

Online Probabilistic Load Forecasts Considering Data Gaps

Pengfei Zhao, Weihao Hu, *Senior Member, IEEE, Senior Member, CSEE*, Di Cao, *Senior Member, IEEE*, Longcheng Dai, Qi Huang, *Fellow, IEEE*, and Zhe Chen, *Fellow, IEEE*

Abstract—Existing load forecasting methods typically assume that recent load data are available for prediction. This is not in conformity with reality since there is a time gap between the flow date (when power is consumed) and when measurement values are obtained. To this end, this letter proposes an online learning-based probabilistic load forecasting method considering the impact of the data gap. Specifically, an adaptive ensemble backpropagation-enabled online quantile regression algorithm is developed to optimize the parameters of the attention network recursively using the newly obtained load observations. To further improve the reliability and sharpness of prediction intervals under significant data gaps, we introduce an online interval calibration technique. The proposed online learning method allows us to adaptively capture the dynamic changes in load patterns and alleviate the information lags caused by data gaps. Comparative tests utilizing real-world datasets reveal the superiority of the proposed method.

Index Terms—Data gaps, online learning, probabilistic load forecasting.

I. INTRODUCTION

ELECTRICAL load forecasting plays a crucial role in the planning and decision-making of system operators [1]. In recent years, numerous statistical- and machine-learning-based load forecasting methods have been proposed in the literature [2]. Typically, these methods rely on recent load data for prediction [3], [4]. For example, to estimate the electrical load at time point $t + H$, the load values at time t , $t - 1$, $t - 2$, etc. are required. However, there exists the data availability issue in real-world applications. To actually

measure power consumption, we use meters that are aggregated to understand exactly what happened in the past. In distribution systems, Advanced Metering Infrastructure (AMI) and modern communication networks are essential for real-time data uploads to Energy Management Systems (EMS) for load forecasting. In many countries and regions, particularly in developing countries or remote areas with limited infrastructure and resources, although distribution grids may have EMS, AMI is not yet widely installed (approximately 57% globally as of late 2023), traditional electromechanical meters and basic electronic meters, which lack real-time communication capabilities and rely on labor-intensive manual readings to record consumption data, are still in use [5]. This can result in a data gap between the flow date (when the power is consumed) and when measurement values are obtained [6]. Considering the data gap, it would be challenging for existing forecasting methods to provide accurate load prediction values.

In this context, this paper proposes an online learning-based probabilistic load forecasting method to deal with the data gap, which has been rarely considered in existing literature. The main contributions are as follows: 1) The proposed online learning method can provide accurate forecasting outcomes even under large data gaps. An attention-based Neural Network (NN) is, first, developed to fuse low-fidelity historical load values under data gaps with high-fidelity weather data collected on the fly. The NN is, then, integrated into a novel Adaptive Ensemble BackPropagation (AEBP)-enabled online learning algorithm for online parameter updating. The online learning strategy allows the proposed model to adaptively capture the dynamic changes in load patterns and compensate for missing historical load information caused by data gaps, which differentiates it from traditional offline learning methods that struggle to provide accurate forecasting values when considering the data gap. 2) The proposed method can achieve effective uncertainty quantification even in large data-gap scenarios. This is realized through the proposed online quantile regression method for the calculation of the Prediction Intervals (PIs) and the designed nonparametric Online Interval Calibration (OIC) method that enhances the reliability and sharpness of the PIs. The effective uncertainty quantification capability enables the operators to make risk-aware decisions even when a large data gap exists.

II. METHODOLOGY

A. Problem Statement

Learning-based load forecasting models aim to capture

Manuscript received August 20, 2024; revised October 28, 2024; accepted December 25, 2024. Date of publication July 3, 2025; date of current version September 30, 2025. This work was supported in part by National Natural Science Foundation of China under Grant 72401055, in part by National Natural Science Foundation of China under Grant 52277083, and in part by the joint founding of Guangdong, and Dongguan under Grant 2023A15110939.

P. F. Zhao, W. H. Hu, and D. Cao (corresponding author, email: caodi@uestc.edu.cn) are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.

L. C. Dai is with the Electric Power Research Institute of State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750011, China.

Q. Huang is with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China. He is also with the Chinese Academy of Science and Technology for Development, Beijing 100038, China.

