scientific reports

Check for updates

OPEN

Adaptive electricity consumption forecasting approach for universal environments

Shiqi Zhou[1,2], Saisai Ni[1,2], Yifeng Han[1], Zhekang Dong[1,2]✉ & Chun Sing Lai[3]

The development of an accurate electricity consumption forecast model is crucial for stable operation and intelligent management of power systems. Traditional methods often overlook user heterogeneity and lack measures to address concept drift caused by distribution changes in electricity data over time. We propose an adaptive electricity consumption probability forecasting method tailored to universal environments. The method includes a nonmonotonic correlation elimination-based recursive feature selection that adaptively determines the optimal feature combination. Our model incorporates a joint loss function combining point and probability forecasting evaluations to accurately quantify online batch errors. It also features a buffer to store batch data showing pattern changes and dynamically adjusts weights to counteract concept drift. We validated our method, adaptive electricity consumption forecast for universal environments (AECF-UC), against some mainstream methods using a multi-environment dataset. Comparative and ablation experiments show that AECF-UC outperforms others, achieving average RMSE, pinball loss and CRPS of 0.3041, 0.0567 and 0.1683 respectively, with the joint loss method improving prediction accuracy by about 6% over the single-loss method. These results indicate that the proposed method exhibits certain advantages in universality and adaptability.

Keywords Electricity consumption forecasting, Concept drift, Probability forecasting, Hidden Markov model

Energy consumption has surged with the rapid economic development of various countries[1,2]. Electricity consumption forecast is essential for understanding its variation trends, facilitating the rational allocation of power resources, and ensuring the safety and economic efficiency of both the generation and consumption processes[3,4]. Consequently, research on electricity consumption forecasts has attracted widespread attention from scholars worldwide.

Accurate electricity consumption forecast requires complete datasets and suitable forecasting models. At the dataset level, the integration of smart meters in different power systems enables the construction and development of complete electricity consumption datasets[5,6]. At the model level, due to the variability between regions and buildings, accurate multiuser forecasting is more challenging than single-user forecasting. Namely, more instabilities and uncertainties are encountered when implementing multiuser electricity consumption forecast[7,8]. Moreover, nonlinear and nonstationary electricity consumption data always make traditional forecasting methods ineffective[9]. To address these issues, data-driven forecasting methods are widely used. These methods aim to provide effective solutions to address the volatility of electricity consumption data by continuously receiving new data. According to the different forms of prediction results, data-driven electricity consumption forecast methods can be broadly categorized into two types[10]: point forecasting and probabilistic forecasting. Specifically, point forecasting methods take a single value as the predicted electricity consumption value. This forecasting technology is widely used and relatively mature[11–16]. However, as electricity consumption patterns become more uncertain or variable, electricity consumption demand becomes difficult to accurately forecast by using only point forecasting methods. Consequently, an increasing number of scholars are focusing their research on probabilistic forecasting. Different from those of point forecasting, the outputs of probabilistic forecasting are probability density functions or confidence intervals of electricity consumption. Compared with point forecasting, probabilistic forecasting provides the possible distribution of future electricity consumption and effectively evaluates the uncertainties of forecasting[17].

[1]College of Electronic Information, Hangzhou Dianzi University, Hangzhou 310018, China. [2]Zhejiang Provincial Key Laboratory of Intelligent Vehicle Electronics Research, Hangzhou 310018, China. [3]Department of Electronic and Computer Engineering, Brunel University, London UB8 3PH, UK. ✉email: englishp@hdu.edu.cn

