Probabilistic Wind Power Forecasting with Missing Data Tolerance: An End-to-End Nonparametric Approach

Zichao Meng, *Member, IEEE,* Ye Guo, *Senior Member, IEEE,* and Chenhao Zhao

***Abstract*—Missing data occurs due to sensor failures, communication issues, or temporary gaps in the measurement process, which may significantly reduce the performance of the wind power forecasting system. Therefore, an end-to-end nonparametric approach is proposed for probabilistic wind power forecasting (WPF) incorporating missing data imputation. The proposed method comprises both end-to-end training and online application procedures. In the end-to-end training phase, a deep learning-based nonparametric forecast model undergoes iterative processes involving imputation for missing data and model training with revised loss. In the online application phase, the trained forecast model is deployed online to provide multi-step-ahead probabilistic WPF through continuously imputating online observations. Compared with other state-of-the-art benchmarks, the advantages of the proposal include: 1) the proposed method is nonparametric, i.e., no hypotheses on distribution types are needed, and 2) the proposed end-to-end training process will automatically regulate the imputed value from the deep learning-based forecast model for higher probabilistic forecast performance, thereby mitigating the negative impact of missing observations. This approach leverages the advantages of the nonparametric method and the deep learning-based end-to-end structure. As a result, the proposed approach showcases an outstanding approximation capability for the future probability distribution of nonstationary wind power while simultaneously addressing missing values. Experiments validate that the proposed end-to-end nonparametric approach is more effective in mitigating the negative impact of data missingness on forecast performance compared to other representative two-phase methods integrated with standalone missing data imputation steps. Additionally, it outperforms its parametric end-to-end counterpart across various missing rate scenarios, especially in multi-step-ahead probabilistic forecasting.**

***Index Terms*—Probabilistic wind power forecasting, missing data imputation, end-to-end structure, quantile regression**

I. Introduction

The advancement of wind energy holds a key position in our endeavor toward achieving carbon neutrality. As of 2022, the worldwide installed capacity of wind energy has exceeded 898 GW, and additional growth is on the horizon [1]. The swift expansion of wind power generation resources introduces heightened uncertainties in the generation front for power systems [2]. Consequently, the implementation of robust wind power forecasting (WPF) techniques becomes paramount for dispatch centers. This empowers them to enhance decision-making and optimize energy scheduling for the future, effectively navigating the inherent variability of wind resources.

Machine learning methods [3]–[9] have gained widespread adoption in WPF, owing to their heightened flexibility and superior nonlinear approximation capabilities compared to curve-fitting methods [10], [11] and traditional statistical techniques [12], [13]. In the realm of machine learning-based WPF, two primary categories have emerged: point forecasting and probabilistic forecasting [14]. Point forecasting models, including support vector machines [3], extended polynomial networks [4], and Markov chains [5], provide single-point expectations. Probabilistic WPF approaches offer a more comprehensive set of information, such as quantiles, intervals, or densities, enhancing decision-making for power system operators [6]–[9], as opposed to point forecasts which offer only expectations [3]–[5]. Notable examples of probabilistic forecast models include model predictive control strategy [9], Gaussian processes [6], nonparametric Bayesian methods [7], and kernel density estimation [8]. Recent advancements have introduced deep learning architectures, with a focus on recurrent neural network (RNN)-based models [15]. Notably, the RNN [16] and bidirectional long short-term memory (LSTM) network with deep concatenated residual structure [17] have been employed for point WPF. Bayesian LSTM [18] and autoregressive RNN [19], [20] have contributed to probabilistic WPF. Moreover, advanced techniques such as ensemble convolutional neural networks (CNN) [21], deep mixture density networks [22], temporal attention networks [23], and Transformer-based models [24] have also been harnessed for probabilistic WPF applications.

An inherent limitation in existing WPF methods is that they are vulnerable to missing observations, which may frequently happen due to sensor failures and communication errors [25]. Notably, offshore wind farms, particularly those operating under complicated weather conditions, frequently experience communication network congestion or disruptions, resulting in varying degrees of data unavailability [26]. A comprehensive analysis of two-year-length data sets encompassing wind speed and power data from 30 European wind farms, with capacities ranging from 41.8 MW to 129.0 MW, was conducted in [27]. The study revealed that real wind power data typically features a median missing rate of 2.70%, with some sites experiencing missing levels as high as 36%. Particularly noteworthy is

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB0906000 and 2020YFB0906005. Part of this work has been presented on the Second International Conference on Cyber-energy Systems and Intelligent Energies in 2024 (ICCSIE 2024). Corresponding author: Ye Guo, e-mail: guo-ye@sz.tsinghua.edu.cn

Z. Meng is with the Smart Grid and Renewable Energy Lab, Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua University, Shenzhen 518055, China, and also with the Electrical Dispatch and Control Center, Guangdong Power Grid, Guangzhou 510220, China.

Y. Guo and C. Zhao are with the Smart Grid and Renewable Energy Lab, Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua University, Shenzhen 518055, China.

that in case where 11.65% of data was missing and directly dropped, a substantial 19% increase in the normalized mean absolute error [1] was observed compared to the case with a full training data set, i.e., the data set is complete, highlighting the severe negative impact of missing data on forecast performance [27]. In such data-missing situations, the performance of conventional deep learning-based methods may be significantly affected in both the training and application stages. In the training stage, substantial data loss in the historical data set may prevent the forecasting model from capturing temporal correlations and inherent uncertainties within the original data. In the application stage, incomplete input data also degrades the prediction performance of the model. Therefore, addressing missing data at both training and application stages is crucial for WPF, particularly when aiming to model the intricate probability distribution of wind power outputs.

There are few studies concerned with missing value estimation related to WPF. For example, paper [28] adopts linear interpolation (LI) to reconstruct various types of missing data collected from wind farms. Paper [27] utilizes a multiple imputation technique based on the Markov Chain Monte Carlo method for WPF considering missing data recovery. Despite the simplicity and efficiency of conventional statistical methods such as LI [27], [28], they may face challenges in adequately representing the intricate patterns, including inherent variability and unpredictability, within highly volatile time series [29]. More sophisticated machine learning-based approaches have been developed to achieve more accurate missing data restoration, leveraging their excellent nonlinear approximation capabilities. For instance, a k nearest neighbor (KNN)-based imputation and forecast algorithm combined with CNN and LSTM is designed for wind power generation forecasting [30]. Moreover, the deep learning-based autoencoder, initially utilized in recommender systems for feature extraction and data reconstruction in an unsupervised manner [31], is also applied for missing data imputation in the wind farm Supervisory Control And Data Acquisition (SCADA) system [32].

Outside the scope of WPF, various missing data imputation techniques have also been applied to other research fields. For example, Kalman filters and weighted moving average are employed for solar irradiance forecasting considering imputating missing values, demonstrating robustness in forecasting with different resolutions [28].

In [33], a random forest (RF)-based method is designed to handle missing data of mixed types in multiple biological data sets, which includes both continuous and discrete variables. In [34], autoreplicative RF is developed with computational efficiency for missing data imputation and shows effectiveness on small data sets. RF-based methods are verified as panaceas for imputing missing data, especially when data are highly skewed [35]. In [36], a modified KNN method is used to impute spatial and temporal missing traffic data with high computational efficiency. In [37], the KNN technique is used for missing value imputation in gene expression data, which helps an early detection of cancer. For autoencoder-based methods, they are applied to imputing missing values in electric load data in smart grids [38], food composition databases [39], industrial time-series data [40], and provide superior results. Additionally, in [41], a revised CNN with enhanced resolution perception is employed to impute the missing data in point solar photovoltaic (PV) generation forecasting.

While there have been several approaches for missing data imputation, existing methods may lack a specific focus on the unique challenges of missing data estimation in time-series forecasting. When applied to forecasting with missing data, these methods typically necessitate a separation of the imputation and forecast models into two distinct phases. In this setup, imputed data are treated as the observation in the prediction procedure, potentially leading to error propagation and reduced forecasting accuracy. To mitigate such errors propagated from imputation to forecasting, paper [42] proposed an end-to-end method for probabilistic PV forecasting based on autoregressive LSTM, where a single model concurrently performs both the imputation and probabilistic forecasting tasks. This study holds relevance to our concerns, given that the same autoregressive structure based on LSTM has demonstrated success in addressing probabilistic WPF in [43]. However, one critical issue in designing this algorithm in [42] is that it assumes the probability distribution of the renewable outputs (viewed as random variables) follows the Gaussian distribution. This strong assumption on the forecasted probability distribution may not hold because distribution types of nonstationary time series may change with time. Although our previous work [44] proposed a one-step-ahead nonparametric forecasting method considering missing data imputation, it mainly focused on the offline procedure without considering the imputation error and multi-step-ahead forecasting scenarios were not fully addressed.