Z. Chen is with the Department of Energy Technology, Aalborg University, DK-9220 Aalborg, Denmark.

DOI: 10.17775/CSEEJPES.2024.06300

the mapping relationship between the specifically designed input features $\mathbf{x}_t = \{S_t, M_t, C_t\}$ at time t and the future electrical load $\hat{y}_{t+H}$, where H denotes the forecasting horizon. In general, the input features include past load observations $S_t = \{s_{t-D-J}, \dots, s_{t-D-1}, s_{t-D}\}$, auxiliary variables such as weather data M_t , and time information C_t . Here, J represents the length of history loads, D denotes the magnitude of the data gap. For the past load observation, existing load forecasting methods typically assume that the recent load consumption data are available. For example, when past electrical load data spanning from $t - D - J$ to $t - D$ are used for forecasting, the value of the data gap D is typically set to a small number. However, in real-world applications, there is a time gap between the data when power is consumed and when measurement values are obtained, which means that the value of the time gap may be very large in practice. Since the correlation between past observations and future consumption decreases with the increase of the time gap D , it would be challenging for traditional load forecasting methods to provide accurate predictions under such data gaps. To deal with the challenges caused by the data gap, this paper proposes an AEBP-enabled online learning-based probabilistic load forecasting method.

B. AEBP-Based Online Learning

In the presence of the data gap, the information source of historical loads is referred to as low-fidelity data due to the reduced correlations. On the other hand, weather data and calendar variables are easy to collect on the fly, categorizing these information sources as high-fidelity data. The overall framework of the proposed method is illustrated in Fig. 1. It can be seen that both high- and low-fidelity data are concatenated and passed through a dense layer. Subsequently, an attention mechanism is applied to fuse the high-fidelity data with the latent representations learned from the dense layer. The data then pass through several dense layers, each coupled with an output layer to generate a forecast quantile. The final prediction is the weighted sum of these forecast quantiles. The loss function is computed based on the predictions and

newly revealed load observations. During backpropagation, the parameters of each layer are updated using the gradients calculated from the corresponding loss and weight.

Specifically, the designed NN employs a dense layer initially to extract feature representations from both high- and low-fidelity data, as:

$$\mathbf{g} = \text{Dense}(S, M, C, \boldsymbol{\theta}_D) \quad (1)$$

where $\mathbf{g}$ denotes the outputs; $\boldsymbol{\theta}_D$ represents the parameters of the dense layer; S , M , and C are history loads, weather data, and time information, respectively. Following this, an attention module [7] is employed to adaptively fuse multi-fidelity data as:

$$\mathbf{att} = \text{Attention}(\mathbf{g}, M, C, \boldsymbol{\theta}_{\text{att}}) \quad (2)$$

where $\mathbf{att}$ denotes the outputs, and $\boldsymbol{\theta}_{\text{att}}$ represents the parameters of the attention module. An AEBP mechanism that can dynamically adjust the model's depth in an online learning environment is designed for NN. It treats each output layer as a quantile predictor, and the final forecast is the weighted sum of these predictors. Specifically, $\mathbf{att}$ is forwarded through L hidden dense layers, and each dense layer generates a feature representation, denoted as $\mathbf{h}^{(l)}$, for $1 \leq l \leq L$. Then, it proceeds to an output layer and produces a q -th quantile forecast value, denoted as $y_q^{(l)}$, for $1 \leq l \leq L$. The final q -th quantile prediction $\hat{y}_q$ is a weighted combination of $y_q^{(l)}$, denoted as $\hat{y}_q = \sum_{l=1}^L \alpha^{(l)} \hat{y}_q^{(l)}$, where $\alpha^{(l)}$ is a learnable parameter that represents the weight assigned for the l -th predictor.

The parameter $\alpha^{(l)}$ is updated using the Hedge algorithm [8]. Specifically, at the initial stage, all α values are uniformly distributed, i.e., $\alpha^{(l)} = \frac{1}{L}$. At the end of each time point, the prediction $\hat{y}_{q,t}$ is obtained, and the actual load measurement $y_{t'}$ can be revealed, where $t' = t - D - H + 1$. Then, the l -th layer loss $\mathcal{L}_{q,t}^{(l)}$ for q -th quantile is calculated by the pinball loss function:

$$\mathcal{L}_{q,t}^{(l)} = \begin{cases} (1-q)(\hat{y}_{q,t'}^{(l)} - y_{t'}), & \hat{y}_{q,t'}^{(l)} \geq y_{t'} \\ q(y_{t'} - \hat{y}_{q,t'}^{(l)}), & \hat{y}_{q,t'}^{(l)} < y_{t'} \end{cases} \quad (3)$$

Following this, the parameters of NN are updated at each time step using the newly arrived observations, as follows:

$$\begin{cases} \boldsymbol{\theta}_{\text{Out},t+1}^{(l)} \leftarrow \boldsymbol{\theta}_{\text{Out},t}^{(l)} - \eta \alpha_t^{(l)} \nabla_{\boldsymbol{\theta}_{\text{Out},t}^{(l)}} \mathcal{L}_{q,t}^{(l)} \\ \boldsymbol{\theta}_{\text{HD},t+1}^{(l)} \leftarrow \boldsymbol{\theta}_{\text{HD},t}^{(l)} - \eta \sum_{j=1}^L \alpha_t^{(j)} \nabla_{\boldsymbol{\theta}_{\text{HD},t}^{(j)}} \mathcal{L}_{q,t}^{(j)} \\ \boldsymbol{\theta}_{\text{att},t+1} \leftarrow \boldsymbol{\theta}_{\text{att},t} - \eta \alpha_t^{(1)} \nabla_{\boldsymbol{\theta}_{\text{att},t}} \mathcal{L}_{q,t}^{(1)} \\ \boldsymbol{\theta}_D,t+1 \leftarrow \boldsymbol{\theta}_D,t - \eta \alpha_t^{(1)} \nabla_{\boldsymbol{\theta}_D,t} \mathcal{L}_{q,t}^{(1)} \end{cases} \quad (4)$$