a. Julian et al.[18] proposed an online probabilistic load forecasting method that considers temporal correlations and is suitable for producing multivariate forecasts. Lin et al.[19] proposed a dual-stage attention-based long short-term memory (LSTM) network for probabilistic forecasting, and a feature attention-based encoder was built in the first stage. A. Bracale based on multivariate quantile regression (MQR), which refines the probabilistic forecasts of individual electricity consumption online[20]. D. Salinas et al.[21] proposes DeepAR, a methodology for producing accurate load probabilistic forecasts, based on training an autoregressive recurrent neural network model on many related time series.
b. A. Faustine et al.[22] presented a parameterized quantile regression approach for short-term probabilistic load forecasting at the distribution level, along with a scoring metric proposed to assess the issue of robustness when applied in real time. C. Wang et al.[23] proposed a Bayesian neural network called the Bayesian decoder transformer (BNNDeT), which achieves joint probabilistic forecasting for multi-energy loads by considering their complex coupling relationships and related uncertainties. S. Zhou et al.[24] proposed a multi-dimensional adaptive short-term forecasting method for electrical load based on Bayesian Autoformer network (BSAuto), which can address the issues of data volatility and model uncertainty. The hidden Markov model (HMM) structure is adept at capturing potential state changes and probabilistic transfer relationships in sequences for probabilistic inference purposes. It is widely used in probabilistic forecasting tasks due to its flexibility and scalability[25–30]. Among these, V. Álvarez et al.[30] proposed an adaptive probabilistic load forecasting (APLF) method based on HMM. It can adapt to dynamic changes in electricity consumption data through online learning and accurately assess the uncertainty of electricity consumption. This approach has exhibited great superiority in electricity consumption forecasts. However, certain limitations are still encountered when applying the abovementioned methods to multi-environment forecasting scenarios.

User heterogeneity: In electricity consumption forecast tasks, heterogeneity refers to the diversity of electricity consumption patterns resulting from differences among users in aspects such as their living habits, residential environments, and economic conditions. This diversity necessitates that forecasting models adapt to the specific characteristics of different users to accurately predict their electricity consumption.

Concept drift: The distribution of electricity consumption data changes over time, and the phenomenon of concept drift inevitably occurs[31]. If the utilized forecasting methods are not sufficiently flexible for addressing such a dynamic data distribution, its accuracy cannot be guaranteed.

Some of the electricity consumption forecast methods are summarized in Table 1 attempt to address the challenges of user heterogeneity and concept drift. However, despite their efforts, the existing methods still struggle to fully address these challenges, particularly in multi-environment. Therefore, this work proposes an adaptive electricity consumption forecast method with good generalizability for use in cases with different user environments and time dimensions. The main contributions of this study can be summarized as follows.

- To address the problem of user heterogeneity, a nonmonotonic correlation-based recursive method is designed. This method selects the optimal observation features for users and makes the model universal.
- To address the problem of concept drift, we add a buffer module based on a joint loss function. This module enables the model to quickly adapt to the given data distribution and improves the comprehensive performance of the prediction method.
- To improve the performance of the model, and address generalization capability, multi-environment datasets are conducted some analysis (e.g., a heterogeneity analysis and ablation analysis).

The remainder of this paper is organized as follows. "Load data analysis" describes the composition of the utilized datasets and analyses the relationships among multi-environment loads and their uncertainties. "Design and Implementation" introduces the proposed model and elaborates on the specific design and implementation

Methods	Forecasting		Brief description	Addressed issue	
	Point	Probabilistic		Heterogeneity	Drift
OARNN	√		An electricity consumption forecasting capable of continuously learning from newly arriving data and adapting to new patterns		√
GNN	√		A load forecasting model based on spatiotemporal attention convolutional mechanism	√	
TFT	√		A novel attention-based load forecasting model that combines high-performance multi-horizon forecasting		√
QLSTM		√	A probabilistic Electricity consumption forecasting for individual consumers using pinball loss guided LSTM	√	
DeepAR		√	A probabilistic forecasting model based on training an autoregressive recurrent neural network model on many related time series		√
MQR		√	A multivariate quantile regression forecasting that refines probabilistic forecasting of individual electricity consumptions online		√
BNNDeT		√	An electricity consumption forecasting based on multi-task BNN that extracts data features through Bayesian multiple encoder	√	
BSAuto		√	A multi-dimensional adaptive short-term forecasting method for electrical load based on Bayesian Autoformer network		√
APLF		√	A probabilistic electricity consumption forecasting based on the adaptive online learning of hidden Markov models		√
This work		√	An adaptive approach for electricity consumption forecasting in universal environments	√	√

Table 1. Descriptions of the related electricity consumption forecasting methods.

