method, introduces the state estimation technique based on multi-source data fusion, and analyzes the advantages and disadvantages of traditional imputation methods and LSTM-based methods.

A. Multi-source Data Fusion and Synchronization

In modern distribution networks, multi-source measurements with different time scales and accuracies inevitably lead to data asynchrony that must be resolved before state estimation, as illustrated in Fig. 1. In the proposed FASE framework, multi-source measurement data fusion first synchronizes the data by imputing delayed RTU and DTU measurements to match μ PMU sampling rates, and then ensures consistent data quality by accounting for device accuracy and potential measurement errors across all sources.

B. Data Imputation Methods

Traditional data imputation methods have limitations in dealing with complex dynamics of modern distribution networks. To comprehensively evaluate the performance of LSTM-based data imputation method, we compare it against three well-established traditional data imputation methods based on the approach proposed in [32]-[35].

LI estimates missing values by connecting adjacent known data points with straight lines. For a missing value at time t between known values at time t_1 and t_2 , the interpolated value is calculated as:

$$\hat{x}_i^{\text{LI}} = x_h + \frac{(x_q - x_h)(t_i - t_h)}{t_q - t_h} \quad (1)$$

where $\hat{x}_i^{\text{LI}}$ is the linear interpolation estimate at time i ; x_h is the nearest known measurement before the missing value; x_q is the nearest known RTU measurement following the missing value; t_i is the missing measurement moment; and t_h and t_q are two adjacent valid sampling points surrounding the missing-measurement moment.

While computationally efficient, LI assumes linear relationship and fails to capture nonlinear patterns in power system dynamics.

HA replaces missing values with the average of historical authentic samples from nearby measurement devices as follows.

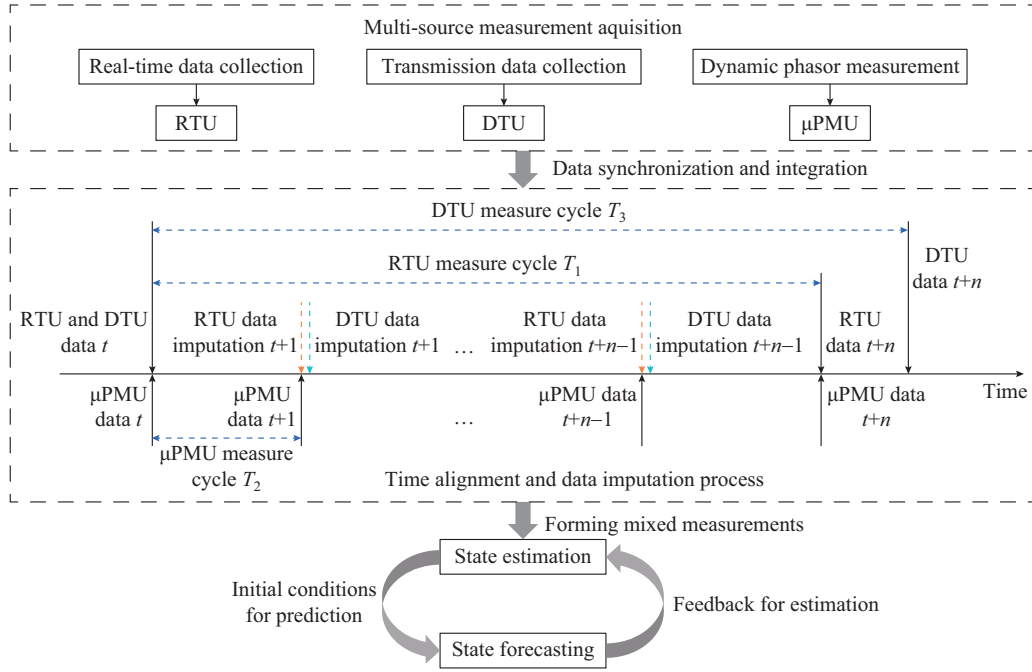


Fig. 1. State estimation architecture in a multi-source measurements data fusion environment.

$$\hat{x}_i^{\text{HA}} = \frac{1}{N_h} \sum_{q \in R} x_q \quad (2)$$

where $\hat{x}_i^{\text{HA}}$ is the historical average estimate at time i ; N_h is the number of historical samples used for averaging calculations; R denotes the historical sample collection; and x_q is the q^{th} measurement value in the historical sample collection.

HA works well for regular patterns but fails under abnormal conditions or rapid system changes.

OWA combines LI and HA methods through weighted averaging, providing good performance for both short and long periods:

$$\hat{x}_i^{\text{OWA}} = e^{-\alpha g_i} \hat{x}_i^{\text{LI}} + (1 - e^{-\alpha g_i}) \hat{x}_i^{\text{HA}} \quad (3)$$

where $\hat{x}_i^{\text{OWA}}$ is the optimally weighted averaging estimate at time i ; α is a positive weight parameter; and g_i is the distance from the estimated value to the nearest available sample. For small g_i , OWA relies mainly on LI values; for large g_i , it depends more on HA values.

The LSTM network effectively captures long-term dependencies in long sequence data by introducing memory units, and is able to model more complex nonlinear time dependencies, showing the potential for improving accuracy in distribution network state estimation applications. The architecture of the LSTM network for multi-source measurement data imputation, as shown in Fig. 2, together with its imputation process, is described as follows.

1) Forget gate

The forget gate is used to control how much information from the previous moment is retained in the current moment. The formula is given as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (4)$$

where t is the time step; f_t is the output of the forget gate; W_f is the weight matrix; h_{t-1} is the hidden state at the last moment; x_t is the input data at the current moment (includ-

ing RTU and DTU data); b_f is the bias term; and σ is the sigmoid activation function.

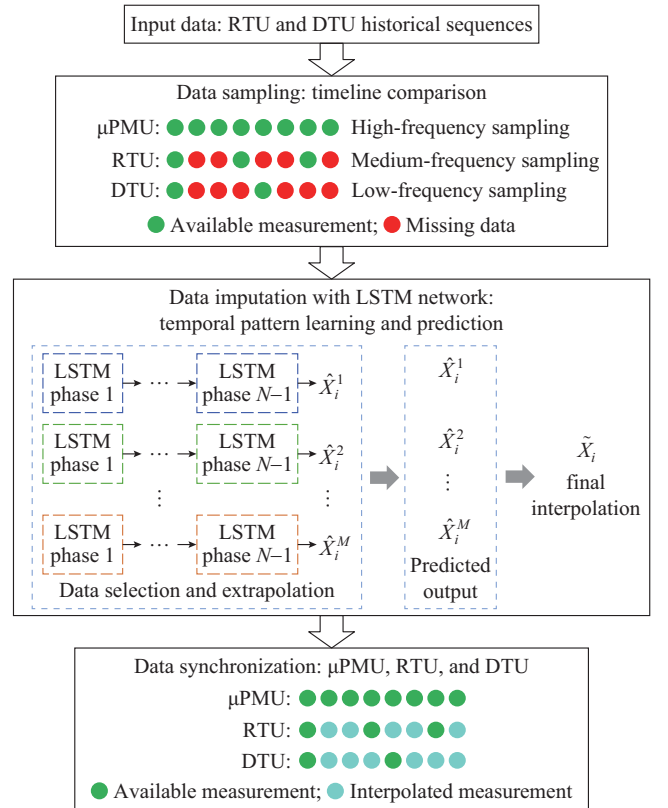


Fig. 2. Architecture of LSTM network for multi-source measurement data imputation.

2) Input gate

The input gate determines how much new information is written into the memory cell as:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (5)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \quad (6)$$

where $\mathbf{i}_t$ is the output vector of the input gate; $\tilde{\mathbf{C}}_t$ is the candidate memory cellstate; $\mathbf{W}_i$ and $\mathbf{W}_c$ are the weight matrices; and $\mathbf{b}_i$ and $\mathbf{b}_c$ are the bias terms.

Next, the state of memory cell is updated, where the current state of the memory cell is determined by the state at the previous moment and the new candidate state as:

$$\mathbf{C}_t = f_i \mathbf{C}_{t-1} + \mathbf{i}_t \tilde{\mathbf{C}}_t \quad (7)$$

where $\mathbf{C}_{t-1}$ is the state of the memory cell at the previous moment.

3) Output gate

The output gate determines the effect of the memory cell state on the hidden state. The specific calculation is given as:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad (8)$$

where $\mathbf{o}_t$ is the output vector of the output gate, which controls the information passed to the hidden state; $\mathbf{W}_o$ is the weight matrix associated with the input vector; and $\mathbf{b}_o$ is the bias term corresponding to the output gate.

The hidden state at the current moment $\mathbf{h}_t$ is given as:

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{C}_t) \quad (9)$$

4) Data imputation

The LSTM model learns time dependencies in RTU and DTU data. It takes historical RTU and DTU sequences as input and generates imputed values as outputs to synchronize them with μ PMU data.