In this paper, to deal with the missing data problem and realize high-performance probabilistic forecasts without requirements on any a priori hypotheses on the distribution types, an end-to-end nonparametric method is proposed for probabilistic WPF with missing data tolerance. Specifically, the core contributions are summarized as follows:

1) An end-to-end training approach is designed for a nonparametric forecast model with incomplete data sets. Therein, a continuous imputation process is implemented first to calculate the forecasted quantiles with the forecast model and impute the missing value in the wind power series based on the forecasted median. Then, a revised loss, involving pinball loss for probabilistic forecasting and mean square error loss for imputation, is formulated and minimized to train the forecast model. This model training process includes both imputation and forecasting, indicating an end-to-end manner.
2) An online application method for multi-step-ahead probabilistic WPF is proposed based on the trained forecast model facing missing data online. The online application method considers historical dynamic information in the wind power series and executes multi-step-ahead forecasts through the imputation process. This leverages the capability of the forecast model to continuously handle

[1]The normalized mean absolute error denotes the mean absolute WPF error divided by the capacity of the corresponding wind farm.

missing data, as wind power in future time spots can be viewed as missing observations.

The remainder of this paper is organized as follows. Section II presents the problem modeling and evaluation metrics. Section III introduces the methodology for probabilistic WPF with missing data tolerance. Section IV details simulations. Section V finally draws conclusions.

II. Problem Modeling and Evaluation Metrics

In this section, the problem modeling for a deep learning-based nonparametric probabilistic forecast model with missing data inputation is introduced first. Then, evaluation metrics for the probabilistic forecasted results are presented.

A. Problem Modeling

Probabilistic WPF with missing data aims to provide the future probability distribution of wind power outputs based on missing value imputation. The historical observations of wind power series at $\{t-\delta+1,\, t-\delta+2,\, \cdots,\, t-1,\, t\}$ with missing data can be denoted as

$$\tilde{\boldsymbol{P}}_t = [p_{t-\delta+1},\, p_{t-\delta+2},\, /,\, \cdots,\, /,\, p_{t-2},\, p_{t-1},\, /], \tag{1}$$

where p_i, $i \in [t-\delta+1,\, t]$, denotes the observation of wind power at time spot i and "/" denotes the observation is missing at the corresponding time spot. This incomplete vector $\tilde{\boldsymbol{P}}_t$ with no missing data will be imputed as a complete vector $\boldsymbol{P}_t$ before the forecast, which is defined as

$$\boldsymbol{P}_t = [p_{t-\delta+1},\, p_{t-\delta+2},\, \hat{p}_{t-\delta+3},\, \cdots,\, \hat{p}_{t-3},\, p_{t-2},\, p_{t-1},\, \hat{p}_t], \tag{2}$$

where $\hat{p}_*$ (denoting variable $\hat{p}$ with different subscripts) denotes an imputed result for the missing value of historical wind power data.

After missing data imputation, the probability distribution of wind power can be obtained via a function, i.e., the probabilistic forecast model. Given the advantages of nonparametric methods, which do not rely on any a priori hypotheses on distribution types, we adopt quantile regression [45], a widely used nonparametric approach, for modeling the probabilistic forecast model. Quantile regression estimates quantiles associated with different confidential levels of the conditional distribution of the forecasting target based on a regression model, i.e., the probabilistic forecast model in our method. This gives a more detailed view of the distribution, allowing for the modeling of different parts of the forecasted distribution (e.g., median, lower quartile, upper quartile, etc.). We model the probabilistic forecast model in quantile regression as a deep neural network (NN) with multiple layers considering the universal approximation capability of the deep NN. This deep learning-based forecast model maps historical observations to quantiles of the forecasting target as

$$p^{\alpha}_{t+l|t} = F(\boldsymbol{P}_t,\, \alpha;\, \theta),\, \alpha \in Q, \tag{3}$$

where α denotes a nominal proportion ranging from 0 to 1 and it specifies the quantile to be estimated, i.e., the α-th quantile $p^{\alpha}_{t+l|t}$. Q denotes the proportion set including all proportions α concerned. $F(\cdot;\, \theta)$ denotes the deep learning-based probabilistic forecast model parameterized by θ and θ include weights and bias for different layers of the deep NN. Here, the proportion set Q is $\{0.05,\, 0.10,\, ...,\, 0.90,\, 0.95\}$ where the element α, i.e., the nominal proportion for quantile, in Q ranges from 0.05 to 0.95 in 0.05 increments. For example, when α is 0.5, the corresponding quantile denotes the median, i.e., the 0.5-the quantile. If there is no missing data in $\boldsymbol{P}_t$, probabilistic WPF is implemented directly based on the historical observations without considering imputation.

The parameters θ of the probabilistic forecast model can be optimized via minimizing the pinball loss $\mathcal{L}$, which is formulated as

$$\mathcal{L} = \sum_{\alpha \in Q} \alpha \cdot \max(0, p_{t+l} - p^{\alpha}_{t+l|t}) + (1-\alpha) \cdot \max(0, p^{\alpha}_{t+l|t} - p_{t+l}), \tag{4}$$

where α is a proportion between 0 and 1, specifying the quantile to be estimated, i.e., the α-th quantile $p^{\alpha}_{t+l|t}$, with p_{t+l} being the observation. The pinball loss penalizes forecasting errors based on whether the forecasted quantile is below or above the observation. The proportion α or $(1-\alpha)$ determines the penalty weight for under- or over-forecasting, respectively:

1) When $p^{\alpha}_{t+l|t} < p_{t+l}$: The loss increases linearly with the distance between p_{t+l} and $p^{\alpha}_{t+l|t}$, weighted by α, as α proportion of forecasted values are expected to be below the observation.

2) When $p^{\alpha}_{t+l|t} \geq p_{t+l}$: The loss increases linearly with the distance between $p^{\alpha}_{t+l|t}$ and p_{t+l}, weighted by $(1-\alpha)$, as $(1-\alpha)$ proportion of forecasted values are expected to be above the observation.

By minimizing the pinball loss, the forecast model learns provide a forecast with an α probability of under forecasting the observation and a (1-α) probability of over forecasting the observation, which reflects real quantile values.

B. Evaluation Metrics

The forecasted quantiles are evaluated within the probabilistic forecast evaluation framework in [46].

1) Reliability: Evaluating the reliability of a probabilistic forecast model involves measuring the average deviations between expected and observed frequencies below the forecasted quantiles

$$\overline{A_l} = \frac{1}{I} \cdot \sum_{i=1}^{I} |\alpha_i - \frac{1}{N} \cdot \sum_{n=1}^{N} H(p^{\alpha_i}_{n+l|n} - p_{n+l})|, \tag{5}$$

where N represents the sample count in the testing data set. α_i denotes the expected frequency, i.e., the nominal proportion, spanning from 5% to 95% ($I = 19$) in 5% increments. $p^{\alpha_i}_{n+l|n}$ indicates the α_i-th forecasted quantile with an observation of p_{n+l}, while $H(x)$ denotes the unit step function.

2) Sharpness: Assessing sharpness involves quantifying the average width of prediction interval (PI) across different levels (1-α_i)

$$\overline{\delta_l} = \frac{1}{I \cdot N} \cdot \sum_{i=1}^{I} \sum_{n=1}^{N} (p^{1-\alpha_i/2}_{n+l|n} - p^{\alpha_i/2}_{n+l|n}). \tag{6}$$

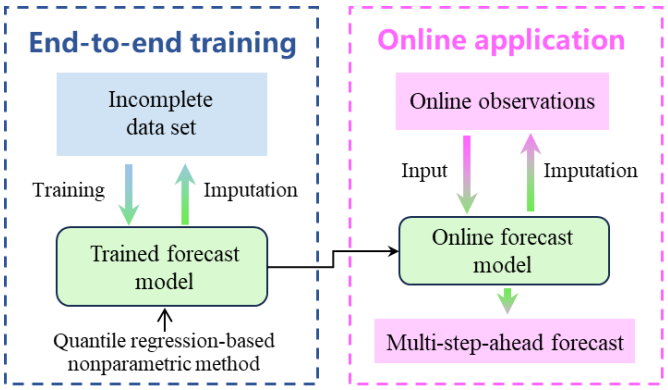


Fig. 1. End-to-end nonparametric framework for probabilistic WPF with missing data imputation.

3) Skill Score: The skill score integrates reliability and sharpness, and an average skill score over N time spots is computed as

$$\overline{SS_l} = \frac{1}{N} \cdot \sum_{n=1}^{N} \sum_{i=1}^{I} \{[H(p_{n+l|n}^{\alpha_i} - p_{n+l}) - \alpha_i] \cdot (p_{n+l} - p_{n+l|n}^{\alpha_i})\}. \tag{7}$$

III. Methodology

A. Framework

The overview of the end-to-end nonparametric probabilistic WPF method with missing data tolerance is depicted in Fig. 1. The method consists of end-to-end training and online application procedures, and a quantile regression-based nonparametric forecast model presented in (3) is adopted within our framework. In the end-to-end training stage, the nonparametric forecast model is trained considering missing data imputation. Therein, the forecast model is employed to impute the incomplete data set, followed by the training of the forecast model based on the imputed data set. These imputation and training steps are executed iteratively in an end-to-end manner. In the online application stage, the trained forecast model is utilized for multi-step-ahead probabilistic WPF via imputing the online observations.