where $\boldsymbol{\theta}_{\text{HD}}^{(l)}$ denotes the weights of the l -th hidden layer; $\boldsymbol{\theta}_{\text{Out}}^{(l)}$ denotes the weights of the l -th output layer; η denotes the learning rate; $\alpha_t^{(1)}$ represents the weight assigned to the output of the first output layer; $\mathcal{L}_{q,t}^{(1)}$ denotes the loss function calculated based on the predictions from the first output layer and actual values. Note that the parameters $\boldsymbol{\theta}_{\text{HD},t+1}^{(l)}$ of each hidden dense layer are updated through the weighted sum of the error derivations of all hidden layers.

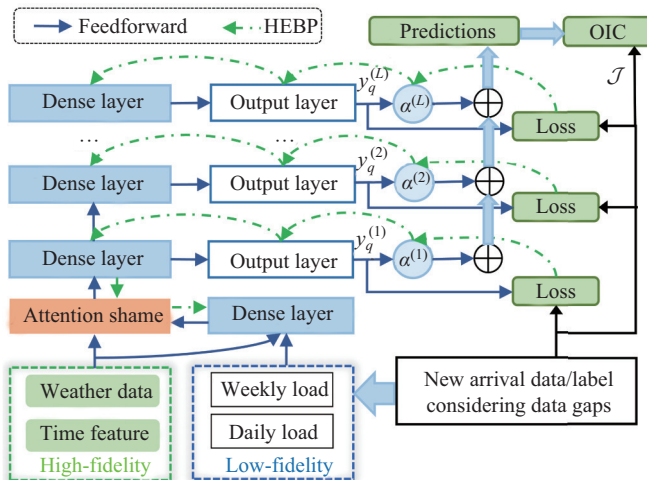


Fig. 1. A graphical illustration of the proposed method.

The weight of the l -th predictor is updated based on its loss $\mathcal{L}_{q,t}^{(l)}$ and a discount factor $\xi \in (0, 1)$, denoted as $\alpha_{t+1}^{(l)} \leftarrow \alpha_t^{(l)} \xi^{\mathcal{L}_{q,t}^{(l)}}$. Additionally, smoothing and normalization steps are conducted for $\alpha_{t+1}^{(l)}$ to prevent overfitting issues, denoted as $\alpha_{t+1}^{(l)} = \max(\alpha_{t+1}^{(l)}, \frac{s}{L})$ and $\alpha_{t+1}^{(l)} = \frac{\alpha_{t+1}^{(l)}}{Z_t}$, respectively. Here, $Z_t = \sum_{l=1}^L \alpha_{t+1}^{(l)}$. It can be observed that the weight for each predictor is dynamically adjusted to allocate the contribution of each layer. Therefore, the AEBP enables adaptive adjustment of the model's depth in an online learning setting, thus avoiding overfitting issues and maximizing the potential of the designed multi-layer NN.

Unlike traditional load forecasting methods trained offline on batch data, we propose solving the data gap issues via the online learning strategy. It updates parameters using new load information and generates forecasts with the most recent parameters, even during the testing procedure. As such, despite the data gap issue causing information to be missing in historical loads, *the online learning method can promptly track the newly arrived load information and pass on it to the subsequent learning and forecasting process*, thereby effectively filling the data gap.

C. Online Interval Calibration Method

A nonparametric OIC method is further developed to guarantee the reliability and sharpness of the PIs under data gaps. Given the Nominal Coverage Probability (NCP) $100(1 - \delta)\%$, the PI is computed as $\mathcal{C}_{\delta,t} = [\hat{\mathcal{Y}}_{\frac{\delta}{2},t}, \hat{\mathcal{Y}}_{1-\frac{\delta}{2},t}]$. Considering the data gaps, at the end of time t , the OIC set is revealed and indexed by $\mathcal{I} = \{t' - K, \dots, t' - 1, t'\}$, where K denotes the length of the OIC set. Then the residuals are calculated on the calibration set as:

$$\mathcal{R}_i = \max\{\hat{\mathcal{Y}}_{\frac{\delta}{2},i} - y_i, y_i - \hat{\mathcal{Y}}_{1-\frac{\delta}{2},i}\} \quad (5)$$