Assuming N time steps of data, the input sequence is defined as:

$$\mathbf{X}_t = [\mathbf{x}_{t-N+1}, \mathbf{x}_{t-N+2}, \dots, \mathbf{x}_t] \quad (10)$$

where $\mathbf{x}_t$ is the feature vector at time step t containing all relevant measurements.

The LSTM-based training dataset combines historical RTU and DTU measurements from practical distribution networks and simulated data from IEEE test feeders. The measurements over a one-year period, including voltages, active/reactive power and currents at representative nodes, are cleaned by outlier removal and normalization, and then converted into sliding-window sequences of length $N=24$ (6 hours at 15-min intervals). For each measurement type, approximately 50000 sequences are obtained. Seventy percent-age of the sequences are for training, 15% for validation, and 15% for testing, enabling the LSTM model to learn both short-term fluctuations and longer-term temporal dependencies in RTU and DTU data.

After training, the LSTM model can be used for RTU and DTU data imputation. The specific process is as follows: historical RTU and DTU are input data into the trained LSTM model, and the model will output the predicted value of the next time step.

$$\hat{\mathbf{y}}_{t+1} = \mathbf{h}_t \quad (11)$$

where $\hat{\mathbf{y}}_{t+1}$ is the actual measured value of the next time step predicted by the LSTM model. This predicted value will be used as an imputed value to fill in the missing parts in the

RTU and DTU data to synchronize them with the μ PMU data.

In addition, in order to measure the difference between the predicted value and the true value of the LSTM model, mean squared error (MSE) is used as the loss function, expressed as:

$$MSE = \frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2 \quad (12)$$

where T is the time series sample length used in the statistics; y_t is the predicted value of LSTM model; and $\hat{y}_t$ is the actual measured value.

III. MEASUREMENT EQUATIONS AND GMM-BASED STATE TRANSITION MODEL

For distribution network state estimation, this section constructs a unified state transition model to integrate μ PMU, RTU, and DTU data with the objective of estimating the auxiliary state defined above. The quasi-steady-state characteristics of the distribution network are used to simplify the system dynamics, and a GMM is introduced to improve the robustness and accuracy of auxiliary state estimation. In this study, we define the auxiliary state as the complete set of system state variables that characterize the electrical condition of the distribution network at any given time instant. The auxiliary state serves as the primary target variable that the proposed FASE framework seeks to determine from available multi-source measurements.

A. μ PMU, RTU, and DTU Measurement Equations

For nodes equipped with μ PMU, the measurement data are initially provided in polar coordinates. These data must be converted into rectangular coordinates using appropriate coordinate transformation formula, as illustrated in (13).

$$\begin{cases} V \cos(\theta_v) + jV \sin(\theta_v) = V_\mu^r + jV_\mu^x \\ I \cos(\theta_v) + jI \sin(\theta_v) = I_\mu^r + jI_\mu^x \end{cases} \quad (13)$$

where V is the magnitude of the voltage at the node equipped with μ PMU; I is the magnitude of the current injected into the node equipped with μ PMU; r denotes the real part of the branch current; x denotes the imaginary part of the branch current; θ_v is the voltage phase angle of the node equipped with μ PMU; V_μ^r and V_μ^x are the real and imaginary parts of the voltage at the node equipped with μ PMU, respectively; and I_μ^r and I_μ^x are the real and imaginary parts of the current injected into the node equipped with μ PMU, respectively.

The coordinate transformation from polar to rectangular coordinates introduces correlation in the measurement errors. When μ PMU measurements are originally obtained in polar form with independent errors in magnitude and phase, the transformation to rectangular coordinates results in correlated errors between the real and imaginary components.

For a voltage measurement with magnitude and phase, assuming independent Gaussian errors, the covariance matrix in rectangular coordinates is expressed as:

$$\mathbf{R}_\mu = \mathbf{J}^T \mathbf{R}_\mu^p \mathbf{J} \quad (14)$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial V_\mu^r}{\partial V} & \frac{\partial V_\mu^r}{\partial \theta_V} \\ \frac{\partial V_\mu^x}{\partial V} & \frac{\partial V_\mu^x}{\partial \theta_V} \end{bmatrix} \quad (15)$$

$$\mathbf{R}_\mu^p = \text{diag}(\sigma_V^2, \sigma_{\theta_V}^2) \quad (16)$$

where $\mathbf{J}$ is the Jacobian matrix of the transformation; and $\mathbf{R}_\mu^p$ is the diagonal covariance matrix in polar coordinates; σ_V^2 is the variance of voltage magnitude measurement error; and $\sigma_{\theta_V}^2$ is variance of voltage angle measurement error.

For the three-phase power from RTU and DTU, the active power P_R , P_D and reactive power Q_R , Q_D measured by RTU and DTU can be expressed as:

$$\begin{cases} P_R = V_R I_R \cos(\phi_R) \\ Q_R = V_R I_R \sin(\phi_R) \\ P_D = V_D I_D \cos(\phi_D) \\ Q_D = V_D I_D \sin(\phi_D) \end{cases} \quad (17)$$

where V_R and V_D are the voltage amplitudes measured by RTU and DTU, respectively; and I_R and I_D are the current amplitudes measured by RTU and DTU, respectively.

The real and imaginary parts of the equivalent injected current of RTU and DTU are $I_R^{r(eq)}$, $I_D^{r(eq)}$ and $I_R^{i(eq)}$, $I_D^{i(eq)}$, respectively:

$$\begin{cases} I_R^{r(eq)} = \frac{I_R}{V_R} \cos(\phi_R) V^r + \frac{I_R}{V_R} \sin(\phi_R) V^i \\ I_R^{i(eq)} = \frac{I_R}{V_R} \cos(\phi_R) V^i + \frac{I_R}{V_R} \sin(\phi_R) V^r \\ I_D^{r(eq)} = \frac{I_D}{V_D} \cos(\phi_D) V^r + \frac{I_D}{V_D} \sin(\phi_D) V^i \\ I_D^{i(eq)} = \frac{I_D}{V_D} \cos(\phi_D) V^i + \frac{I_D}{V_D} \sin(\phi_D) V^r \end{cases} \quad (18)$$

where V^r and V^i are the real and imaginary parts of voltage, respectively; and ϕ_R and ϕ_D are the phase angles measured by RTU and DTU, respectively.

B. Hybrid Measurement Equation of Equivalent Transformation

All measured values are transformed in accordance with (19) and (20) to obtain a hybrid measurement equation as:

$$\mathbf{E}_M = [\mathbf{V} \quad \mathbf{I} \quad \mathbf{I}']^T \quad (19)$$

$$\begin{cases} \mathbf{V} = \begin{bmatrix} V_\mu^r & V_R^{r(eq)} & V_D^{r(eq)} \\ V_\mu^x & V_R^{x(eq)} & V_D^{x(eq)} \end{bmatrix} \\ \mathbf{I} = \begin{bmatrix} I_\mu^r \\ I_\mu^x \end{bmatrix} \\ \mathbf{I}' = \begin{bmatrix} I_R^{r(eq)} \\ I_R^{x(eq)} \\ I_D^{r(eq)} \\ I_D^{x(eq)} \end{bmatrix} \end{cases} \quad (20)$$

where $\mathbf{E}_M$ is the hybrid measurement vector; $\mathbf{V}$ is the node voltage matrix; $\mathbf{I}$ is the branch current vector; and $\mathbf{I}'$ is the

injected current vector.

C. GMM-based State Transition Model

In this subsection, the state variable is represented as:

$$\mathbf{x} = [I_{ij,l}^r \quad I_{ij,l}^x] \quad (21)$$

where $\mathbf{x}$ is an n -dimensional state variable; ij is the branch from node i to node j ; and l is the phase (Phase A, B or C) of the distribution network.

GMM is a powerful probabilistic model that can effectively capture the multimodal characteristics of complex systems [36] - [37]. By introducing GMM, the transition process of the system state can be modeled more flexibly, thereby improving the accuracy and robustness of state estimation.

The auxiliary state serves as the fundamental variable in our GMM-based state transition model. The transition process models how the auxiliary state evolves over time under different operational modes, enabling the proposed FASE framework to capture complex system dynamics including normal operation, renewable energy fluctuations, and varying load conditions. GMM assumes that the transition of system state is composed of a weighted combination of multiple Gaussian distributions, each corresponding to a potential mode of system state. The probability density function of GMM can be expressed as:

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k) = \sum_{i=1}^g \pi_i \mathcal{N}(\mathbf{x}_{k+1} | \mathbf{A}_i \mathbf{x}_k + \mathbf{b}_i, \boldsymbol{\Sigma}_i) \quad (22)$$

where $\mathbf{x}_k$ is the state vector of the system at time step k ; $\mathbf{A}_i$ and $\mathbf{b}_i$ are the linear transformation matrix and bias term of the i^{th} Gaussian distribution, respectively, describing the linear transition of the state from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$; $\boldsymbol{\Sigma}_i$ is the covariance matrix of the i^{th} Gaussian distribution, which describes the randomness of state transition; π_i is the mixing coefficient of the i^{th} Gaussian distribution, satisfying $\sum_{i=1}^g \pi_i = 1$, and indicates the weight of each mode; $\mathcal{N}(\cdot)$ is the Gaussian distribution function; and g is the number of mixed Gaussian distributions.