It is noteworthy that the proposed end-to-end framework is model-agnostic. Other nonparametric learning-based forecast models can be incorporated into it by replacing the forecast model $F(\cdot;\theta)$ in (3) and the loss function $\mathcal{L}$ in (4) with the target ones. Next, we provide detailed explanations of the end-to-end training and online application procedures within our proposed framework in the following two sections, respectively.

B. End-to-End Training With Missing Data Tolerance

We formulate the historical time series into multiple successive inputs preparing for model training. The end-to-end training for the forecast model based on successive inputs with missing data imputation is shown in Fig. 2, mainly including continuous imputation and model training with revised loss.

1) Continuous Imputation: On the top of Fig. 2, green boxes represent actual values observed in the wind power series at corresponding timestamps above them, while grey boxes indicate missing observations. Pink boxes with dashed borders demonstrate imputed values. During one training epoch for the end-to-end training process, a batch of successive inputs with a size T (much smaller than the length of the total training data) is sampled first considering the training efficiency. These T successive inputs $[\boldsymbol{P}_t, \boldsymbol{P}_{t+1}, \cdots, \boldsymbol{P}_{t+T-1}]$ are generated from the incomplete time series continuously. For any time spot i, where $i \in [t, \cdots, t+T-1]$, the input $\boldsymbol{P}_i$ represents a wind power series with a lag interval δ from time spot $i-\delta+1$ to i.

Specifically, denoting the input $\boldsymbol{P}_t$ (the first one in successive inputs) at time t as $[p_{t-\delta+1}, \cdots, p_t]$ (assuming there are no missing data or they have been imputed), the next input $\boldsymbol{P}_{t+1}$ is formulated as $[p_{t-\delta+2}, \cdots, \hat{p}_{t+1}]$ with the new information $\hat{p}_{t+1}$ given by

$$\hat{p}_{t+1} = p_{t+1} \cdot \gamma_{t+1} + (1 - \gamma_{t+1}) \cdot p_{t+1|t}^{0.5}, \tag{8}$$

where γ_{t+1} is a binary variable denoting the presence (1) or absence (0) of the observation at time spot $t+1$. In situations where data is missing ($\gamma_{t+1} = 0$), the missing observation is imputed as $p_{t+1|t}^{0.5} = F(\boldsymbol{P}_t, 0.5; \theta)$, specifically, the forecasted median at the last time spot t. In contrast, when data is available ($\gamma_{t+1} = 1$), the actual value p_{t+1} is used to formulate $\boldsymbol{P}_{t+1}$. Eq. (8) establishes the link between the imputation and forecasting processes in our method, as the probabilistic forecasting result (forecasted median) is employed for imputing the missing observation.

We continuously apply the imputation process described in (8) from time spot t to $t+T-1$, generating all the required inputs. Meanwhile, these inputs are forwarded to the forecast model F_θ (abbreviated form of $F(\cdot;\theta)$). Hidden vectors $[\mathbf{h}_t^\alpha, \mathbf{h}_{t+}^\alpha, \cdots, \mathbf{h}_{t+T-1}^\alpha]$ are calculated and passed through the forecast model over T time spots and corresponding forecasted quantiles $[p_{t+1|t}^\alpha, p_{t+2|t+1}^\alpha, \cdots, p_{t+T|t+T-1}^\alpha]$ are computed via feedforward calculation. This continuous imputation process empowers the forecast model to fill in multiple missing observations from time spots t to $t+T-1$. In the next section, we introduce the training technique for controlling the imputation error that propagates from missing data imputation to probabilistic forecasting.

2) Model Training With Revised Loss: A revised loss involving the pinball loss and the mean square error (MSE) loss is formulated here. Taking into account all inputs across T time spots, the revised loss over the entire T successive inputs is defined as

$$\begin{aligned}
\mathcal{L}_{re} &= \frac{1}{T} \sum_{i=t}^{T+t-1} \left(\mathcal{L}_p^i + \mathcal{L}_m^i\right), \\
\mathcal{L}_p^i &= \begin{cases} \gamma_{i+1} \cdot \sum_{\alpha \in Q} \alpha \cdot (p_{i+1} - p_{i+1|i}^\alpha), & p_{i+1} \geq p_{i+1|i}^\alpha \\ \gamma_{i+1} \cdot \sum_{\alpha \in Q} (1-\alpha) \cdot (p_{i+1|i}^\alpha - p_{i+1}), & p_{i+1} < p_{i+1|i}^\alpha \end{cases} \\
\mathcal{L}_m^i &= \gamma_{i+1} \cdot (p_{i+1|i}^{0.5} - p_{i+1})^2,
\end{aligned} \tag{9}$$

where $\mathcal{L}_p^i$ denotes the pinball loss, $\mathcal{L}_m^i$ denotes the MSE loss at time spot i, $i \in [t, T+t-1]$, p_{i+1} denotes the observation

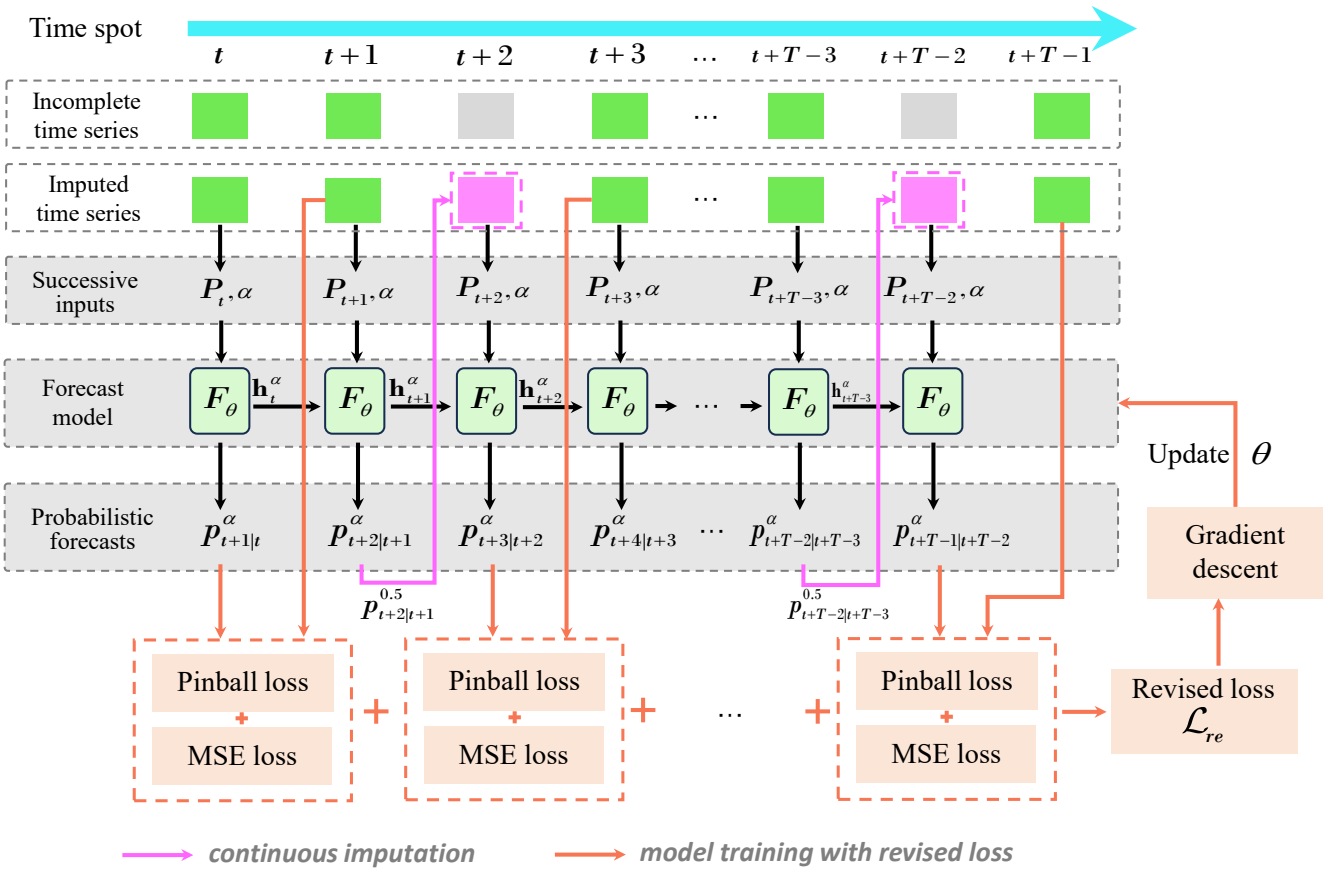


Fig. 2. End-to-end training with missing data tolerance.

of $p^{\alpha}_{i+1|i}$. In (9), the pinball loss $\mathcal{L}^i_p$ is used to calibrate the probabilistic forecast performance based on forecasted quantiles and corresponding observations. The the MSE loss $\mathcal{L}^i_m$ is introduced to avoid nonsignificant results and make the forecasted median as close as possible to the corresponding actual value. Here, we solely record the pinball loss when the actual value is observed at time spot $i+1$, because we do not have the necessary observation p_*.