for each $i \in \mathcal{I}$. The calibration residual set can be interpreted as follows: if the actual value falls below the lower bound of the PI, $y_i < \hat{\mathcal{Y}}_{\frac{\delta}{2},i}$, then $\mathcal{R}_i = \hat{\mathcal{Y}}_{\frac{\delta}{2},i} - y_i$ represents the penalty for this uncovered mistake. Similarly, if the actual value exceeds the upper bound of the PI, the penalty is $\mathcal{R}_i = y_i - \hat{\mathcal{Y}}_{1-\frac{\delta}{2},i}$. Moreover, when the actual value lies within the PI, $\mathcal{R}_i$ is determined as the larger of two non-positive numbers: $\hat{\mathcal{Y}}_{\frac{\delta}{2},i} - y_i$ and $y_i - \hat{\mathcal{Y}}_{1-\frac{\delta}{2},i}$. Thus, the calibration residual set accounts for both undercoverage and overcoverage scenarios [9]. Finally, the PI after calibration is calculated as:

$$\mathcal{C}_{\delta,t}^* = [\hat{\mathcal{Y}}_{\frac{\delta}{2},t} - Q_{1-\delta}(\mathcal{R}), \hat{\mathcal{Y}}_{1-\frac{\delta}{2},t} + Q_{1-\delta}(\mathcal{R})] \quad (6)$$

where

$$Q_{1-\delta} = \mathcal{R}_{(\lceil (1-\delta)(|\mathcal{I}|+1) \rceil)} \quad (7)$$

where $\mathcal{R}_{(k)}$ denotes the k -th order statistic of the calibration residual set $\mathcal{R} = \{\mathcal{R}_i : i \in \mathcal{I}\}$ sorted in ascending order; and $\lceil \cdot \rceil$ denotes the ceiling operator. The scaling factor $(|\mathcal{I}| + 1)$ corresponds to the conformal correction, which ensures valid marginal coverage at the confidence level $1 - \delta$ without relying on asymptotic assumptions. The detailed implementation of the proposed online learning method is outlined in Algorithm 1.

Algorithm 1: AEBP-Enabled Online Learning

Input: Input features $\mathbf{x}_t = \{S_t, M_t, C_t\}$, discount factor ξ , smoothing factor s , length K of OIC set, significance level δ , quantile set $\{1, 2, \dots, q, \dots, Q\}$

Output: Prediction interval $\mathcal{C}_{\delta,t}^*$

- 1 Initialize neural network parameters $\theta_D, \theta_{\text{att}}, \theta_{\text{HD}}^{(l)}, \theta_{\text{Out}}^{(l)}$, $\alpha^{(l)} = \frac{1}{L+1}, \forall l = 1, \dots, L$
- 2 **for** $t = 1, \dots, T$ **do**
- 3 Receive input features $\mathbf{x}_t$;
- 4 Predict q -th quantile $\hat{\mathcal{Y}}_{q,t} = \sum_{l=1}^L \alpha_t^{(l)} \hat{y}_{q,t}^{(l)}$, for q in quantile set;
- 5 Construct prediction interval $\mathcal{C}_{\delta,t} = [\hat{\mathcal{Y}}_{\frac{\delta}{2},t}, \hat{\mathcal{Y}}_{1-\frac{\delta}{2},t}]$;
- 6 Reveal OIC set and indexed by $\mathcal{I}$;
- 7 Compute residuals $\mathcal{R}$ from (5);
- 8 Calculate calibrated prediction interval $\mathcal{C}_{\delta,t}^*$ based on (6) and (7);
- 9 Obtain new load observation $y_{t'}$;
- 10 Calculate loss function $\mathcal{L}_{q,t}^{(l)}$ according to (3);
- 11 Update $\theta_{D,t+1}, \theta_{\text{att},t+1}, \theta_{\text{HD},t+1}^{(l)}, \theta_{\text{Out},t+1}^{(l)}, \forall l = 1, \dots, L$ based on (4);
- 12 Update $\alpha_{t+1}^{(l)} \leftarrow \alpha_t^{(l)} \xi^{\mathcal{L}_{q,t}^{(l)}}$, $\forall l = 1, \dots, L$;
- 13 Smoothing $\alpha_{t+1}^{(l)} = \max(\alpha_{t+1}^{(l)}, \frac{s}{L})$;
- 14 Normalize $\alpha_{t+1}^{(l)} = \frac{\alpha_{t+1}^{(l)}}{Z_t}$, where $Z_t = \sum_{l=1}^L \alpha_{t+1}^{(l)}$
- 15 **end**

III. CASE STUDY

A. Experimental Setups

Real load data from New South Wales (NSW), Australia, and Arizona State, USA¹, are collected for the simulation. The former dataset comprises aggregated loads from various households in distributed systems, spanning from 2018 to 2020, with a one-hour resolution. The latter consists of hourly load data from the Tempe campus of Arizona State University, covering the period from 2017 to 2019. For the purpose of this study, we assume that these two datasets exhibit data gap issues.