The GMM parameters are updated in real time to adapt to changing network conditions. This dynamic updating mechanism is crucial for maintaining accurate state estimation as the distribution network experiences varying operational modes.

The parameters $\{\pi_i, \mathbf{A}_i, \mathbf{b}_i, \boldsymbol{\Sigma}_i\}$ are updated using a modified online expectation-maximization (EM) algorithm as follows:

1) Expectation step: at each time step k , the responsibility of each Gaussian component is calculated to determine how much each component contributes to explaining the current state transition, expressed as:

$$\gamma_{i,k} = \frac{\pi_{i,k-1} \mathcal{N}(\mathbf{x}_{k+1} | \mathbf{A}_{i,k-1} \mathbf{x}_{k-1} + \mathbf{b}_{i,k-1}, \boldsymbol{\Sigma}_{i,k-1})}{\sum_j \pi_{j,k-1} \mathcal{N}(\mathbf{x}_{k+1} | \mathbf{A}_{j,k-1} \mathbf{x}_{k-1} + \mathbf{b}_{j,k-1}, \boldsymbol{\Sigma}_{j,k-1})} \quad (23)$$

where $\gamma_{i,k}$ is the responsibility of the i^{th} Gaussian component for the current observation $\mathbf{x}_k$.

2) Maximization step: the mixing coefficients are updated using an exponential moving average:

$$\pi_{i,k} = \lambda \pi_{i,k-1} + (1-\lambda) \gamma_{i,k} \quad (24)$$

where λ is the forgetting factor, which controls the balance between historical information and current observations. This formulation allows the model to adapt to changing operational conditions while maintaining stability.

The entries of the transition matrix $\mathbf{A}_i$ are estimated by solving a weighted least squares problem as:

$$\mathbf{A}_{i,k} = \min \sum_{t=1}^k \tau^{k-t} \gamma_{i,t} \|\mathbf{x}_{t+1} - \mathbf{A}_i \mathbf{x}_t - \mathbf{b}_i\|^2 \quad (25)$$

where τ is the temporal decay weight; $\mathbf{b}_i$ is the bias vector. The objective function minimizes the weighted prediction errors across all historical data points, with weights determined by both temporal decay factor and component responsibility.

The GMM enables multimodal representation where different components capture distinct operational behaviors such as normal operation, renewable energy fluctuations, and load variations. The real-time adaptation occurs as new measurement data arrive, with responsibility values continuously calculated and mixing coefficients updated accordingly, thus enabling automatic switching between operational modes as network conditions evolve.

IV. IMPROVED CKF WITH ADAPTIVE ROBUST FUNCTIONS

CKF is applied to address challenges in distribution network state estimation by effectively handling nonlinear systems with better numerical stability and accuracy than EKF and UKF. To counter measurement noise and outliers, an adaptive robust function is integrated, which dynamically adjusts data point weights to reduce outlier influence. The process involves generating cubature points, propagating them through nonlinear state transitions with adaptive weighting, making measurement predictions, and applying the Kalman gain to update state and covariance. This approach significantly improves estimation accuracy and robustness under nonlinear and uncertain conditions in distribution networks.

A. CKF

In state estimation for nonlinear systems, traditional Kalman filters and their extensions like EKF and UKF often struggle with accuracy and complexity challenges. CKF offers an effective alternative by employing numerical integration techniques. The core innovation of CKF is using third-order spherical-radial cubature integration rules to calculate nonlinear integrals during prediction and update phases. Rather than relying on Taylor expansion linearization like the EKF, CKF directly utilizes cubature sample points to estimate nonlinear function integrals, resulting in superior filtering accuracy.

Core steps of CKF consist of two key procedures: state predicting and state updating, where the core operation of state updating is to update the state estimates and their covariance matrix.

1) State Predicting

The state predicting is one of the core links of the CKF. This process requires transferring the state information of the previous moment to the current moment while consider-

ing the nonlinear characteristics of the system.

$$\mathbf{x}_{j,k+1|k} = f(\boldsymbol{\xi}_{j,k}, \mathbf{u}_k) \quad (26)$$

where $\mathbf{x}_{j,k+1|k}$ is the sample point that forms a new sample point after propagation through the state transfer function; $\boldsymbol{\xi}_{j,k}$ is the j^{th} cubic sampling point; $\mathbf{u}_k$ is the control input vector, representing the external influence or control action applied to the system at time step k ; and $f(\cdot)$ is the GMM state transition function.

Finally, we need to calculate the predicted state value of the system at the current time step k based on the propagated sampling points and its uncertainty in the form of a covariance matrix.

$$\hat{\mathbf{x}}_{k+1|k} = \sum_{j=1}^{2N} \frac{1}{2N} \mathbf{x}_{j,k+1|k} \quad (27)$$

$$\mathbf{P}_{k+1|k} = \sum_{j=1}^{2N} \frac{1}{2N} (\mathbf{x}_{j,k+1|k} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{j,k+1|k} - \hat{\mathbf{x}}_{k+1|k})^T + \mathbf{Q}_k \quad (28)$$

where $\hat{\mathbf{x}}_{k+1|k}$ is the state prediction matrix for the next time step $k+1$; $\mathbf{P}_{k+1|k}$ is the prediction covariance; $\mathbf{Q}_k$ is the process noise covariance matrix at time step k ; and N is the number of state variables in the system.

2) State Updating

The state updating step is crucial for incorporating new measurement information to refine our state estimates.

First, the filter propagates each sample point through the measurement function to get predicted measurement value, expressed as:

$$\boldsymbol{\zeta}_{j,k} = h(\mathbf{x}_{j,k+1|k}) \quad (29)$$

This transforms state estimates into the measurement space, allowing us to compare predictions with actual measurements.

These individual predictions are then combined to calculate the overall predicted measurement value as:

$$\hat{\mathbf{z}}_{k+1} = \sum_{j=1}^{2N} \frac{1}{2N} \boldsymbol{\zeta}_{j,k} \quad (30)$$

Next, we need to assess the uncertainty in our predictions by computing the measurement prediction covariance matrix as:

$$\mathbf{S}_{k+1} = \sum_{j=1}^{2N} \frac{1}{2N} (\boldsymbol{\zeta}_{j,k} - \hat{\mathbf{z}}_{k+1})(\boldsymbol{\zeta}_{j,k} - \hat{\mathbf{z}}_{k+1})^T + \mathbf{R}_k \quad (31)$$

$$\mathbf{R}_k = \text{diag}(\mathbf{R}_\mu, \mathbf{R}_r, \mathbf{R}_d) \quad (32)$$

where $\mathbf{R}_k$ is the measurement noise covariance matrix. $\mathbf{R}_r$ and $\mathbf{R}_d$ remain diagonal as their measurements are directly obtained in rectangular form. This covariance matrix accounts for both the spread in our predictions and the measurement noise.

To understand how state changes relate to measurement changes, we calculate the cross-covariance matrix and the Kalman gain, which determines how much we should trust our new measurements versus our predictions:

$$\mathbf{C}_{k+1} = \sum_{j=1}^{2N} \frac{1}{2N} (\mathbf{x}_{j,k+1|k} - \hat{\mathbf{x}}_{k+1|k})(\boldsymbol{\zeta}_{j,k} - \hat{\mathbf{z}}_{k+1})^T \quad (33)$$

$$\mathbf{K}_{k+1} = \mathbf{C}_{k+1} \mathbf{S}_{k+1}^{-1} \quad (34)$$

where C_{k+1} is the cross-covariance matrix between states and measurements; and K_{k+1} is the Kalman gain.

Finally, after the cross-covariance matrix and Kalman gain are obtained, the state estimates and covariance matrix are updated as:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} (z_{k+1} - \hat{z}_{k+1}) \quad (35)$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1} S_{k+1} K_{k+1}^T \quad (36)$$

where $\hat{x}_{k+1|k+1}$ is the result of updating the state estimate using the measurement value; and $P_{k+1|k+1}$ is the updated prediction covariance matrix.

B. Detailed Implementation of Improved CKF

The improved CKF enhances traditional CKF through systematic integration of adaptive robust functions in both prediction and correction steps. Unlike traditional CKF that treats all sample points and measurements equally, the improved CKF assigns dynamic weights based on their statistical reliability.

The improved CKF maintains the traditional CKF structure while introducing adaptive weighting mechanisms. For each cubature point χ_j , it calculates a reliability measure based on its distance from the predicted state as:

$$d_j = \|\chi_j - \hat{x}_k\|^2 \quad (37)$$

$$\omega_j = e^{-\frac{\beta d_j}{\sigma^2}} \quad (38)$$

where d_j is the square of the Euclidean distance between the j^{th} cubature point χ_j and the current state estimate $\hat{x}_k$; ω_j is the dynamic weight coefficient assigned to the j^{th} cubature point; β is the adaptation parameter; and σ^2 is the state uncertainty.