C. Online Application

In this stage, the trained forecast model is used for multi-step-ahead probabilistic WPF during the online application. Assuming the forecast range as L, i.e., forecast the next L steps quantiles covering future time spots from $t+1$ to $t+L$, a historical range is assigned before the forecast range to estimate the initial hidden vector for the forecast model in multi-step-ahead probabilistic WPF. Here, the historical range covers time spots from $t-L$ to $t-1$, which is designed with the same length as that of the forecast range to capture dynamic information of wind power series during recent L steps. Specifically, in the historical range, we implement the same continuous imputation process as introduced in Section III-B1. Specifically, successive inputs $[\boldsymbol{P}_{t-L}, \boldsymbol{P}_{t-L+1}, \cdots, \boldsymbol{P}_{t-1}]$ are formulated and the hidden vector $\mathbf{h}^{\alpha}_{t-1}$ for the initialization of forecast range is computed via the feedforward calculation.

Next, we elaborate on the multi-step-ahead probabilistic WPF process within the forecast range from time spots t to $t+L-1$ during the online application, as shown in Fig. 3. Specifically, the forecast model takes $\mathbf{h}^{\alpha}_{t-1}$ as the hidden vector in the LSTM and we implement another continuous imputation process when the binary variable γ_{t+1} in (8) is always 0 through the forecast range. In detail, denoting $\boldsymbol{P}_j$ as the input at the time spot j, where $j \in [t, t+L-1]$, we feed $\boldsymbol{P}_j$ into the forecast model F_θ and generate probabilistic forecasts p^{α}_{j+1} at different proportions α. Subsequently, the input $\boldsymbol{P}_{j+1}$ at $j+1$ is formulated as $[p_{j-\delta+2}, \cdots, \hat{p}_{j+1}]$ with the new information $\hat{p}_{j+1}$ given by the forecasted median at last time spot as

$$\hat{p}_{j+1} = p^{0.5}_{j+1|j}, \tag{9}$$

and the probabilistic forecasts for time spot $j+1$ are calculated again via F_θ based on $\boldsymbol{P}_{j+1}$ as $y^{\alpha}_{j+2|j+1}$.

We continuously implement the imputation process until the end of the forecast range. In this way, we obtain many quantile traces of $[y^{\alpha}_{t+1|t}, \cdots, y^{\alpha}_{t+L|t+L-1}]$ and finally achieve multi-step-ahead forecasting.

IV. Experiments

A. Description of Data Set and Mechanism of Missingness

We collected a real-world wind power data set from the Australian National Electricity Market [47]. The data set records wind power data for the Willogoleche wind farm

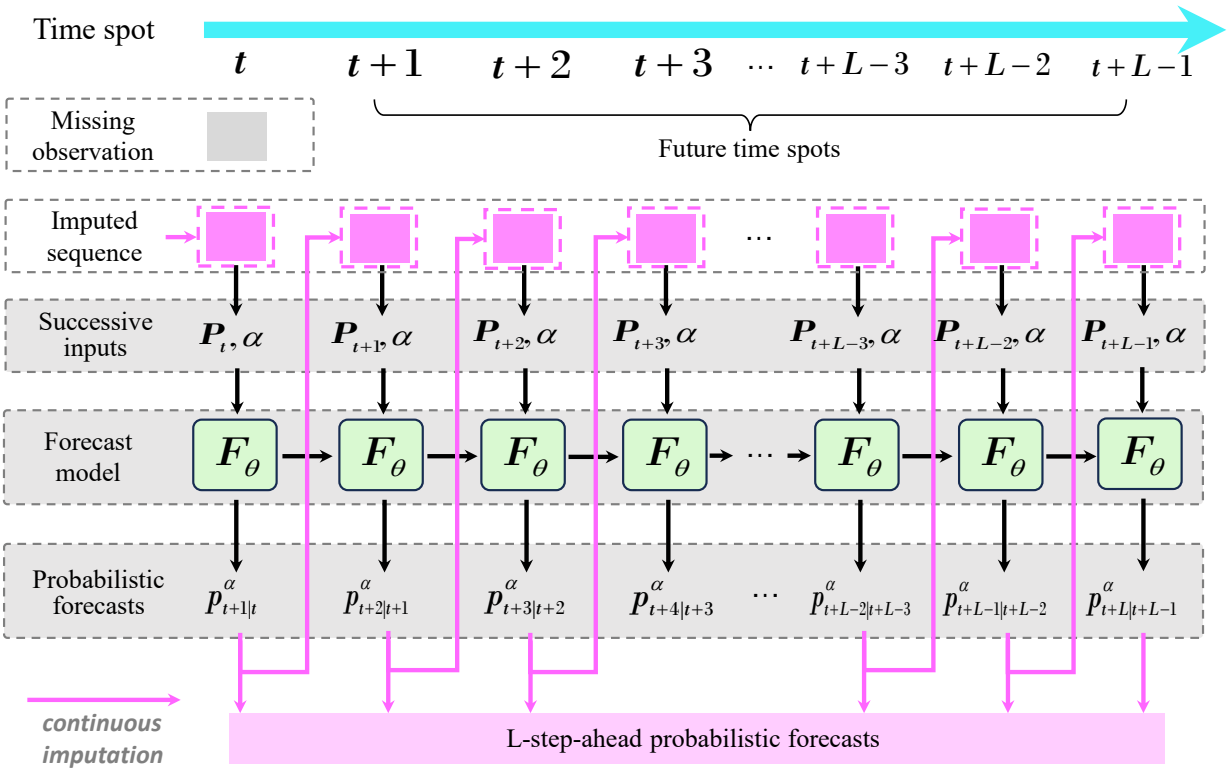


Fig. 3. Multi-step-ahead probabilistic WPF with continuous imputation during online application.

located in South Australia ranging from 2018 to 2020 whose capacity is 119.4 MW. The time resolution of the wind power data is 5 min. The input $\boldsymbol{P}_t$ of forecast models contains historical wind power series and also takes *time of the day* and *day of the year* of the forecasting target p_{t+1} as extra features as mentioned in subsection IV-B of [48] to consider the diurnal and seasonal effects. We allocated 60% historical samples for the forecast model training, the middle 20% for validation, and the last 20% for testing. Therefore, the data size for training data (used for offline end-to-end training) is [126144, 3], i.e., a matrix with a length of 126144 and a width of 3. For the testing data (used for online forecasting), the data size is [42048, 3], and we alternatively implement the online forecasting based on the testing sample one by one in a rolling window manner. All features in $\boldsymbol{P}_t$ as well as p_{t+1} have been normalized via min-max normalization [49] before the forecast.

We assume that the missingness in wind power data is consistent with a mechanism of missing completely at random (MCAR), which implies that the absence of data is unrelated to the values [50]. The data missing happens at both the end-to-end training and online application stages. The missing rate ranges from 5% to 50%, with 5% for each step, encompassing both low and high missing rate scenarios. In each scenario, to implement the MCAR mechanism with varying missing rates, we set the probability of data being missing at each time point equal to the predefined missing rate. This ensures that the positions of missing data are randomly allocated, allowing us to achieve the predetermined overall missing rate. Fig. 4

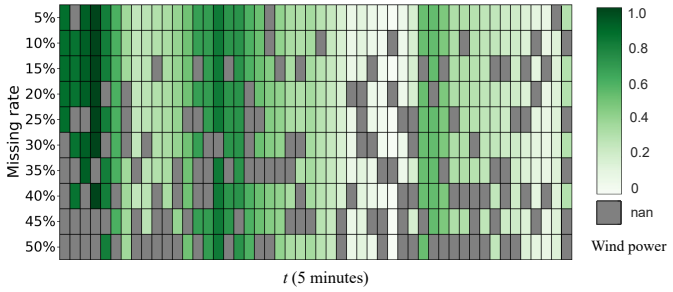


Fig. 4. Visualization of missing patterns under different missing rates.

visualizes missing patterns over 50 successive time spots with increasing missing rates. The grey rectangles denote missing data, and the green ones represent the normalized value of wind power generation. Moreover, the system model indicating data size and features associated with different missing rate scenarios is presented in Fig. 5.