Sub-dataset	Group	Number of users	Contained information	Number of timestamps	Frequency
Individual	Std	300	EC, EP, and WI	17,520	Every 30 min
	DTou	200			
Campus	Downtown	2	EC and WI	8760	Every 1 h
	Polytechnic	2			
Region	Dayton	1	EC and WI	8760	Every 1 h
	New England	9			

Table 2. Basic information of the multi-environment datasets. EC∀, electricity consumption; EP∀, electricity price; WI∀, weather information (including temperature, pressure, dewpoint, humidity, and wind bearing.

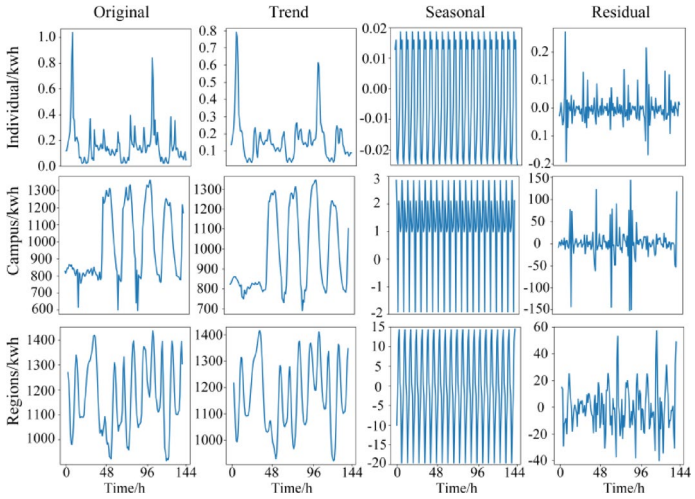


Fig. 1. Trend, seasonal and remainder components of multi-environment electricity consumption in 1 week.

process. For verification purposes, a series of comparative experiments and ablation analyses are conducted in "Experiments and results". Finally, "Conclusion" draws the conclusion.

Load data analysis

Datasets

The multi-environment dataset is summarized and described in Table 2. It comprises three sub-datasets: an individual dataset (containing electricity consumption, electricity price, and weather information data), a campus dataset (containing electricity consumption and weather information data), and a region dataset (containing electricity consumption and weather information data). Specifically, the individual sub-datasets can be further divided into standard (Std) groups and dynamic time-of-use (DTou) groups. The data are derived from the Low Carbon London project performed by UK Power Networks[31]. The campus dataset is also divided into Downtown Campus and Polytechnic Campus datasets, and the data are obtained from the campus system of Arizona State University in the United States[32]. The region dataset consists of load demand data from Dayton and New England, which are separately provided by PJM Interconnection and the ISO-NE organization, respectively[33,34]. The datasets generated and analysed during the current study are available in the https://data.london.gov.uk/blog/electricity-consumption-in-a-sample-of-london-households/, https://cm.asu.edu, and https://www.iso-ne.com repository.

Heterogeneity analysis

The electricity consumption data can be decomposed into trend, seasonal and remainder components using the seasonal-trend decomposition with loess (STL) method[35]. This method is used to decompose the electricity consumption of multiple environments in 1 week, and each decomposed component is shown in Fig. 1.

For individual electricity consumption, the trend component displays fluctuations with sharp peaks, indicating irregular short-term usage patterns. In contrast, the campus-level consumption shows clear and relatively stable periodic trends, highlighting regular daily electricity usage behaviors. Regional-level data exhibit broader, smoother fluctuations, suggesting aggregated and averaged usage patterns over larger user groups, thus reflecting less abrupt and more gradual changes. The differences in the data distributions across the environments are evident in their trend, seasonal, and residual components, especially in the trend and residual components. These differences indicate the heterogeneity between the electricity load data.

To ascertain these differences, an analysis of variance (ANOVA) is utilized. It can assess the disparities among the means of the multi-environmental electricity consumption data[36,37]. In the ANOVA, to quantify the differences between groups, the F statistic is calculated by

$$F = \frac{(N-k)\sum_{i=1}^{k} n_i \left(\bar{X}_i - \bar{X}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_i\right)^2} \tag{1}$$

where F is the F statistic, N represents the total sample size, k denotes the number of groups, n_i is the amount of data in the ith group, X_{ij} is the jth data point of the ith group, $\bar{X}_i$ is the mean of the ith data group, and $\bar{X}$ is the overall population mean. The F statistic for the multi-environment electricity consumption data is shown in Fig. 2. The red blocks indicate smaller values, and the blue and green blocks indicate higher values. Most of the F statistics are represented by blue and green, indicating that variations in the means are prevalent across the multi-environmental data. Subsequent hypothesis testing is conducted to determine the significance of the differences. With the significance level set at 0.05, 90% of the P values are below this threshold, which indicates substantial heterogeneity among the data.