The specific improved CKF is shown in Algorithm 1, where $\omega_{r,j}$ is the robust weight for the j^{th} residual; r_j is the j^{th} residual vector; and R is a scalar normalization factor. The parameters β and γ are updated online using:

$$\beta_{k+1} = \alpha \beta_k + (1 - \alpha) \hat{\beta}_k \quad (39)$$

$$\gamma_{k+1} = \alpha \gamma_k + (1 - \alpha) \hat{\gamma}_k \quad (40)$$

where α is the forgetting factor; β_k is the adaptive parameters in the state prediction step; γ_k is the adaptive adjustment parameter for measurement residuals; and $\hat{\beta}_k$ and $\hat{\gamma}_k$ are estimated from measurement statistics.

C. Adaptive Robust Function

We propose an adaptive robust function that enhances the ability of CKF to handle abnormal data and noise by integrating it into both prediction and correction steps. The improved CKF offers better resistance to outliers and non-Gaussian noise, dynamically adjusts sample point and measurement weights for greater stability, and optimizes computational efficiency for high-dimensional state estimation tasks.

In the state predicting step, the adaptive robust function is used to adjust the weights of the sample points so that the sample points far away from the current estimated value have less influence on the prediction results. Specifically, an adaptive weight function $w(\xi_j)$ can be introduced, expressed as:

$$w(\xi_j) = \frac{1}{1 + v \|\xi_j - \hat{x}_{k|k}\|} \quad (41)$$

Algorithm 1: improved CKF

Initialization: set $\beta = 1.0$ and initialize covariance scaling factor $\alpha = 1.0$

1. **for** $j = 1$ to n
2. Generate cubature points: $\chi_{j,k} = \hat{x}_{k-1} + \sqrt{n S_{j,k-1}}$
3. Calculate adaptive weights based on (27) and (28)
4. Normalize weights: $\tilde{\omega}_j = \frac{\omega_j}{\sum_j \omega_j}$
5. Propagate through GMM based on (16)
6. Predict state: $\hat{x}_{j,k|k-1} = \sum_j \tilde{\omega}_j \chi_{j,k|k-1}$
7. Calculate prediction covariance matrix with adaptive weights
8. Update μ PMU covariance matrix: $R_\mu = J^T R_\mu^p J$, and update the corresponding block in R_k
9. Generate new cubature points and propagate through measurement function
10. Calculate measurement residuals: $r_j = z_k - h(\chi_{j,k|k-1})$
11. Apply robust weighting: $\omega_{r,j} = e^{-\frac{\gamma \|r_j\|^2}{R}}$
12. Update state: $\hat{x}_k = \hat{x}_{k|k-1} + K_k (z_k - \hat{z}_{k|k-1})$
13. **end for**

where v is a tuning parameter used to control the decay rate of the weight; ξ_j is the sample point of the current time step; and $\hat{x}_{k|k}$ is the estimated value of the current state.

After the sample point is propagated through the state transfer function, the state prediction and covariance matrices after weight adjustment can be expressed as:

$$\hat{x}_{k+1|k} = \sum_{j=1}^{2n} w(\xi_j) \chi_{j,k+1|k} \quad (42)$$

$$P_{k+1|k} = \sum_{j=1}^{2n} w(\xi_j) (\chi_{j,k+1|k} - \hat{x}_{k+1|k}) (\chi_{j,k+1|k} - \hat{x}_{k+1|k})^T + Q_k \quad (43)$$

While in the state correction step, an adaptive robust function is applied to the measurement residuals to further enhance the robustness against abnormal measurement values. Specifically, an adaptive weight is introduced to correct the measurement update process, expressed as:

$$w(z_{k+1}) = \frac{1}{1 + \gamma \|\mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1}\|} \quad (44)$$

where $\mathbf{z}_{k+1}$ is the actual measurement value.

The calculation formula of Kalman gain after taking weight into account is given as:

$$K_{k+1} = C_{k+1} S_{k+1}^{-1} w(z_{k+1}) \quad (45)$$

By integrating adaptive robust functions in CKF, the filter can dynamically adjust the influence weights of sample points and measurements, thereby effectively resisting the interference of outliers and noise. This method not only improves the accuracy of state prediction and correction, but also enhances the robustness of the filter when dealing with complex systems, especially suitable for scenarios with high-precision state estimation in non-ideal environments.

V. PROPOSED FASE FRAMEWORK WITH MULTI-SOURCE MEASUREMENT DATA FUSION

This section integrates the previously described imputation

method, transition model, and filtering algorithm into the proposed FASE framework for distribution networks. The proposed framework performs multi-source measurement data fusion by combining LSTM-based data imputation for asynchronous RTU and DTU measurements, the GMM-based state transition model, and the improved CKF with adaptive robust functions. In this way, synchronized multi-source measurements are processed to provide accurate and real-time state estimation while mitigating the impacts of data asynchrony, noise, and outliers.

A. Implementation of Proposed FASE Framework

To enhance the real-time performance and accuracy of FASE using measurement data from various terminals, this subsection details the implementation of the proposed FASE framework that leverages data imputation for asynchronous measurements. Initially, the RTU and DTU in Fig. 1 capture measurement values at time step t and $t+n$. Subsequently, a data imputation technique is applied to fill in the missing RTU and DTU measurements for the time interval from $t+1$ to $t+n-1$, thereby synchronizing these measurements within the same cycle as the μ PMU. Following this, a hybrid measurement FASE (H-FASE) model is developed using the interpolated data, allowing for a unified state estimation. The complete CKF is then implemented, which includes state prediction, measurement prediction, and filter correction. Specifically, H-FASE model concurrently collects data from μ PMU, RTU, and DTU at regular intervals, and CKF is utilized to perform state estimation throughout the entire process.

To estimate the auxiliary state of the distribution network, a state estimator based on CKF is implemented. This estimator integrates measurements from μ PMU, DTU, and RTU devices to infer the complete auxiliary state vector, including electrical quantities at both measured and unmeasured network locations. The auxiliary state estimation enables comprehensive network monitoring despite limited measurement infrastructure in distribution networks. The flowchart of the proposed FASE framework is depicted in Fig. 3.

B. Overall Estimation Performance Index

In order to measure the accuracy of the improved CKF, (46) and (47) are defined to represent the performance of the sampling time.

$$\sigma = \hat{z}_{m,t} - z_{m,y} \quad (46)$$

$$\delta_t = \sqrt{\frac{1}{N} \sum_{m=1}^N \sigma^2} \quad (47)$$

where $\hat{z}_{m,t}$ and $z_{m,y}$ are the estimated value and true value of the m^{th} state variable at time step t and y , respectively; σ is the absolute error (AE) of the state variable estimate; and δ_t is the root mean square error (RMSE) of the estimate at time step t .

The estimation function of the overall performance is defined as:

$$\delta = \frac{1}{T} \sum_{t=1}^T \delta_t \quad (48)$$

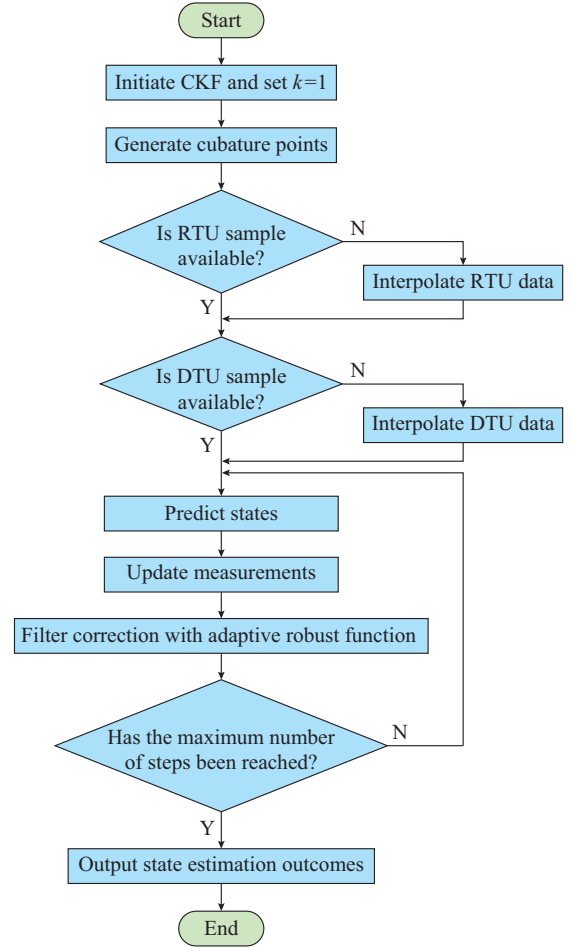


Fig. 3. Flowchart of proposed FASE framework.

where δ is equal to the RMSE for the multi-sampling period; and D is the total number of sampling instances within the simulation.