B. Benchmarks

In our experiments, we examine several state-of-the-art imputation and forecasting benchmarks, which include both two-phase and end-to-end approaches. Two-phase approaches separate the imputation and prediction into two procedures, combining various commonly used techniques for imputation and prediction. Therein, the imputation methods were chosen from LI [51], KNN [36], RF [33], and autoencoder [31]. The forecasting methods were designed with LSTM owing to

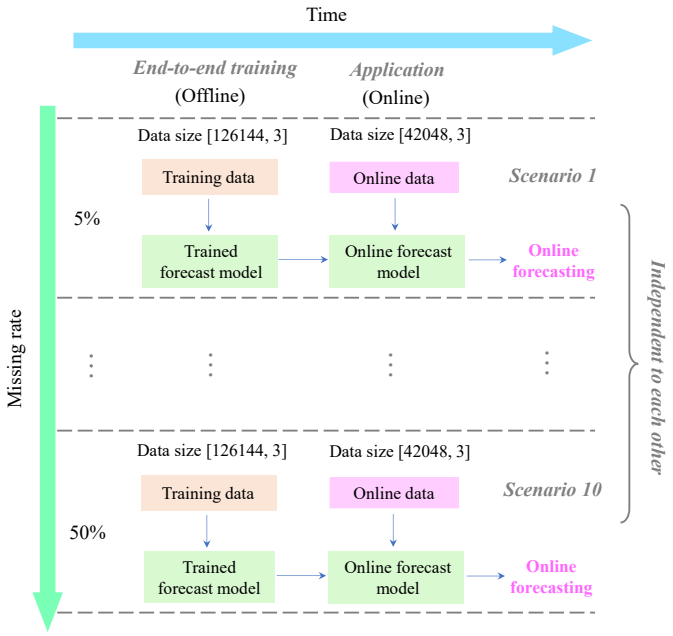


Fig. 5. The system model.

its superiority in renewable forecasting over other algorithms verified in [42], [43] and the quantiles were computed via a fully connected layer. We represent the number of the layer for LSTM as N_F and the width for each layer as W_F in the forecast model. For a fair comparison, all benchmarks and our proposal share the same LSTM-based nonparametric structure mentioned above. Specifically, we refer to these two-phase benchmarks as LI-LSTM, KNN-LSTM, RF-LSTM, and Autoencoder-LSTM. For the end-to-end approach, the Recursive LSTM (Rec-LSTM) [42], variational autoencoder (VAE) [52], and adaptive quantile regression (AQR) [53] were adopted as benchmarks. In Rec-LSTM [42], the forecast model was employed to simultaneously execute imputation and prediction in a parametric way. In VAE [52], an encoder estimated the posterior distribution of latent variables with missing inputs, and a decoder reconstructed the missing values using the estimated latent variables based on Gaussian distribution. In AQR [53], the missing data features were captured by a NN-based feature extraction module with missingness indicators, and forecasted quantiles were then output by a non-crossing quantile regression NN based on the extracted features. Apart from Rec-LSTM and VAE, these benchmarks are nonparametric approaches. The multi-step-ahead forecasts in these benchmarks were conducted using a continuous imputation strategy.

C. Model Training and Hyperparameter Tuning

We allocated 60% historical samples for the forecast model training, the middle 20% for validation, and the last 20% for testing. To mitigate the effect of intrinsic randomness in model training and validation, we adopt a four-fold cross-validation technique [54]. If the average loss on the training splits (three folds from the original data set) has been lower than that on the validation split (the rest one fold from the original data set) over 20 successive epochs, the training process will be stopped to avoid overfitting. A grid-search method was adopted to find the optimal structure of the forecast model (associated with N_F and W_F), the number of successive inputs T during the end-to-end training process, the lag interval δ, and the learning rate l_r. Specifically, N_F was chosen from {8, 16, 32, 64}, W_F from {16, 32, 64, 128}, T from {12, 24, 48, 96}, δ from {5min, 10min, 15 min, 20 min}, and l_r from {0.0001, 0.001, 0.01, 0.1}. The optimal $\{N_F, W_F, T, \delta, l_r\}$ was determined when the average cross-validation loss was the lowest during the grid-search procedure for different data-missing scenarios, respectively, which was recorded in Table I. The aforementioned training processes were carried out on CentOS 7.6 using 8 TITAN V GPUs, and Adam was selected as the optimizer for gradient descent in all experiments.

TABLE I
OPTIMAL HYPERPARAMETERS

Missing rate	N_F	W_F	T	δ	l_r
5%	16	32	24	15 min	0.001
10%	16	64	24	10 min	0.001
15%	16	32	48	15 min	0.001
20%	16	64	24	15 min	0.001
25%	16	32	48	10 min	0.001
30%	16	32	48	15 min	0.001
35%	16	64	24	20 min	0.001
40%	32	64	48	15 min	0.001
45%	16	64	48	15 min	0.001
50%	32	128	48	15 min	0.001

D. Prediction Results of the Proposed Method

Probabilistic forecasting results generated by our proposal, reflecting confidence levels from 10% to 90%, are illustrated in Fig. 6, which indicates PIs between different forecasted quantiles. Fig. 6(a), (b), (c), and (d) showcase forecasting outcomes over 1000 successive time spots under scenarios with 10%, 20%, 30%, and 40% missing rates, respectively. In each subfigure, the red line represents the observed forecasting target, while the horizontal axis delineates successive time intervals with a 5-min resolution. It is evident that our proposed method consistently delivers highly reliable probabilistic forecasting results across varying missing rates. The forecasted PIs adeptly include the observations and precisely capture the upward and downward trends in the wind power series. This impressive performance of probabilistic WPF results can be attributed to the exceptional capacity of our end-to-end nonparametric method to learn missing data patterns across various missing rate scenarios, which mitigates the negative impacts on missing data and substantiates the robustness of our proposed method to handle missing values in probabilistic WPF.

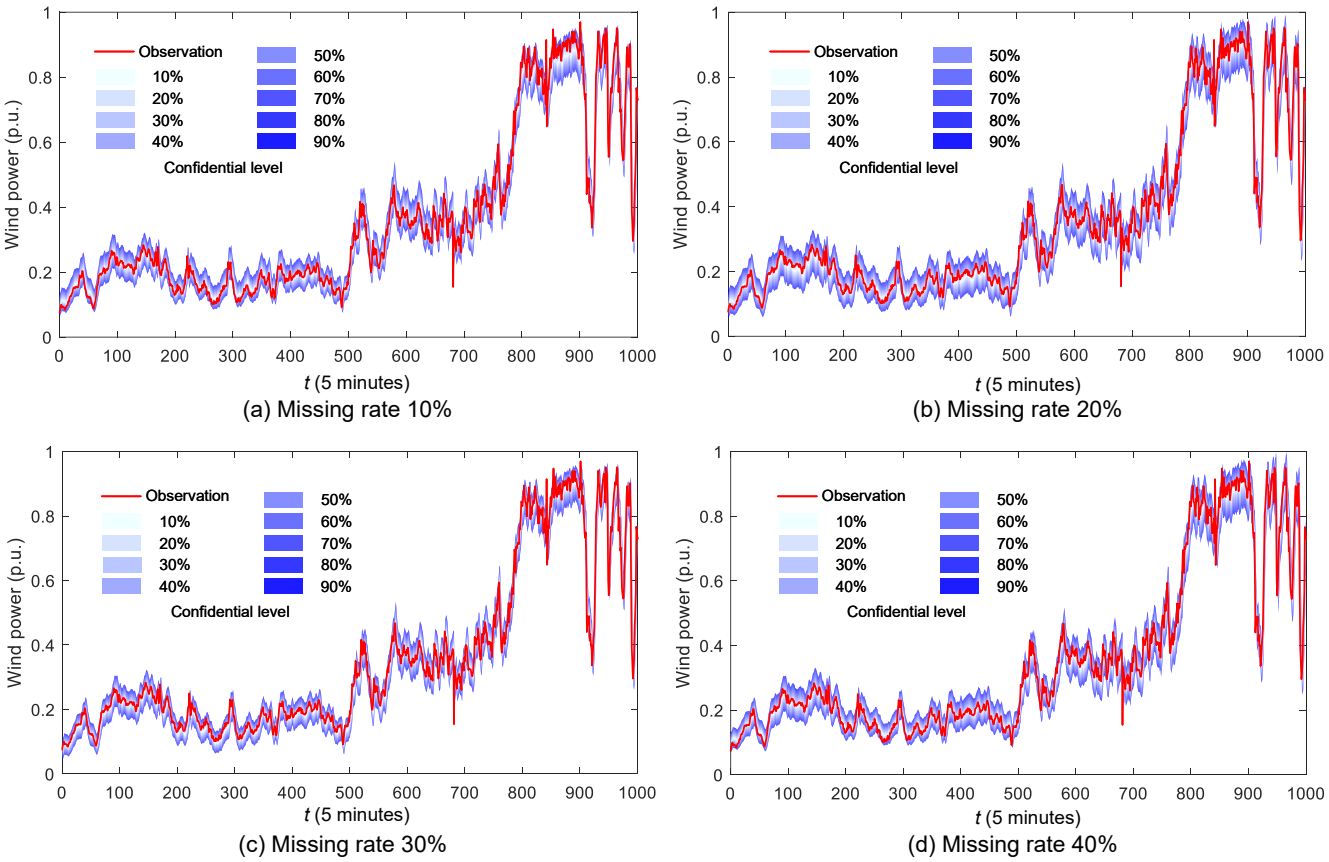
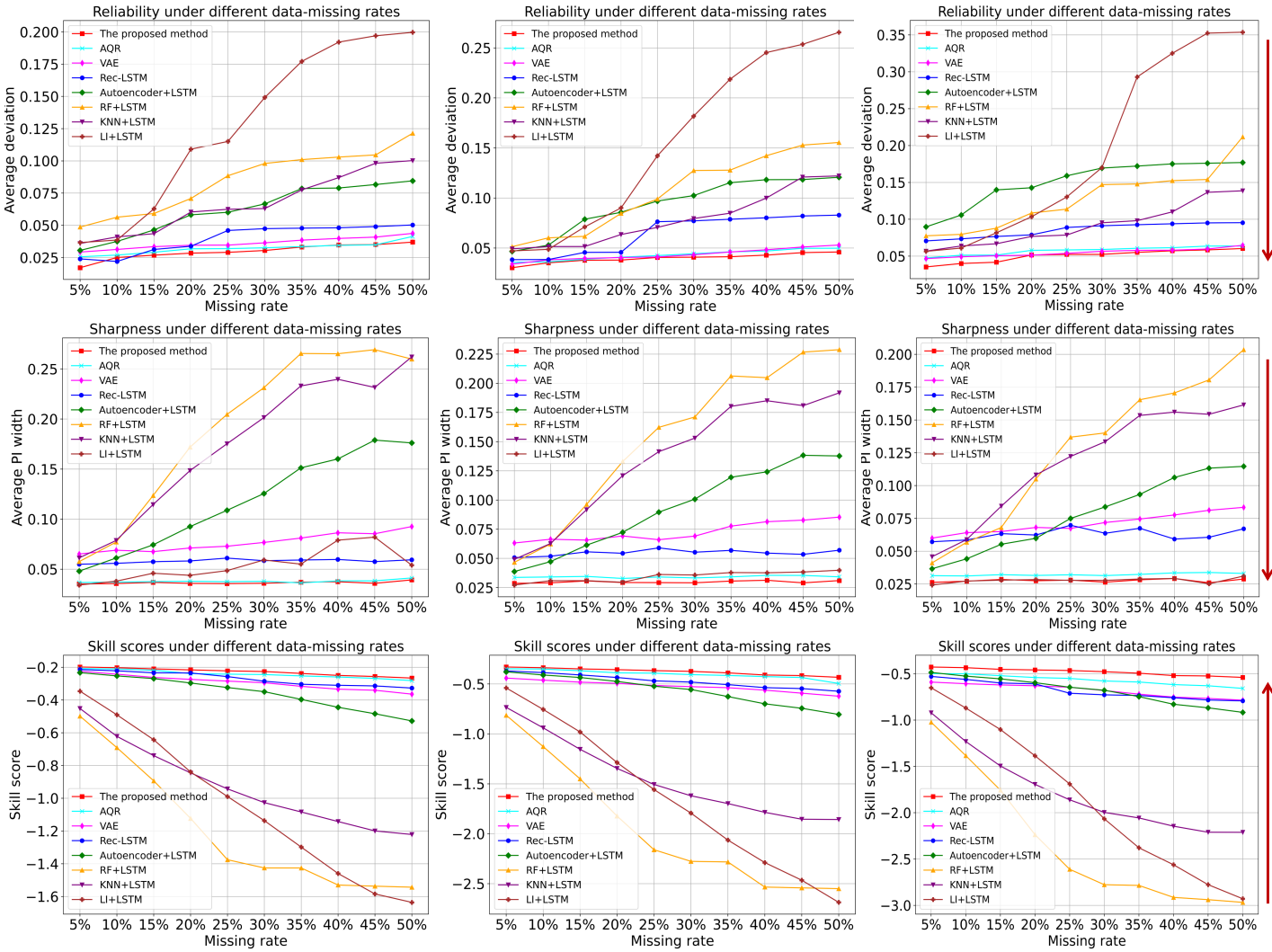