Concept drift analysis

The proposed buffer module addresses the issue of concept drift arising from variations in the data distribution. In this work, three existing detection methods are applied: the drift detection method (DDM)[38], an adaptive window (ADWIN)[39] and the Page–Hinkley (PH) method[40]. These methods are utilized collectively to assess the presence of concept drift within the experimental data. Figure 3 illustrates the joint evaluation results obtained by applying these concept drift detection methods to the electricity consumption data contained in the multi-environment dataset. The top graph shows the number of drifts that occur during the entire detection process. To provide a more tangible understanding of concept drift, we randomly select eight households, and their specific detection outcomes are illustrated in the subsequent eight graphs. The instances of detected concept drift are marked by colored vertical lines corresponding to each method, while the gray rectangular areas denote the drift warning zones identified by the DDM. In Fig. 3a, significant changes can be seen in the data distribution. The drifts occur at timestamps 2600, 7600, and 12,400, and all three methods detect drift around these points. The three methods detect different amounts of concept drift not only in different time series but also in different locations within the same series. It indicates that concept drift widely exists in electricity consumption data.

Design and implementation

To address the problems of user heterogeneity and concept drift, this paper proposes an adaptive electricity consumption forecast method for universal environments (AECF-UC). The framework of AECF-UC, which includes data flow, feature adaptation, learning and prediction, and joint buffer design modules, is illustrated in Fig. 4. The specific contents of each component are described below.

Data flow

AECF-UC forecasts the probability distribution of multi-environmental electricity consumption data on the next day online, with a resolution of 1 h. The temporal characteristics, denoted by $m(t)$, indicate the hour of the day for weekdays (ranging from 1 to 24), weekends and holiday (ranging from 25 to 48) and serve as position

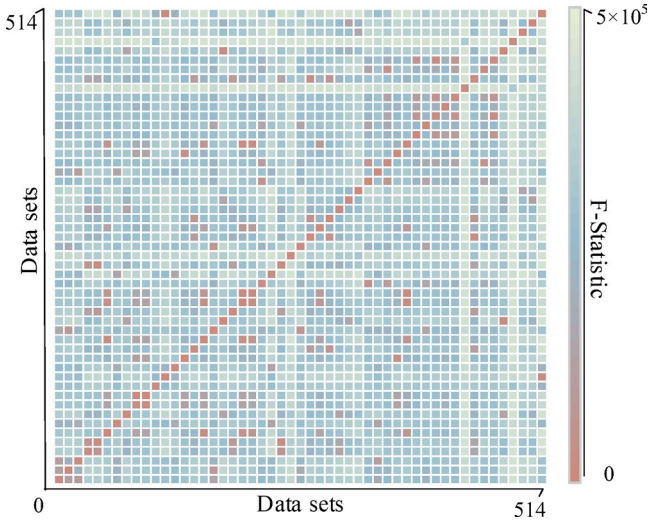


Fig. 2. F statistic for a multi-environment dataset.