Comparative tests are carried out among various baselines, including: 1) Persistent model, where the averaged values of historical same-time loads collected under data gaps are set as the predictions; 2) Improved Quantile Regression Neural Network (iQRNN) [3]; 3) Long Short-Term Memory (LSTM) Neural Network [10]; 4) Sparse Variational Gaussian Process (SVGP) [11]; 5) Fully Parameterized Sequence to Quantile Regression (FPSeq2Q) [12]; 6) Complex Neural Network (ComNN) [4]; 7) Online-Offline Deep Kernel Learning (ODKL) [13]; 8) Online Contrastive Learning of Disentangled Seasonal-Trend Representations (OCOSt) [14]. These methods are trained, validated, and tested on 60%, 20%, and 20% samples of the datasets, respectively. The hyperparameters of the above benchmarks are initially set according to the recommended values in the corresponding references and then fine-tuned via the grid search method. The hyperparameters of the proposed method are pre-set based on expert knowledge and subsequently also fine-tuned using the grid search method. The final hyperparameters of various methods are listed in Table I. 1-hour-ahead load forecasting is investigated in this

¹<http://cm.asu.edu/>

TABLE I
MAIN HYPERPARAMETERS OF VARIOUS METHODS

Models	Main Hyperparameters
iQRNN	Units of fully connected layer (FCL) 1: 400; units of FCL2: 400; learning rate: 0.0001
LSTM	Units of LSTM layer: 50; units of FCL: 32; learning rate: 0.001
SVGP	Kernel: Squared Exponential; inducing points: 100
FPSeq2Q	Attention head: 4; encoder layer: 4; decoder layer: 4; hidden dimension: 32; dimension of cosine embedding: 32; learning rate: 0.001
ComNN	Units of each FCL: 50; convolutional kernel size: 3; output filter: 32; learning rate: 0.001
ODKL	Spiking neuron network (SNN) layer: 2; units of each SNN: 64; Size limit of basis vector set: 100; learning rate: 0.0005
OCOST	Number of DilatedConv block: 10; kernel size: 3; channel size: 64; units of FCL: 64; representation dimension: 320; learning rate: 0.001
Proposed	Dense layer: 6; units of each dense layer: 50; discount factor ξ : 0.99; smoothing factor s : 0.2; length of OIC set: 24; learning rate: 0.001

study. Each experiment is repeated five times, and the mean values along with Standard Deviations (SDs) are reported.

B. Performance on Deterministic Forecasting

The Mean Absolute Percentage Errors (MAPEs) and SDs obtained by various methods with data gaps set to 0 (no data gap), 0.5, 1, 3, 7, 14, and 30 days for the Arizona and NSW datasets are listed in Tables II and III, respectively. From the tables, it is evident that most forecasting methods deliver satisfactory performance in the absence of data gaps. Owing to the reduced correlations between the past observations and future power consumption, a significant accuracy decline is observed for the Persistent method when 12-hour and one-day data gaps are considered. Its performance becomes even worse when the data gap is increased to 7, 14, and 30 days. This illustrates the huge negative impact of the data gap on the performance of the forecasting models. Machine learning

methods, including iQRNN, LSTM, SVGP, FPSeq2Q, and ComNN, achieve better performance than that obtained by the Persistent method under various conditions. However, the above methods optimize their parameters using the batch of historical data in an offline manner. The values of newly arriving load consumption data cannot be fully exploited. Therefore, similar performance degradations are observed for these machine learning methods when the data gap exists. The deficiency of traditional learning-based load forecasting methods in dealing with the data gap issue is illustrated here. In contrast to offline learning methods, the online learning approaches ODKL and OCoST utilize newly arrived load data to update model parameters and incorporate this information into subsequent learning, thereby alleviating the negative effects of data gaps. However, these two methods maintain a fixed model depth during the learning process, where a model that is too complex may converge slowly, while a model that is too simple may restrict learning capability. The proposed method, AEBP, differentiates itself by enabling adaptive adjustment of the model's depth in an online learning setting, thus nicely combining the benefits of online learning and deep learning. By fully exploiting available data and reducing the impact of data gaps, the proposed method experiences only slight performance degradation when data gaps are considered. Its MAPEs achieved under data gap outperform that obtained by the Persistent, iQRNN, LSTM, SVGP, FPSeq2Q, ComNN, ODKL, and OCoST methods by 54.90%, 48.10%, 49.52%, 47.04%, 41.38%, 30.55%, 24.48%, and 14.85%, respectively. These results demonstrate the superiority of the proposed method in dealing with the data gap.