VI. CASE STUDY

This section evaluates the proposed FASE framework based on the improved CKF on the IEEE 33-bus and IEEE 118-bus systems. In the IEEE 33-bus system shown in Fig. 4, a 0.6 MW photovoltaic (PV) unit is connected at bus 28 and a 1.5 MW wind turbine (WT) is connected at bus 15; the base voltage and power are 4.8 kV and 1 MVA, respectively; and the PV, wind, and load power fluctuates by 30% within 30 s.

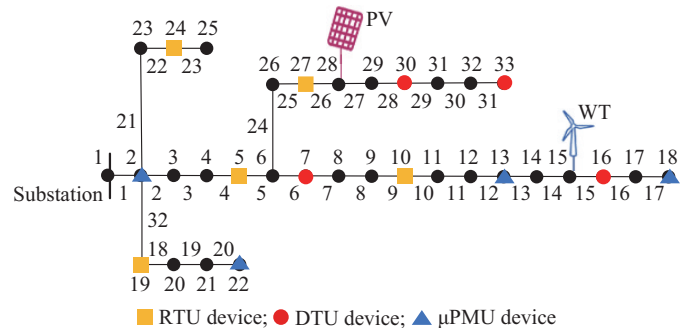


Fig. 4. IEEE 33-bus system.

The IEEE 118-bus system [38] is modified as a radial distribution configuration shown in Fig. 5, with a 2 MW PV unit and a 4 MW wind turbine at selected buses, a 10 kV reference voltage and 10 MVA base power, and all loads perturbed by $\pm 30\%$ to emulate large-scale dynamic operating conditions.

The research compares four data imputation methods (LI, HA, OWA, and LSTM), and evaluates state estimation accuracy using EKF, UKF, and the improved CKF.

In the IEEE 33-bus system, RTUs are deployed at several trunk buses to monitor voltages, currents, and power; DTUs are placed at representative load and distributed generation

buses; and μ PMUs are installed at key buses to provide time-synchronized phasor measurements. The IEEE 118-bus case follows the same deployment principle. To emulate realistic measurement conditions, independent Gaussian noise is added according to typical device accuracy: for μ PMUs, standard deviations of 0.1% and 0.2% for voltage and angle, respectively, while 0.2% and 0.3% for current magnitude and angle, respectively; for RTUs, 0.5% and 1.0% for voltage and power; and for DTUs, 1.0% and 1.5% for voltage and power, respectively. Noise and measurements are fed to the estimator, whereas the noise-free profiles serve as the ground truth for performance evaluation.

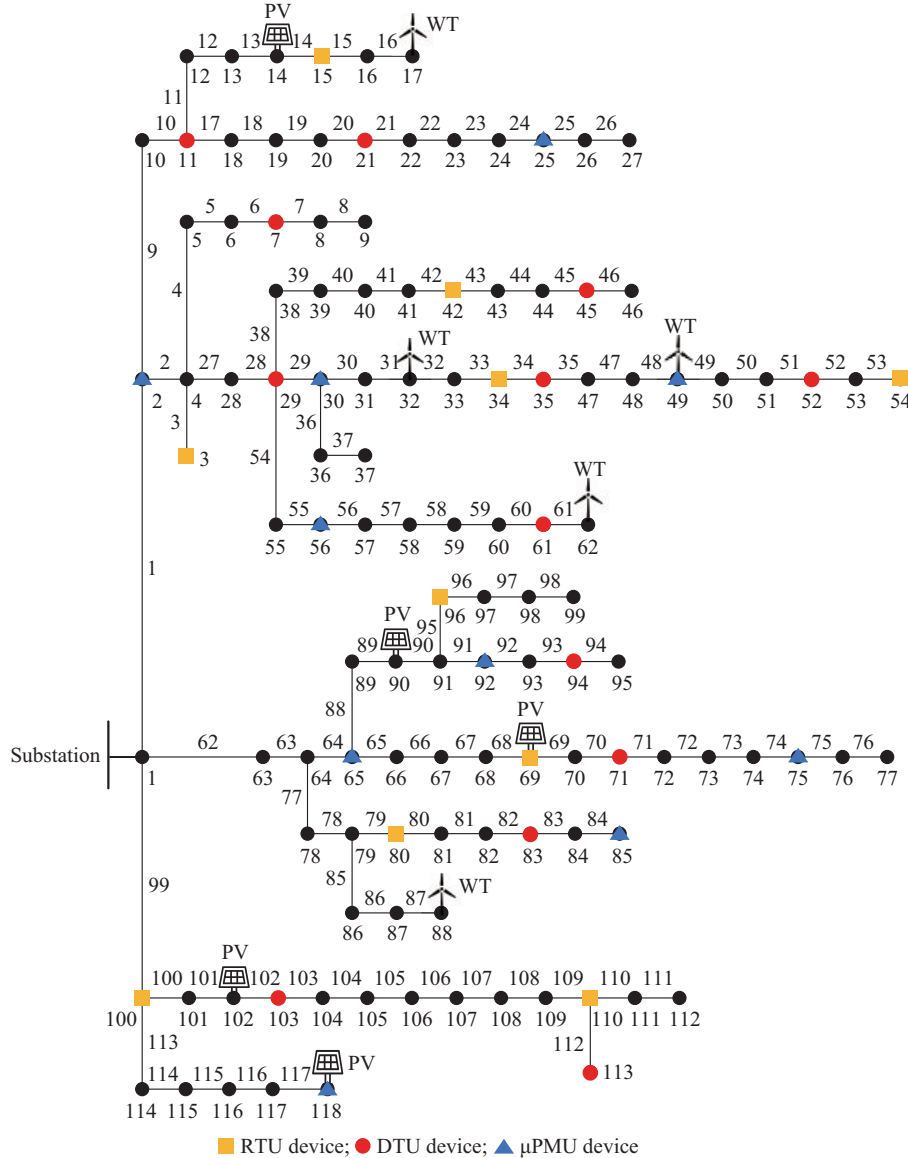


Fig. 5. IEEE 118-bus system.

A. Comparison of Different Data Imputation Methods

In Case A, four data imputation methods (LI, HA, OWA, and LSTM-based method) are compared in the IEEE 33-bus system using the maximum/minimum AE and RMSE at bus 3, bus 15, and all buses, as shown in Tables I and II. LSTM consistently achieves the smallest maximum AE and the low-

est RMSE across all locations, LI shows the largest errors, while OWA and HA fall between them.

Overall, LI performs worst because its linear assumption cannot capture nonlinear dynamics during rapid load or generation changes. HA benefits from historical patterns but deteriorates under abnormal conditions.

TABLE I
ESTIMATION PERFORMANCE OF DIFFERENT DATA IMPUTATION METHODS

Bus No.	The maximum AE				The minimum AE			
	LI	HA	OWA	LSTM-based	LI	HA	OWA	LSTM-based
3	9.0563×10^{-4}	6.5963×10^{-4}	2.1369×10^{-4}	8.6354×10^{-6}	1.9862×10^{-7}	3.5324×10^{-8}	9.6845×10^{-9}	2.1796×10^{-9}
15	8.5987×10^{-4}	5.4356×10^{-4}	2.0569×10^{-4}	6.3177×10^{-6}	1.7596×10^{-7}	2.6354×10^{-8}	9.3215×10^{-9}	1.2436×10^{-9}
All buses	2.8546×10^{-4}	2.3326×10^{-4}	1.9684×10^{-4}	3.4567×10^{-6}	9.0563×10^{-8}	7.6359×10^{-8}	1.2686×10^{-8}	1.8561×10^{-9}

TABLE II
RMSE OF DIFFERENT DATA IMPUTATION METHODS

Bus No.	RMSE			
	LI	HA	OWA	LSTM-based
3	8.5763×10^{-5}	2.4235×10^{-5}	9.6548×10^{-6}	5.3623×10^{-7}
15	3.8946×10^{-5}	1.9386×10^{-5}	9.3546×10^{-6}	3.2358×10^{-7}
All buses	3.9643×10^{-5}	2.3259×10^{-5}	1.6632×10^{-5}	4.6345×10^{-7}

OWA improves by adaptively combining LI and HA, yet it still relies on linear and historical assumptions. LSTM, in contrast, learns complex temporal dependencies directly from data and reduces RMSE by roughly one order of magnitude compared with the best traditional method, establishing deep learning interpolation as the preferred option for distribution network applications.

The comprehensive comparison of three traditional imputation methods (LI, HA, and OWA) reveals distinct performance characteristics in distribution network applications. LI demonstrates the poorest performance due to its fundamental limitation of assuming linear relationships between data points, making it inadequate for capturing the complex nonlinear dynamics inherent in power system measurements, particularly during rapid load or generation changes. HA provides moderate performance by effectively leveraging temporal patterns from previous periods, but suffers significant degradation during abnormal conditions or system disturbances when historical patterns fail to represent current system behavior. Among these methods, OWA achieves the best performance by intelligently combining linear interpolation and historical averaging through distance-based weighting schemes, though it remains constrained by its reliance on linear assumptions and historical data patterns.