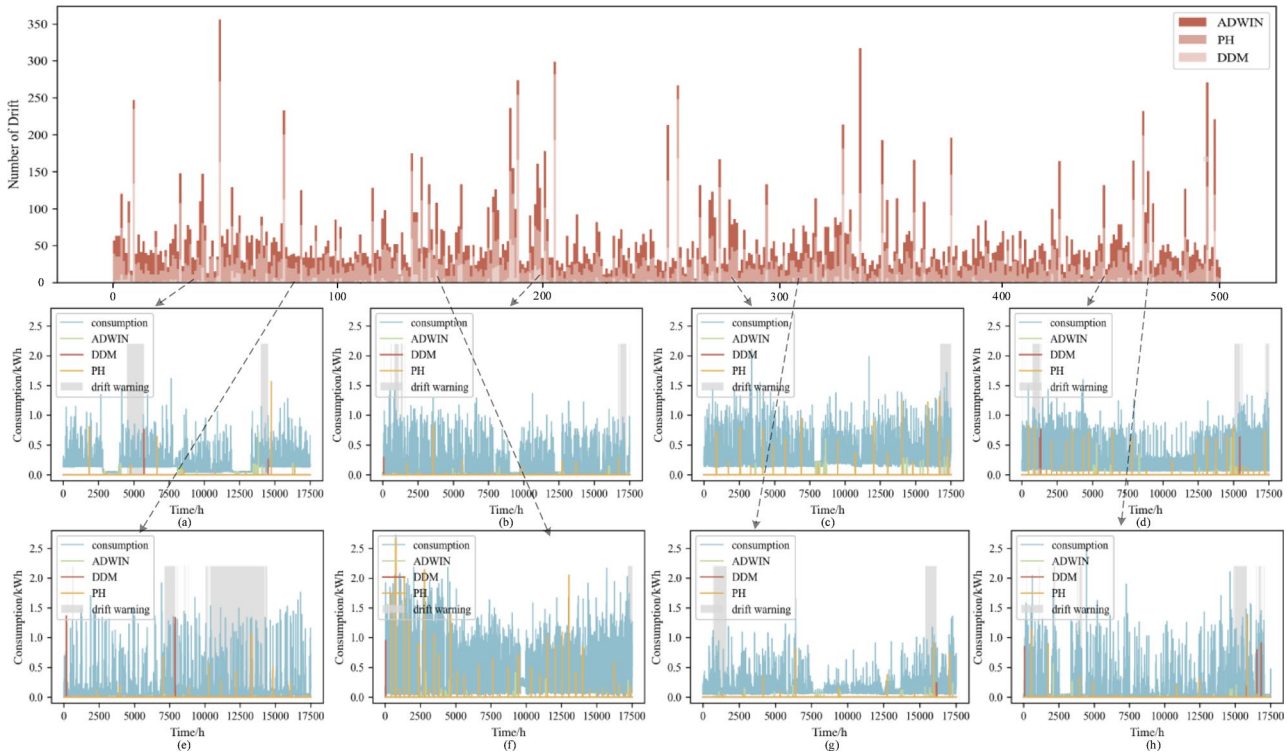


Fig. 3. Detection results produced by three concept drift occurrence detection algorithms on electricity consumption data from multi-environment dataset.