In terms of predictive stability, the SDs of the Persistent and SVGP models are zero due to their nonparametric nature. The SDs of offline learning-based parametric models—iQRNN, LSTM, FPSeq2Q, and ComNN—are relatively high, influenced by data gap issues. Online learning methods, such as

TABLE II
MAPEs (%) AND SDs (%) OBTAINED BY VARIOUS METHODS ON ARIZONA DATA

Data gap	0 days	12 hours	1 day	7 days	14 days	30 days
Persistent	4.72±0.00	7.03±0.00	7.11±0.00	9.37±0.00	11.73±0.00	16.52±0.00
iQRNN [3]	3.33±0.12	7.72±0.17	8.91±0.38	10.52±0.43	11.70±0.22	11.45±0.20
LSTM [10]	3.04±0.19	6.93±0.85	7.15±0.91	10.21±1.39	10.35±1.22	12.59±1.06
SVGP [11]	2.65±0.00	5.51±0.00	6.23±0.00	8.71±0.00	9.42±0.00	10.31±0.00
FPSeq2Q [12]	2.53±0.14	5.30±0.44	6.01±0.31	8.22±0.52	9.60±0.33	11.22±0.53
ComNN [4]	2.47±0.15	4.17±0.24	4.74±0.94	6.24±1.18	7.89±0.43	9.30±0.41
ODKL [13]	2.94±0.08	4.06±0.13	4.82±0.08	5.57±0.09	6.84±0.14	8.69±0.20
OCoST [14]	2.41±0.10	3.44±0.11	4.03±0.24	4.60±0.13	6.61±0.12	7.74±0.22
Proposed	2.86±0.07	3.21±0.14	3.32±0.06	3.95±0.10	4.42±0.10	6.25±0.19

TABLE III
MAPEs (%) AND SDs (%) OBTAINED BY VARIOUS METHODS ON NSW DATA

Data gap	0 days	12 hours	1 day	7 days	14 days	30 days
Persistent	11.96±0.00	12.14±0.00	12.28±0.00	14.11±0.00	15.09±0.00	20.92±0.00
iQRNN [3]	5.43±0.20	9.37±0.36	10.16±0.28	11.87±0.35	12.32±0.32	13.34±0.31
LSTM [10]	7.88±0.74	10.61±0.96	11.34±1.10	12.56±0.83	13.20±1.05	15.28±0.97
SVGP [11]	6.27±0.00	10.93±0.00	11.70±0.00	13.25±0.00	14.88±0.00	16.14±0.00
FPSeq2Q [12]	6.33±0.25	8.85±0.48	9.41±0.60	11.35±0.44	11.94±0.52	13.93±0.63
ComNN [4]	5.18±0.18	7.58±1.02	8.14±0.33	10.24±0.56	11.31±0.94	12.21±0.74
ODKL [13]	6.48±0.03	7.10±0.18	7.24±0.23	8.93±0.11	9.74±0.12	11.70±0.41
OCoST [14]	5.15±0.06	5.63±0.16	6.90±0.12	8.05±0.13	9.33±0.14	10.93±0.26
Proposed	5.09±0.04	5.57±0.11	6.20±0.09	7.43±0.20	7.50±0.05	8.84±0.31

ODKL and OCoST, reduce SDs by incorporating new load information to alleviate the data gap. The proposed method further improves stability, achieving an average SD of 0.12, demonstrating that the predictive stability of the proposed method is at a high level.

C. Impact Analysis of Different Functional Modules

Comparative tests are performed among several baselines to assess the effectiveness of our designed functional modules: 1) M1: Proposed+offline learning, which updates the parameters of our designed NN in an offline learning manner. 2) M2: Proposed+OGD, where traditional Online Gradient Descent (OGD) is used for updating our model's parameters. 3) M3: FCL+AEBP+OL, where the attention-based fusion module in our proposed method is replaced with several FCLs. 4) M4: LSTM+AEBP+OL, where all attention layers and dense layers in the proposed model are replaced with LSTM layers. 5) M5: Attention+AEBP+OL, where all dense layers in the model are replaced with attention layers.

The MAPEs for various methods when the data gap is set to 12 hours and 7 days are listed in Table IV. Since the network in M1 is trained in an offline learning manner, it struggles to capture the dynamic consumption pattern changes using newly arriving load data. Therefore, its performance is influenced by the data gap and is worse than that of the online learning-based proposed method. Since the M2 cannot dynamically adjust the model's capacity in an online learning environment, its MAPEs are higher than that obtained by the proposed method on both datasets. When the attention network is removed, it is difficult for FCL to fuse the multi-fidelity data effectively. As a result, the performance obtained by the M3 is worse than that of the proposed method. Although the LSTM and attention mechanism in M4 and M5 can fuse multi-fidelity data and achieve better performance than FCNN, an excessive number of LSTM and attention layers increases model complexity. This leads to slower convergence in online learning, ultimately resulting in lower accuracy than our model. These results demonstrate the effectiveness of the designed functional modules.