LSTM consistently outperforms these traditional methods by effectively modeling complex temporal dependencies and nonlinear relationships that characterize distribution network measurements. The deep learning approach demonstrates superior capability in capturing both short-term fluctuations and long-term trends that traditional linear methods cannot adequately represent. While OWA provides the best traditional solution by balancing local spatial information with historical temporal patterns, the data-driven ability to learn intricate temporal patterns of LSTM yields substantially superior performance across all evaluation metrics, achieving RMSE improvements of approximately 85%-95% compared to OWA, thereby establishing deep learning approaches as the preferred solution for distribution network data imputation tasks.

B. Comparison of State Estimation Accuracy Using EKF, UKF, and improved CKF

Case B compares FASE implementations using EKF, UKF, and the improved CKF on the IEEE 33-bus system. Figures 6-9 plot the real and imaginary parts of phase currents at representative branches. EKF exhibits noticeable deviations in strongly nonlinear and fast-changing regions. UKF improves the tracking but still suffers from visible errors under rapid dynamics. The improved CKF maintains the closest agreement with the true trajectories during both steady and transient periods, confirming its superior accuracy and robustness for real-time distribution state estimation.

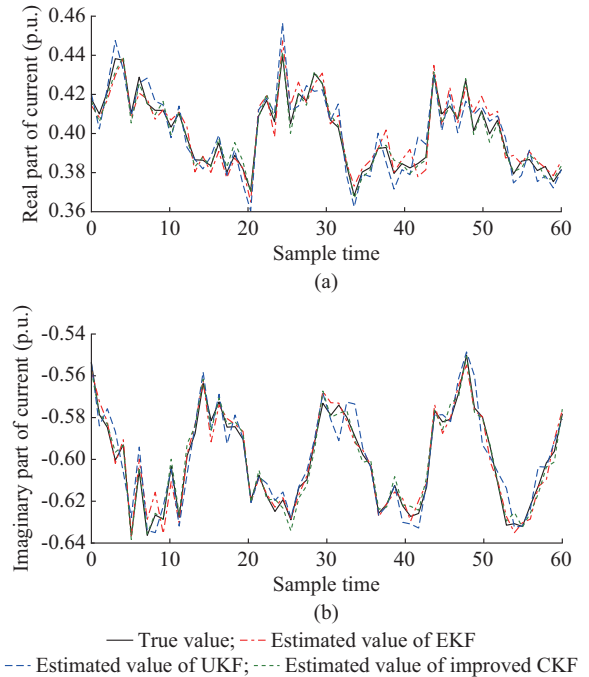


Fig. 6. Real and imaginary parts of current of Phase A at branch 20. (a) Real part. (b) Imaginary part.

In order to further verify the effectiveness of the proposed FASE framework in large-scale distribution networks, this paper conducts an extended verification on the IEEE 118-bus system. Figures 8 and 9 show the comparison of the estimation results of the real and imaginary parts of the current of Phased A at branch 32 and the current of Phase B at branch 62, respectively.

At branch 32, the estimated value of the improved CKF is highly consistent with the true value. Especially during the high-frequency fluctuation period of 0-20 sampling time, it can accurately track the current change trend.

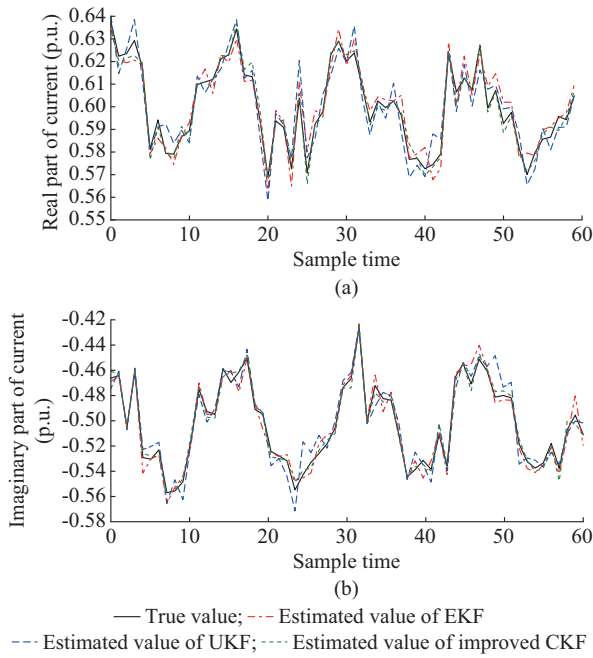


Fig. 7. Real and imaginary parts of current of Phase B at branch 15. (a) Real part. (b) Imaginary part.

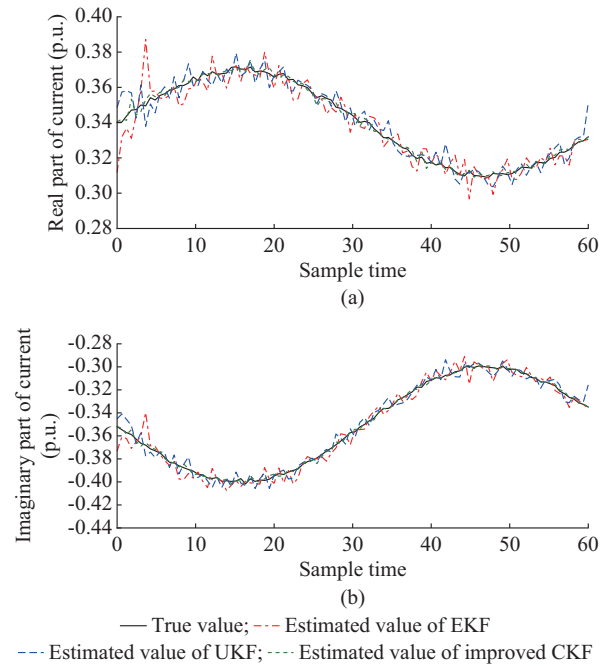


Fig. 9. Real and imaginary parts of current of Phase B at branch 62. (a) Real part. (b) Imaginary part.

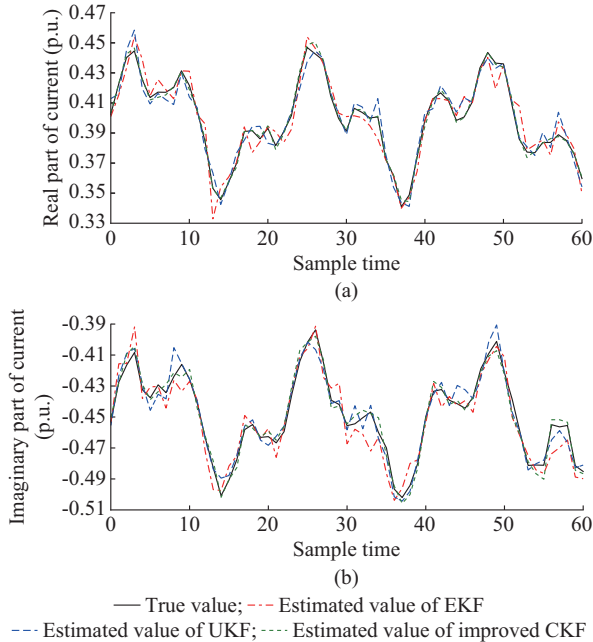


Fig. 8. Real and imaginary parts of current of Phase A at branch 32. (a) Real part. (b) Imaginary part.

In the imaginary part estimation, the improved CKF also shows excellent tracking accuracy, and the estimated curve almost coincides with the true value.

The branch 62 is far away from the distributed energy resources, so its current fluctuation shown in Fig. 9 is relatively small and shows obvious periodic characteristics. From the estimation results, it can be observed that the real part of the current of Phase B at branch 62 shows regular periodic changes in the range of 0.30–0.38 p.u.. The improved CKF can accurately capture this periodic pattern, and the estimation error is always kept within 0.003 p.u..

The imaginary part of estimation results also shows good periodic tracking ability. In the range of -0.28 to -0.42 p.u., the improved CKF accurately tracks the periodic fluctuations of the current.

C. Estimation Errors Under Different Algorithms

Figures 10 and 11 summarize the RMSE of real part of phase current under EKF, UKF, and the improved CKF on the IEEE 33-bus system and IEEE 118-bus system, respectively. The improved CKF achieves the lowest error and the fastest convergence, whereas EKF and UKF suffer from larger bias and fluctuations, especially under strong nonlinearities.

Figures 12 and 13 present the comparison of estimation performance between the improved CKF and traditional CKF on the IEEE 33-bus system and IEEE 118-bus system, respectively. As shown in the Fig. 12 and Fig. 13, the improved CKF maintains lower RMSE and smoother convergence than the traditional CKF, confirming that the adaptive weighting strategy for cubature points significantly enhances noise/outlier tolerance and scalability to large-scale distribution networks.

As a result, the superior performance of the improved CKF can be attributed to several key algorithmic enhancements. The adaptive robust function effectively reduces the influence of outliers and measurement noise by dynamically adjusting the weights of cubature points based on their statistical reliability.