Fig. 6. One-step-ahead probabilistic WPF results with different missing rates.

E. Evaluation and Comparison of Performance

We conducted a comparative analysis of the performance of LI-LSTM, KNN-LSTM, RF-LSTM, Autoencoder-LSTM, Rec-LSTM, VAE, AQR models, and our proposed method concerning one-step-ahead, six-step-ahead, and twelve-step-ahead probabilistic WPF across 10 missing rates ranging from 5% to 50%, respectively. Evaluation metrics including the average deviation $\overline{A_l}$ in (5), average PI width $\overline{\delta_l}$ in (6), and the average skill score $\overline{SS_l}$ in (7) were employed to assess the probabilistic forecasting outcomes on the testing data set. Corresponding evaluation results with different missing rates are presented in Fig. 7 and the average sums of evaluation metrics over all missing rates are recorded in Table II. Based on the results, we have the following findings.

1) Highest Reliability, Sharpness, and Skill Score for the Proposed Method Over Different Data-Missing Scenarios: Fig. 7(a) reports evaluation results of one-step-ahead probabilistic WPF. In the reliability evaluation, our proposed end-to-end nonparametric approach consistently demonstrates low deviations in different tested data-missing scenarios, achieving the lowest average deviation $\overline{A_l}$ (2.97%) as shown in Table II across all cases, thus indicating the highest reliability. Other benchmarks exhibit relatively lower $\overline{A_l}$ close to our proposal when the missing rate is below 15%, while one of them, i.e. KNN-LSTM, deteriorates significantly on reliability with deviations exceeding 10% when the missing rate surpasses 20%. In the sharpness evaluation, depicted in Fig. 7(a) and summarized in Table II, our proposal demonstrates the smallest average PI width $\overline{\delta_l}$ in most experiments across missing rates ranging from 5% to 50%. In contrast, methods such as KNN-LSTM, RF-LSTM, and Autoencoder-LSTM showcase higher $\overline{\delta_l}$ (larger PI widths) as the missing rate increases. In terms of the comprehensive metric, the average evaluation results of the skill score presented in the fourth column of Table II affirm the superiority of our proposed end-to-end nonparametric approach. With the highest average skill score $\overline{SS_l}$ (-0.228) among all methods, it underscores the superior overall performance of our proposal. From Fig. 7(b), (c), and Table II, when we consider six-step-ahead and twelve-step-ahead probabilistic WPF, our proposed method maintains its superiority associated with the highest skill score and reliability with the lowest PI widths. In addition, we trained the forecast models for different methods using a randomly shuffled dataset, where the chronological order of the original time series data was disrupted, to showcase the model's performance on randomly chosen data. Table III presents the average skill scores, denoted as $\overline{SS_l}'$, for forecasting scenarios at different forecasting steps based on randomly chosen data with missing rates ranging from 5% to 50%. Our proposed method consistently achieves the highest skill scores in all scenarios. This superior performance can be attributed to its ability to effectively manage imputation errors arising from



(a) One-step-ahead probabilistic WPF (b) Six-step-ahead probabilistic WPF (c) Twelve-step-ahead probabilistic WPF

Fig. 7. Reliability, sharpness, and skill score evaluations related to different forecasting scenarios under 10 missing rates varying from 5% to 50%. (a), (b), and (c) are evaluation results for one-step-ahead, six-step-ahead, and twelve-step-ahead probabilistic WPF, respectively.

TABLE II
AVERAGE SUMS FOR EVALUATION METRICS WITH DIFFERENT AHEAD STEPS OVER ALL MISSING RATES

	One-step-ahead forecasts			Six-step-ahead forecasts			Twelve-step-ahead forecasts		
	$\overline{A_l}\%$	$\overline{\delta_l}$	$\overline{SS_l}$	$\overline{A_l}\%$	$\overline{\delta_l}$	$\overline{SS_l}$	$\overline{A_l}\%$	$\overline{\delta_l}$	$\overline{SS_l}$
LI-LSTM	12.8	0.0540	−1.04	15.6	0.0344	−1.64	19.3	0.0278	−1.84
KNN-LSTM	6.69	0.175	−0.927	7.93	0.136	−1.45	9.22	0.118	−1.78
RF-LSTM	8.52	0.193	−1.20	10.6	0.154	−1.96	12.8	0.127	−2.34
Autoencoder-LSTM	6.23	0.118	−0.358	9.36	0.930	−0.565	15.1	0.0782	−0.685
Rec-LSTM	3.99	0.0582	−0.270	6.46	0.0549	−0.471	8.57	0.0629	−0.681
VAE	3.63	0.0768	−0.294	4.35	0.0727	−0.523	5.49	0.0713	−0.680
AQR	3.22	0.0379	−0.242	4.31	0.0342	−0.403	5.76	0.0323	−0.568
Proposed method	2.97	0.0365	−0.228	3.98	0.0298	−0.376	5.04	0.0276	−0.479

In this table, $\overline{A_l}\%$ denotes the average deviation in reliability evaluation, which is the lower the better. $\overline{\delta_l}$ denotes the average PI width in sharpness evaluation, which is the lower the better. $\overline{SS_l}$ denotes the average skill score, which accounts for both reliability and sharpness, and higher values are desirable.