TABLE IV
MAPEs (%) OF DIFFERENT MODELS

Dataset	Gap	M1	M2	M3	M4	M5	Pro.
Arizona	12 hours	5.02	4.03	3.85	3.30	3.24	3.21
	7 days	5.48	4.75	4.67	4.34	4.20	3.95
NSW	12 hours	8.40	7.69	6.74	6.45	5.32	5.57
	7 days	9.93	8.61	8.40	8.18	7.85	7.43

D. Performance on Probabilistic Forecasting

We employ three metrics, namely Prediction Interval Coverage Probability (PICP) [15], Prediction Interval Normalized Average Width (PINAW) [15], and Quantile Score (QS) [3] to evaluate the uncertainty quantification of the proposed probabilistic forecasting method. Since the empirical coverage probability should be close to the NCP, PICP values with smaller deviations from the NCP indicate better coverage performance [16]. A smaller PINAW indicates a sharper PI. QS simultaneously evaluates the reliability and sharpness of

a PI. A smaller QS value represents a better forecasting outcome. When the data gap is set to 7 days, the evaluation metrics with a 90% NCP obtained by various probabilistic load forecasting methods on two datasets are listed in Table V. It can be seen that various offline learning methods tend to generate unreliable PIs, with the PICP obtained by iQRNN, FPSeg2Q, SVGP, and ComNN methods being lower than 87% in most cases. Additionally, wider PIs are observed for these methods owing to the impact of the data gap. These findings indicate the adverse impact caused by the data gap on the reliability and sharpness of the outcomes achieved by the conventional probabilistic forecasting methods. The online learning method ODKL achieves better PICP performance but with a very wide PI. In contrast, the proposed method without OIC achieves better performance in terms of PICP and QS metrics. However, it tends to be overcovered, leading to a larger PINAW on the NSW dataset. The designed OIC method further improves the PI quality of the proposed method by considering both undercoverage and overcoverage scenarios to calibrate the PIs. It achieves the best probabilistic forecasting performance on all metrics. It is worth noting that the PICP after calibration is very close to NCP, indicating that the PIs provided by the OIC method are reliable for practical operations and benefit for uncertainty-aware decisions. These results demonstrate the effectiveness of the proposed OIC-enabled method in uncertainty quantification considering the impact of the data gap.

TABLE V
PROBABILISTIC FORECASTING PERFORMANCE OF VARIOUS METHODS

Dataset	Arizona			NSW		
	PICP	PINAW	QS	PICP	PINAW	QS
iQRNN [3]	83.17%	0.196	0.0161	80.60%	0.167	0.0486
SVGP [11]	84.17%	0.207	0.0157	83.20%	0.160	0.0433
FPSeg2Q [12]	86.38%	0.203	0.0151	87.35%	0.172	0.0420
ComNN [4]	84.64%	0.190	0.0143	85.17%	0.154	0.0408
ODKL [13]	92.85%	0.218	0.0128	93.27%	0.185	0.0392
Proposed	93.26%	0.188	0.0109	93.13%	0.170	0.0310
Proposed+OIC	89.26%	0.168	0.0094	89.00%	0.146	0.0276

IV. CONCLUSION

This letter proposed a novel NN structure with an AEBP-based online learning method to address data gap issues in load forecasting. It was further extended to a nonparametric OIC method to ensure the reliability and sharpness of the PIs. Simulation results on two real load datasets demonstrated that the proposed method can provide more accurate deterministic load forecasts as well as more reliable probabilistic forecasting outcomes compared to several state-of-the-art load forecasting methods.

Since collecting datasets with data gaps from open sources is challenging, our case study is based on the assumption that such gaps exist in the collected datasets from Arizona and NSW. In future work, we intend to validate our model on real-world distributed systems where the data gap is actually present.

REFERENCES

- [1] H. Zhang, W. Hu, D. Cao, Q. Huang, Z. Chen and F. Blaabjerg, "A

Temporal Convolutional Network Based Hybrid Model for Short-Term Electricity Price Forecasting,” *CSEE Journal of Power and Energy Systems*, vol. 10, no. 3, pp. 1119–1130, May 2024.