The consistent performance improvement across both amplitude and phase angle measurements in all three phases validates the robustness and scalability of the improved CKF for practical deployment in monitoring applications of large-scale distribution networks.

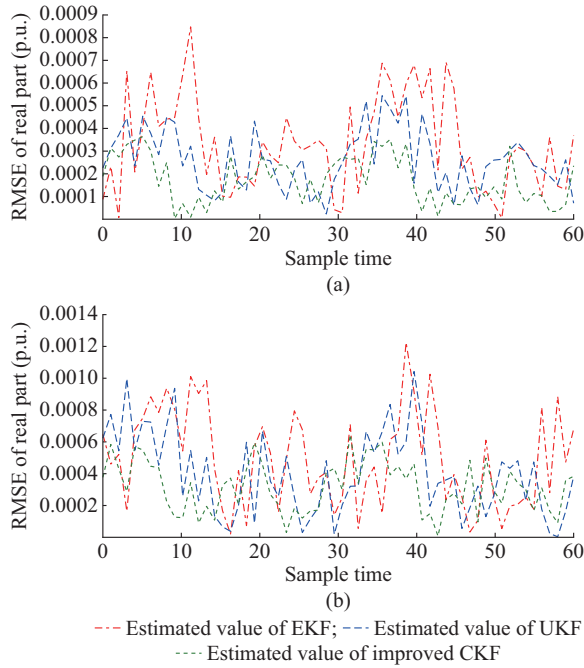


Fig. 10. RMSE of real part of phase current under EKF, UKF, and improved CKF on IEEE 33-bus system. (a) RMSE of real part of Phase A current at branch 20. (b) RMSE of real part of Phase B current at branch 15.

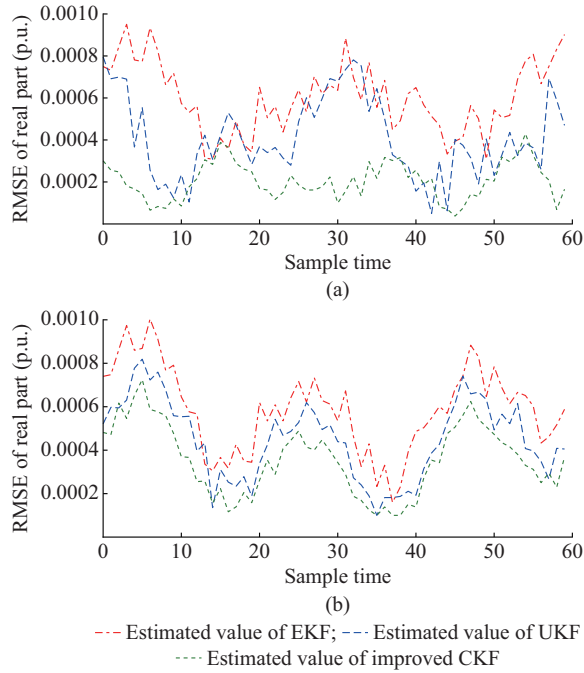


Fig. 11. RMSE of real part of phase current under EKF, UKF, and improved CKF on IEEE 118-bus system. (a) RMSE of real part of Phase A current at branch 32. (b) RMSE of real part of Phase B current at branch 62.

D. Computational Complexity and Real-time Performance Analysis

This subsection evaluates computational complexity and real-time performance. Table III breaks down the average processing time of each major component in one estimation cycle; the complete processing time of proposed FASE framework is within approximately 2-3 ms, satisfying real-time requirements.

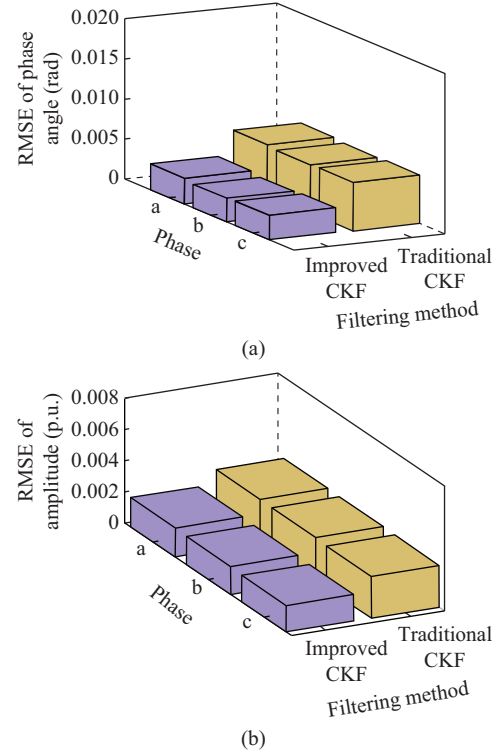


Fig. 12. Estimation performance between improved CKF and traditional CKF on IEEE 33-bus system. (a) RMSE of phase angle. (b) RMSE of amplitude.

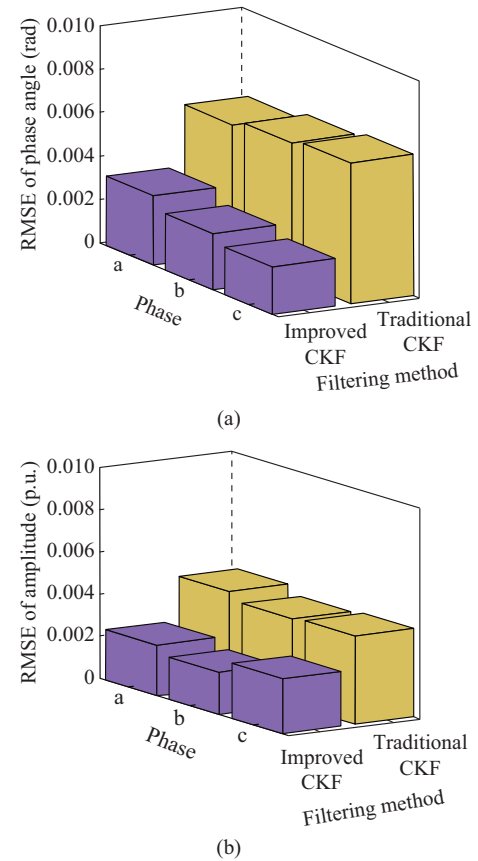


Fig. 13. Estimation performance between improved CKF and traditional CKF on IEEE 118-bus system. (a) RMSE of phase angle. (b) RMSE of amplitude.

Table IV further compares computation time with traditional UKF and EKF and shows that the proposed FASE framework preserves estimation accuracy while keeping the growth of computation cost moderate as the system size increases.

TABLE III
PROCESSING TIME OF PROPOSED FASE FRAMEWORK

Component	Processing time (ms) for system	
	IEEE 33-bus	IEEE 118-bus
LSTM data imputation	20.5	53.4
GMM state transition	17.6	40.7
Improved CKF	25.2	48.6
Total framework	63.3	142.7

TABLE IV
COMPARISON OF COMPUTATION TIME

System	Computation time (ms)		
	Proposed	UKF	EKF
IEEE 33-bus	63.3	76.8	95.7
IEEE 118-bus	142.7	168.4	186.6

VII. CONCLUSION

This paper proposes a unified real-time FASE framework with multi-source measurement data fusion that jointly addresses data asynchrony, measurement heterogeneity, and nonlinear dynamics in distribution networks. The LSTM-based data imputation method cuts RTU and DTU alignment error by 85% versus linear interpolation; a GMM-based state transition model with online EM algorithm enables automatic mode switching; and the improved CKF incorporates adaptive robust functions through dynamic cubature point weighting and robust measurement updates, achieving better computational efficiency compared to UKF.

Simulation results in IEEE 33-bus and 118-bus systems demonstrate superior performance with RMSE values consistently below 0.0004 p.u., 95% operational mode capture accuracy, better noise resistance than traditional CKF, and real-time execution of 2.3 ms per iteration. Current validation emphasizes normal operation with Gaussian noise and largely complete measurements; while the behavior of the proposed FASE framework under gross errors, abrupt events, and severe data loss remains to be fully assessed. Future work will stress-test resilience to bad data, cyber-physical attacks, and measurement failures, evaluate responses to reconfigurations and fault transients, and develop fast adaptive mechanisms for detection and mitigation. Additional efforts will focus on parallelized implementations for sub-millisecond cycles, deep reinforcement learning hyperparameter tuning, and hierarchical GMMs for multi-timescale modeling.