TABLE III
PERFORMANCE COMPARISON FOR SKILL SCORES $\overline{SS_l}'$ BASED ON RANDOMLY CHOSEN DATA

	$\overline{SS_l}'$ for one-step-ahead forecasts	$\overline{SS_l}'$ for six-step-ahead forecasts	$\overline{SS_l}'$ for twelve-step-ahead forecasts
Autoencoder-LSTM	−0.773	−0.872	−1.47
Rec-LSTM	−0.657	−0.783	−1.62
VAE	−0.691	−0.765	−1.74
AQR	−0.688	−0.802	−1.51
Proposed method	−0.621	−0.745	−1.29

the continuous imputation process while simultaneously optimizing forecasting accuracy. By considering both imputation and forecasting errors, the model is able to handle multiple missing values continuously, with future data also treated as potential missing values. Moreover, our method makes the model adaptable to both series-based and random data scenarios as verified by skill scores in Table II and Table III, since no strict constraints are imposed on the data form. In contrast, other methods do not account for imputation errors (Rec-LSTM, VAE, and AQR) or separate the imputation and forecasting processes (Autoencoder-LSTM). As a result, these alternatives struggle to control imputation errors, leading to inferior performance in random data scenarios as well.

2) End-to-End Methods Perform Better Compared to Two-Phase Methods: It is worth noting that the end-to-end approach, i.e., Rec-LSTM, VAE, and AQR, closely rival our proposed end-to-end approach in terms of $\overline{A_l}$ and surpasses other two-phase benchmarks in reliability, as depicted in the top subfigures of Fig. 7(a), (b), and (c), respectively. In the sharpness evaluation, observed from the middle subfigures in Fig. 7(a), (b), (c), and the average PI width $\overline{\delta_l}$ in Table II for different step-ahead scenarios, end-to-end methods, including Rec-LSTM, VAE, AQR, and our proposal, generally exhibit smaller PI widths with higher sharpness. Regarding the comprehensive indicator skill score illustrated at the bottom of Fig. 7(a), our proposed method and other end-to-end methods demonstrate only a slight decrease as missing rates increase from 5% to 50%, whereas significant declines are observed in high missing rate scenarios for those utilizing two-phase methods. The higher average skill scores $\overline{SS_l}$ for end-to-end methods recorded in Table II further emphasize the superiority of end-to-end methods over two-phase approaches. This aligns with the observed high performance in probabilistic WPF, as shown in Fig. 6, across different missing data scenarios.

3) Deep Learning-Based Imputation Techniques Perform Better Compared to Machine Learning Ones: Methods adopting deep learning-based imputation techniques, i.e., our proposed method, AQR, VAE, Rec-LSTM, and Autoencoder-LSTM generally yield better evaluation results compared to those using traditional machine learning approaches, i.e., LI-LSTM, KNN-LSTM, and RF-LSTM. This superiority can be attributed to the enhanced nonlinear approximation capabilities of deep learning techniques, facilitating accurate imputation of missing values and consequently resulting in higher performance in probabilistic WPF. Additionally, we have conducted additional experiments using varying amounts of training data to evaluate performance of our deep learning-based method. Fig. 8 presents the skill scores of our method using different amounts of training data, ranging from 20000 to 120000 samples. The results show significant improvements in forecasting performance as the training data increases from 20000 to 80000 samples. However, only marginal gains are observed when the number of training samples exceeds 100000. This aligns with the empirical observation that the test loss (indicated by the negative skill score) of deep learning models tends to converge as the dataset size grows, following the scaling law [55]. From this, we can conclude that our method is data-efficient, and that two years of data (126144 training samples) is sufficient for effective model training and forecasting.

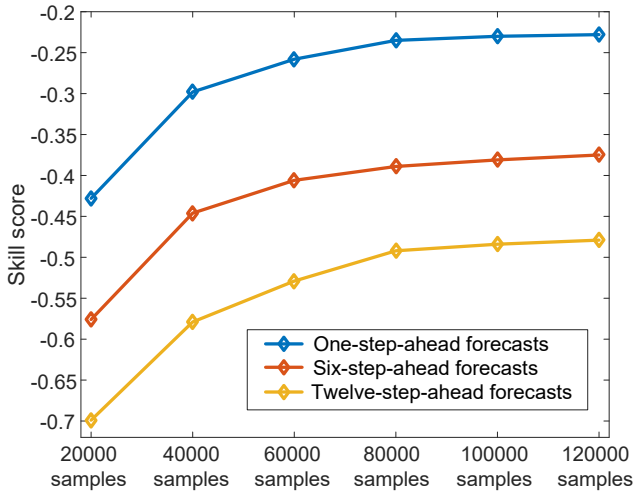


Fig. 8. Skill scores under different amounts of training data.

4) Nonparametric Methods Demonstrate Superiority in Multi-Step-Ahead Forecasts: The results from various methods in Table II demonstrate that skill scores decline (more negative) as the forecasting step increases from one step to twelve due to the growing difficulty of forecasting tasks as the lead time extends, reflecting decreases in forecast accuracy. Our end-to-end nonparametric method outshines its parametric benchmark, i.e., Rec-LSTM and VAE, particularly in multi-step-ahead forecasting scenarios, showcasing the highest skill score. The observed performance degradation in Rec-LSTM and VAE can be attributed to the challenges associated with modeling probabilistic WPF in a parametric manner, especially for multi-step-ahead forecasts. In such cases, the assumption of Gaussian distribution for the multi-step-ahead probability distributions of nonstationary wind power may not always hold, leading to the limitations of Rec-LSTM in such scenarios. Additionally, it is noteworthy that Autoencoder-LSTM demonstrates skill scores more closely aligned with those of Rec-LSTM as we increase the forecast steps from one to twelve, further emphasizing the superiority of the nonparametric method in modeling multi-step-ahead probabilistic WPF. For AQR, it provides skill scores closest to those of our proposed method, which can be attributed to its end-to-end nonparametric structure. However, AQR does not explicitly address the accuracy of missing data imputation.

TABLE IV
IMPROVEMENT COMPARED WITH THE BEST ALTERNATIVE IN TERMS OF SKILL SCORE

	One-step-ahead forecasts	Six-step-ahead forecasts	Twelve-step-ahead forecasts
$+\overline{SS_l}$	5.79%	6.70%	15.7%

$+\overline{SS_l}$ denotes the ratio of the skill score of the best alternative minus that of our proposal to the skill score of the corresponding alternative.

Consequently, the imputation error in AQR may be difficult to control, resulting in inferior performance compared to our proposed method.

To sum up, our proposed method effectively integrates the strengths of the nonparametric approach, end-to-end structure, and deep learning, yielding a highly competitive performance in probabilistic WPF with tolerance for missing data across varying rates of absence. To show the efficacy of our proposal more directly, we present the improvement in terms of the overall skill score $\overline{SS_l}$ compared to the best alternative in Table IV, which elucidates significant improvements our end-to-end nonparametric method brings in the ability to model probabilistic WPF with integrated missing data handling especially in multi-step-ahead forecasts.

V. CONCLUSION

An end-to-end nonparametric approach for probabilistic WPF with missing data tolerance has been designed in this paper. In our presented method, missing data are continuously imputed by probabilistic forecasting results, i.e., forecasted medians. The forecast model is then trained with a revised loss, encompassing both the error in missing data imputation and the pinball loss computed from forecasted quantiles, in an end-to-end manner. Numerical simulations validate the supremacy of our method over other state-of-the-art counterparts, taking into account conformance with real-world data and the width of PIs. The key findings in our work include:

1) The proposed end-to-end nonparametric method demonstrates the highest reliability, sharpness, and skill score in probabilistic WPF across different data-missing scenarios. Notably, the performance of the forecast model only experiences slight declines with increasing missing rates.
2) End-to-end methods perform better compared to two-phase methods.
3) Deep learning-based imputation techniques perform better compared to machine learning ones.
4) Nonparametric methods demonstrate superiority in multi-step-ahead forecasts.

In the future, we will develop our end-to-end nonparametric method to probabilistic WPF considering outliers, which often result from equipment malfunctions or measurement errors and deteriorate the performance of deep learning-based forecast models. Effective identification and recovery strategies can be incorporated into the end-to-end nonparametric method to mitigate the negative impact of outliers.

REFERENCES

[1] A. Dhabi, "Irena (2023), renewable energy statistics 2023." https://www.irena.org/publications/2023/Jul/Renewable-Energy-Statistics-2023. Accessed: 2023-8-3.

[2] S. Feng, Z. Song, Q. Yang, Y. Hou, Z. Wang, F. Liu, B. Wang, and W. Wang, "Long-term changes of wind resources and its impact on wind power development under climate change in china," *Energy Internet*, vol. 1, no. 1, pp. 52–62, 2024.

[3] S. Buhan and I. Cadirci, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1231–1242, 2015.

[4] L. Zjavka and S. Misak, "Direct wind power forecasting using a polynomial decomposition of the general differential equation," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1529–1539, 2018.

[5] M. J. Sanjari, H. B. Gooi, and N.-K. C. Nair, "Power generation forecast of hybrid PV–Wind system," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 703–712, 2020.

[6] H. Wen, J. Ma, J. Gu, L. Yuan, and Z. Jin, "Sparse variational gaussian process based day-ahead probabilistic wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 957–970, 2022.

[7] W. Xie, P. Zhang, R. Chen, and Z. Zhou, "A nonparametric bayesian framework for short-term wind power probabilistic forecast," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 371–379, 2019.

[8] J. Wang, A. AlShelahi, M. You, E. Byon, and R. Saigal, "Integrative density forecast and uncertainty quantification of wind power generation," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 1864–1875, 2021.