- [2] Y. Lei, L. Yu, J. Zhao, S. Ding, X. Guo and H. Bao, “Segmented Forecasting of Electric Load Under Pandemic Period Based on the ESD-ABiLSTMQR Method,” *CSEE Journal of Power and Energy Systems*, vol. 11, no. 5, pp. 2467–2476, Sep. 2025.
- [3] W. J. Zhang, H. Quan, and D. Srinivasan, “An improved quantile regression neural network for probabilistic load forecasting,” *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4425–4434, Jul. 2019.
- [4] P. F. Zhao, D. Cao, W. H. Hu, Y. H. Huang, M. Hao, Q. Huang, and Z. Chen, “Geometric loss-enabled complex neural network for multi-energy load forecasting in integrated energy systems,” *IEEE Transactions on Power Systems*, vol. 39, no. 4, pp. 5659–5671, Jul. 2024.
- [5] IoT Analytics. (2024, Feb. 21). Smart electricity meter market 2024: Global adoption landscape. [Online]. Available: <https://iot-analytics.com/smart-meter-adoption/>
- [6] M. Karimzadeh, G. Zavaliagkos, and K. Panetta, “Feature extraction and data gap analysis of US residential electricity consumption,” in *2022 IEEE 7th International Energy Conference (ENERGYCON)*, 2022, pp. 1–6.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 6000–6010.
- [8] D. Sahoo, Q. Pham, J. Lu, and S. C. H. Hoi, “Online deep learning: learning deep neural networks on the fly,” in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 2660–2666.
- [9] Y. Romano, E. Patterson, and E. J. Candes, “Conformalized quantile regression,” in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019, pp. 3543–3553.
- [10] W. C. Kong, Z. Y. Dong, Y. W. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-term residential load forecasting based on lstm recurrent neural network,” *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [11] H. L. Wen, J. H. Ma, J. Gu, L. Yuan, and Z. J. Jin, “Sparse variational gaussian process based day-ahead probabilistic wind power forecasting,” *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 957–970, Apr. 2022.
- [12] A. Faustine and L. Pereira, “FPSeq2Q: Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates,” *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2440–2451, May 2022.
- [13] Y. Z. Li, F. S. Zhang, Y. Liu, H. L. Liao, H. T. Zhang, and C. Chung, “Residential load forecasting: An online-offline deep kernel learning method,” *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4264–4278, Mar. 2024.
- [14] G. Woo, C. H. Liu, D. Sahoo, A. Kumar, and S. Hoi, “CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting,” arXiv preprint arXiv: 2202.01575, 2022.
- [15] A. Bracale, P. Caramia, P. De Falco, and T. Hong, “Multivariate quantile regression for short-term probabilistic load forecasting,” *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 628–638, Jan. 2020.
- [16] Z. J. Cao, C. Wan, Z. J. Zhang, F. R. Li, and Y. H. Song, “Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting,” *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1881–1897, May 2020.



research interests include forecasting technologies and the applications of artificial intelligence in modern power systems.

Pengfei Zhao received the B.E. degree in Electrical Engineering and Its Automation, as well as the M.S. degree in Electrical Engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 2019 and 2022, respectively. From 2022 to 2023, he served as an Engineer at the Electric Power Research Institute of State Grid Ningxia Electric Power Co., Ltd. He is currently working toward the Ph.D. degree in the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China. His



System Research Program at the same department. His research interests include artificial intelligence in modern power systems and renewable power generation. He has led/participated in more than 15 national and international research projects, and he has more than 200 publications in his technical field. He is currently serving as the Chair for IEEE Chengdu Section PELS Chapter. He is an IEEE Senior Member and IET Fellow.

Weihao Hu received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in Electrical Engineering, and Ph.D. degree from Aalborg University, Denmark, in 2012. He is currently a Full Professor and the Director of the Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China (UESTC). He was an Associate Professor at the Department of Energy Technology, Aalborg University, Denmark, and the Vice Program Leader of the Wind Power



Di Cao received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC) in 2021. He is currently an Associate Professor at UESTC. His research interests include optimization of distribution networks and applications of machine learning in power systems.



Longcheng Dai received the M.S. degree from the Shenzhen University, Shenzhen, China, in 2022. He is currently an Engineer at the Electric Power Research Institute of State Grid Ningxia Electric Power Co., Ltd. His research interests include HV circuit breaker, vacuum switch, and load frequency control.



Chinese Academy of Science and Technology for Development and the Director with the Sichuan State Provincial Laboratory of Power System Wide-area Measurement and Control. His current research and academic interests include power system instrumentation, power system monitoring and control, and informatics for smart electric energy systems.

Qi Hunag received the B.S. degree in Electrical Engineering from Fuzhou University, Fuzhou, China, in 1996, the M.S. degree from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree from Arizona State University, Tempe, AZ, USA, in 2003. He was the Vice President of the Chengdu University of Technology (CDUT) and President of the Southwest University of Science and Technology (SWUST). He is currently a Professor at the University of Electronic Science and Technology of China (UESTC). He is the Director with the



He has led many research projects and has more than 500 publications in his technical fields. His research interests include power systems, power electronics, and electric machines, and his main current research interests include wind energy and modern power systems. Dr. Chen was an Editor for the IEEE Transactions on Power Systems, an Associate Editor for the IEEE Transactions on Power Electronics, a Fellow of the Institution of Engineering and Technology, and a Chartered Engineer in the U.K.

Zhe Chen received the B.Eng. and M.Sc. degrees from the Northeast China Institute of Electric Power Engineering, Jilin, China, in 1982 and 1986, and the Ph.D. degree from the University of Durham, U.K., in 1997. He is a Full Professor with the Department of Energy Technology, Aalborg University, Denmark. He is the Leader of the Wind Power System Research Program at the Department of Energy Technology, Aalborg University, and the Danish Principle Investigator for Wind Energy of the Sino-Danish Centre for Education and Research.