REFERENCES

- [1] K. Dehghanpour, Z. Wang, J. Wang *et al.*, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312-2322, Mar. 2019.
- [2] D. Xu, Z. Wu, J. Xu *et al.*, "A multiarea forecasting-aided state estimation strategy for unbalance distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 806-814, Jan. 2024.
- [3] F. Bai, Y. Cui, R. Yan *et al.*, "Condition monitoring of distribution power network assets using PMU measurements," *IEEE Transactions on Industry Applications*, vol. 60, no. 3, pp. 4642-4653, May 2024.
- [4] A. Shahid, F. Plaum, T. Korotko *et al.*, "AI technologies and their applications in small-scale electric power systems," *IEEE Access*, vol. 12, pp. 109984-110001, Aug. 2024.
- [5] N. Kanagaraj, M. Vijayakumar, M. Ramasamy *et al.*, "Energy management and power quality improvement of hybrid renewable energy generation system using coordinated control scheme," *IEEE Access*, vol. 11, pp. 93254-93267, Jul. 2023.
- [6] D. Sodin, M. Smolnikar, U. Rudež *et al.*, "Precise PMU-based localization and classification of short-circuit faults in power distribution systems," *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp. 3262-3273, Oct. 2023.
- [7] J. Zhao, M. Netto, Z. Huang *et al.*, "Roles of dynamic state estimation in power system modeling, monitoring and operation," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2462-2472, Sept. 2020.
- [8] H. N. Nguyen, T. Lan-Phan and C. J. Song, "Generative adversarial network-based network intrusion detection system for supervisory control and data acquisition system," in *Proceedings of 2024 IEEE International Conference on Consumer Electronics-Asia*, Danang, Vietnam, Nov. 2024, pp. 1-3.
- [9] J. Li, H. Din, L. Li *et al.*, "Communication technology research and application of distributed DTU terminal," in *Proceedings of 2024 6th International Conference on Energy Systems and Electrical Power*, Wuhan, China, Jun. 2024, pp. 1083-1087.
- [10] M. Tsebia, H. Bentarzi, and A. Ratni, "Micro-grids integration using remote synchronisation based on micro-PMU," in *Proceedings of 2024 2nd International Conference on Electrical Engineering and Automatic Control*, Setif, Algeria, May 2024, pp. 1-6.
- [11] V. L. Srinivas and J. Wu, "Topology and parameter identification of distribution network using smart meter and μ PMU measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, p. 9004114, May 2022.
- [12] D. Liu, W. Gao, W. Xu *et al.*, "5-D seismic data interpolation by continuous representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, p. 5922211, Jul. 2024.
- [13] Y. Sun, J. Li, Y. Xu *et al.*, "Deep learning versus conventional methods for missing data imputation: a review and comparative study," *Expert Systems with Applications*, vol. 227, p. 120201, Oct. 2023.
- [14] Y. Raghuvamsi and K. Teeparthi, "Distribution system state estimation with transformer-Bi-LSTM-based imputation model for missing measurements," *Neural Computing and Applications*, vol. 36, no. 3, pp. 1295-1312, Jan. 2024.
- [15] X. Xia, J. Lin, Q. Jia *et al.*, "ETD-ConvLSTM: a deep learning approach for electricity theft detection in smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2553-2568, Apr. 2023.
- [16] Z. Gao, S. Hu, H. Sun *et al.*, "Dynamic state estimation of new energy power systems considering multi-level false data identification based on LSTM-CNN," *IEEE Access*, vol. 9, pp. 142411-142424, Oct. 2021.
- [17] P. Vedavalli and D. Ch, "A deep learning based data recovery approach for missing and erroneous data of IoT nodes," *Sensors*, vol. 23, no. 1, p. 170, Dec. 2022.
- [18] K. Dehghanpour, Z. Wang, J. Wang *et al.*, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312-2322, Sept. 2018.
- [19] J. Zhao, M. Netto, Z. Huang *et al.*, "Roles of dynamic state estimation in power system modeling, monitoring and operation," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2462-2472, Sept. 2020.
- [20] X. Fu, Z. Wei, H. Sun *et al.*, "Agri-energy-environment synergy-based distributed energy planning in rural areas," *IEEE Transactions on Smart Grid*, vol. 15, no. 4, pp. 3722-3738, Feb. 2024.
- [21] C. Fan, S. Nie, L. Xiao *et al.*, "A multi-stage ensemble model for power load forecasting based on decomposition, error factors, and multi-objective optimization algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109620, Jan. 2024.
- [22] Z. Zhang and Y. Ju, "An embedded consensus ADMM distribution algorithm based on outer approximation for improved robust state estimation of networked microgrids," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 4, pp. 1217-1226, Jul. 2024.
- [23] A. Ghosh, A. Honoré, and S. Chatterjee, "DANSE: data-driven nonlinear state estimation of model-free process in unsupervised learning setup," *IEEE Transactions on Signal Processing*, vol. 72, pp. 1824-1838, Mar. 2024.

- [24] Y. Gao, X. Xu, Z. Yan *et al.*, "Gaussian mixture model for multivariate wind power based on kernel density estimation and component number reduction," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 3, pp. 1853-1856, Jul. 2022.
- [25] M. I. Ribeiro. (2004, Feb.). Kalman and extended Kalman filters: concept, derivation and properties. [Online]. Available: <https://people.duke.edu/~hpgavin/SystemID/References/Ribeiro-KalmanFilter-2004.pdf>
- [26] K. Ampountolas, "The unscented Kalman filter for nonlinear parameter identification of adaptive cruise control systems," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 8, pp. 4094-4104, May 2023.
- [27] Y. Wang, Z. Yang, Y. Wang *et al.*, "Resilient dynamic state estimation for power system using cauchy-kernel-based maximum correntropy cubature Kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, p. 9002011, Apr. 2023.
- [28] X. Zhang, W. Yan, M. Huo *et al.*, "Robust interval state estimation for distribution systems considering pseudo-measurement interval prediction," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 1, pp. 179-188, Jan. 2024.
- [29] T. Zhang, W. Zhang, Q. Zhao *et al.*, "Forecasting-aided joint topology/state estimation for distribution systems via Bayesian non-parametric modeling," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1014-1029, Jan. 2024.
- [30] Y. Wang, Y. Sun, and V. Dinavahi, "Robust forecasting-aided state estimation for power system against uncertainties," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 691-702, Jan. 2020.
- [31] H. Zhao and B. Tian, "Robust power system forecasting-aided state estimation with generalized maximum mixture correntropy unscented Kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, p. 9002610, Mar. 2022.
- [32] Y. Peng, Z. Wu, W. Gu *et al.*, "Optimal micro-PMU placement for improving state estimation accuracy via mixed-integer semidefinite programming," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 2, pp. 468-478, Mar. 2023.
- [33] M. Fan, Y. Hu, X. Zhang *et al.*, "Short-term load forecasting for distribution network using decomposition with ensemble prediction," in *Proceedings of 2019 Chinese Automation Congress*, Hangzhou, China, Nov. 2019, pp. 152-157.
- [34] Y. Yi, S. Liu, Y. Zhang *et al.*, "Phase identification of low-voltage distribution network based on stepwise regression method," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 4, pp. 1224-1234, Jul. 2023.
- [35] D. Hou, Y. Sun, V. Dinavahi *et al.*, "Adaptive two-stage unscented Kalman filter for dynamic state estimation of synchronous generator under cyber attacks against measurements," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 5, pp. 1408-1418, Sept. 2024.
- [36] J. Zhu, Y. Miao, H. Dong *et al.*, "Short-term residential load forecasting based on K-shape clustering and domain adversarial transfer network," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 4, pp. 1239-1249, Jul. 2024.
- [37] X. Liu, X. Zeng, L. Yao *et al.*, "Power system state estimation based on fusion of WAMS/SCADA measurements: a survey," in *Proceedings of 2018 2nd IEEE Conference on Energy Internet and Energy System Integration*, Beijing, China, Oct. 2018, pp. 1-6.
- [38] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12-19, Feb. 2011.

Yuanshi Zhang received the B.Eng. and M.A.Sc. degrees from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively, and the Ph.D. degree from the University of British Columbia, Vancouver, Canada, in 2021. Since 2021, he has been an Associate Professor with the School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include power system optimization and control, high-voltage direct current (HVDC) system, and demand-side management.

Qirui Chen received the B.S. degree in electrical engineering from Nanjing Institute of Technology, Nanjing, China, in 2023. He is currently pursuing the M.S. degree in Southeast University. His research interests include optimization of smart terminal configuration of distribution network and state estimation.

Haizhou Liu received the B.Sc. degree in physics from Nanjing University, Nanjing, China, in 2019, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2024. He is currently a Lecturer at the School of Electrical Engineering, Southeast University, Nanjing, China. His research interest includes artificial intelligence (AI)-augmented decision-making framework across heterogeneous energy systems.

Qinran Hu received the B.S. degree in electrical engineering from the Chien-Shiung Wu College, Southeast University, Nanjing, China, in 2010, and the M.S. and Ph.D. degrees from the University of Tennessee at Knoxville, Knoxville, USA, in 2013 and 2015, respectively. He was a Postdoctoral Fellow with Harvard University, Cambridge, USA, from 2015 to 2018. He is currently a Professor with the School of Electrical Engineering, Southeast University. His research interests include power system optimization, demand aggregation, and virtual power plant.

Bingxu Zhai received the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2012. He is a Senior Expert of dispatch and operation at State Grid Jibei Electric Power Company, Beijing, China. His research interest includes operation and control of power system and renewable energy.