[9] P. Lu, Q. Dong, E. Du, C. Fang, Y. Su, and N. Zhang, "Improving wind power accommodation via an interval model predictive control strategy," *CSEE Journal of Power and Energy Systems*, pp. 1–12, 2024.

[10] J. W. Messner, A. Zeileis, J. Broecker, and G. J. Mayr, "Probabilistic wind power forecasts with an inverse power curve transformation and censored regression," *Wind Energy*, vol. 17, no. 11, pp. 1753–1766, 2014.

[11] L. Landberg, "A mathematical look at a physical power predictionmodel," *Wind Energy*, vol. 1, no. 1, pp. 23–28, 2015.

[12] H. Liu, H.-Q. Tian, C. Chen, and Y.-f. Li, "A hybrid statistical method to predict wind speed and wind power," *Renewable energy*, vol. 35, no. 8, pp. 1857–1861, 2010.

[13] Y. Zhao, L. Ye, P. Pinson, Y. Tang, and P. Lu, "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5029–5040, 2018.

[14] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[15] Z. Shi, H. Liang, and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1177–1187, 2018.

[16] M. A. Hossain, R. K. Chakrabortty, S. Elsawah, E. M. Gray, and M. J. Ryan, "Predicting wind power generation using hybrid deep learning with optimization," *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 8, pp. 1–5, 2021.

[17] M.-S. Ko, K. Lee, J.-K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1321–1335, 2021.

[18] M. Zou, N. Holjevac, J. Dakovic, I. Kuzle, R. Langella, V. D. Giorgio, and S. Z. Djokic, "Bayesian CNN-BiLSTM and vine-GMCM based probabilistic forecasting of hour-ahead wind farm power outputs," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1169–1187, 2022.

[19] P. Arora, S. M. J. Jalali, S. Ahmadian, B. K. Panigrahi, P. N. Suganthan, and A. Khosravi, "Probabilistic wind power forecasting using optimized deep auto-regressive recurrent neural networks," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2814–2825, 2023.

[20] K. Wang, Y. Zhang, F. Lin, J. Wang, and M. Zhu, "Nonparametric probabilistic forecasting for wind power generation using quadratic spline quantile function and autoregressive recurrent neural network," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 4, pp. 1930–1943, 2022.

[21] H.-Z. Wang, G.-Q. Li, G.-B. Wang, J.-C. Peng, H. Jiang, and Y.-T. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, 2017.

[22] H. Zhang, Y. Liu, J. Yan, S. Han, L. Li, and Q. Long, "Improved deep mixture density network for regional wind power probabilistic forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2549–2560, 2020.

[23] H. Zhang, J. Yan, Y. Liu, Y. Gao, S. Han, and L. Li, "Multi-source and temporal attention network for probabilistic wind power prediction," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2205–2218, 2021.

[24] X. Li, N. Yang, Z. Li, Y. Huang, Z. Yuan, X. Song, L. Li, and L. Zhang, "Confidence estimation transformer for long-term renewable energy forecasting in reinforcement learning-based power grid dispatching," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 4, pp. 1502–1513, 2024.

[25] Y. Li and X. Shen, "A novel wind speed-sensing methodology for wind turbines based on digital twin technology," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[26] C. Sun, Y. Chen, and C. Cheng, "Imputation of missing data from offshore wind farms using spatio-temporal correlation and feature correlation," *Energy*, vol. 229, p. 120777, 2021.

[27] R. Tawn, J. Browell, and I. Dinwoodie, "Missing data in wind farm time series: Properties and effect on forecasts," *Electric Power Systems Research*, vol. 189, p. 106640, 2020.

[28] C. N. Snchez, J. Enrquez-Zrate, R. Velzquez, M. Graff, and S. Sassi, "Analysis of wind missing data for wind farms in isthmus of tehuantepec," in *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pp. 1–6, 2018.

[29] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[30] C. Wang, B. Ren, X. Li, and L. Chen, "A cnn-bilstm and knn based missing data imputation for wind power generation forecasting," in *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*, pp. 4065–4070, 2023.

[31] G. Zhang, Y. Liu, and X. Jin, "A survey of autoencoder-based recommender systems," *Frontiers of Computer Science*, vol. 14, pp. 430–450, 2020.

[32] X. Liu and Z. Zhang, "A two-stage deep autoencoder-based missing data imputation method for wind farm scada data," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10933–10945, 2021.

[33] D. J. Stekhoven and P. Bühlmann, "Missforest–non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[34] E. Antonenko, A. Carreño, and J. Read, "Autoreplicative random forests with applications to missing value imputation," *Machine Learning*, vol. 113, no. 10, pp. 7617–7643, 2024.

[35] S. Hong and H. S. Lynn, "Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction," *BMC medical research methodology*, vol. 20, pp. 1–12, 2020.

[36] S. Tak, S. Woo, and H. Yeo, "Data-driven imputation method for traffic data in sectional units of road links," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1762–1771, 2016.

[37] P. Keerin and T. Boongoen, "Improved knn imputation for missing values in gene expression data," *Computers, Materials and Continua*, vol. 70, no. 2, pp. 4009–4025, 2021.

[38] S. Ryu, M. Kim, and H. Kim, "Denoising autoencoder-based missing value imputation for smart meters," *IEEE Access*, vol. 8, pp. 40656–40666, 2020.

[39] I. Gjorshoska, T. Eftimov, and D. Trajanov, "Missing value imputation in food composition data with denoising autoencoders," *Journal of Food Composition and Analysis*, vol. 112, p. 104638, 2022.

[40] Z. Pan, Y. Wang, K. Wang, H. Chen, C. Yang, and W. Gui, "Imputation of missing values in time series using an adaptive-learned median-filled deep autoencoder," *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 695–706, 2022.

[41] W. Liu, C. Ren, and Y. Xu, "Pv generation forecasting with missing input data: A super-resolution perception approach," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1493–1496, 2020.

[42] Q. Li, Y. Xu, B. S. H. Chew, H. Ding, and G. Zhao, "An integrated missing-data tolerant model for probabilistic pv power generation forecasting," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4447–4459, 2022.

[43] P. Arora, S. M. J. Jalali, S. Ahmadian, B. K. Panigrahi, P. N. Suganthan, and A. Khosravi, "Probabilistic wind power forecasting using optimized deep auto-regressive recurrent neural networks," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2814–2825, 2023.

[44] M. Chen, Z. Meng, Y. Liu, L. Luo, Y. Guo, and K. Wang, "Nonparametric end-to-end probabilistic forecasting of distributed generation outputs considering missing data imputation," in *2024 Second International Conference on Cyber-Energy Systems and Intelligent Energy (ICCSIE)*, pp. 1–6, 2024.

[45] S. B. Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.

[46] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

[47] "Australian energy market operator." http://www.nemweb.com.au/REPORTS/CURRENT/Dispatch_SCADA/.

[48] T. Hu, Q. Guo, Z. Li, X. Shen, and H. Sun, "Distribution-free probability density forecast through deep neural networks," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 2, pp. 612–625, 2020.

[49] M. J. Islam, S. Ahmad, F. Haque, M. B. I. Reaz, M. A. S. Bhuiyan, and M. R. Islam, "Application of min-max normalization on subject-invariant emg pattern recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[50] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*, vol. 793. John Wiley & Sons, 2019.

[51] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Applied Energy*, vol. 225, pp. 998–1012, 2018.

[52] H. Wen, "Probabilistic wind power forecasting resilient to missing values: an adaptive quantile regression approach," *Energy*, vol. 300, p. 131544, 2024.

[53] H. Wen, P. Pinson, J. Gu, and Z. Jin, "Tackling missing values in probabilistic wind power forecasting: A generative approach," *arXiv preprint arXiv:2403.03631*, 2024.

[54] T.-T. Wong and P.-Y. Yeh, "Reliable accuracy estimates from k-fold cross validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1586–1594, 2020.

[55] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

Zichao Meng (Member, IEEE) received his Ph.D. degree of electrical engineering in Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, in 2024. He is currently employed at Electrical Dispatch and Control Center, Guangdong Power Grid, Guangzhou, China. His research interests include probabilistic forecast and meta-learning-based optimization in power and energy systems.

Ye Guo (Senior Member, IEEE) received his B.E. and Ph.D. degrees from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2008 and 2013, respectively. He is currently an Associate Professor with Tsinghua-Berkeley Shenzhen Institute, Tsinghua University. From 2014 to 2018, he was a Postdoctoral Associate with Cornell University. His research interests include distributed optimization, game and market theory, state estimation, and their applications in power and energy systems. Between 2019 and 2021, he won the IEEE PES General Meeting Best Paper Award four times in three years in a row, with one co-authored paper being selected as the "Best-of-the-Best Papers". He has also received the Best Poster Award at PSERC IAB Meeting 2018, and the Best Paper Award at the 4th IEEE EI2 Conference.

Chenhao Zhao is currently working towards the Ph.D. degree of electrical engineering in Tsinghua-Berkeley Shenzhen Institute, Tsinghua University.