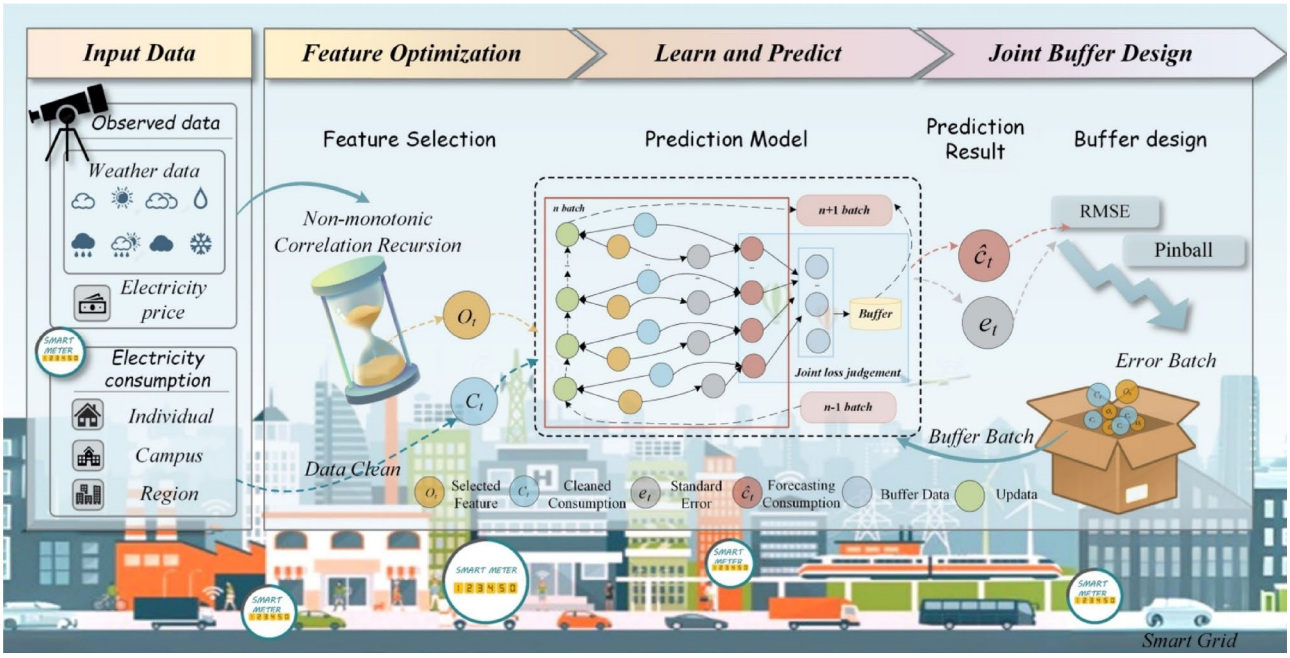


Fig. 4. Framework diagram of the AECF-UC model.

encodings. The related data selected for forecasting include observed variables temperature, pressure, the dew point, humidity, wind bearing, and electricity price.

Data losses and anomalies frequently occur during the processes of data collection and communication, potentially leading to a decline in forecasting performance. To address this issue, anomalies are detected based on Z scores, which measure the deviation of data points from the mean. Specifically, we assume that electricity

consumption data follow a Gaussian distribution, with anomalies typically situated at the extremities of the distribution, far from the center. The deviation of the data points is quantified by the following normalization formula.

$$Z_i = \frac{(c_{t+i} - \mu_{\text{global}})}{\sigma_{\text{global}}} \tag{2}$$

where c_{t+i} represents the electricity consumption data and μ_{global} and σ_{global} are the global mean and standard deviation, respectively. These parameters are adaptively updated according to the following equations:

$$\mu_{\text{global}}{}' = \frac{1}{n}\left[(n-1)\,\mu_{\text{global}} + \mu_n\right] \tag{3}$$

$$\sigma_{\text{global}}{}' = \frac{1}{n}\left[(n-1)\,\sigma_{\text{global}} + \sigma_n\right] \tag{4}$$

where μ_n and σ_n are the mean and standard deviation of the n_{th} batch, respectively. A distance threshold Z_{thr} is established to identify outliers, and data points exceeding this threshold are considered anomalies. The distance threshold Z_{thr} are determined using the probabilistic quantile of standard normal distribution:

$$Z_{\text{thr}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{5}$$

where Φ^{-1} is the inverse cumulative distribution function (CDF) of N (0,1), and α denotes the significance level (empirically set to 0.05 for 95% confidence interval). This theoretically ensures that only 5% of normal data points are misclassified as anomalies. The threshold is dynamically adjusted when global statistics μ_{global} and σ_{global} are updated via Eqs. (3)–(4), enhancing robustness against distribution drift. When a new batch of data is input, missing values and outliers are detected. Then, the online prediction model generates forecasts based on positional encoding information, and these prediction results are used to fill in or replace the missing or anomalous values.

Feature adaptation

The presence of redundant irrelevant information or the omission of important information can lead to a decrease in the accuracy of the electricity consumption prediction model. A feature adaptation module is used to choose appropriate feature subsets, thereby mitigating the differences among them.

A nonmonotonic correlation recursion elimination (NCE) method is proposed to select features. The correlation coefficient between the electricity consumption data Y and the observed features X is captured by measuring the change between the orderings of the random variables[41]. For a dataset of sample size n, the expected value of all possible ranking differences is

$$E\left(\sum\nolimits_{i=1}^{n-1} |r_{i+1} - r_i|\right) = \frac{n\left(n^2-1\right)}{2\,(n+1)} \tag{6}$$

where r_i is the rank of Y_i, and $|r_{i+1}\text{-}r_i|$ denotes the amount of change between the rankings of adjacent observations. Then, Normalization is achieved by dividing the sum of rank differences by the sum of the largest possible rank differences d:

$$d = \frac{\left(n^2-1\right)}{2} \tag{7}$$

and to ensure the correct normalization, the coefficient 3 is added. Finally, the correlation coefficient formula is finally obtained as

$$\xi_n\,(X,Y) = 1 - \frac{3\sum_{i=1}^{n-1} |r_{i+1} - r_i|}{n^2-1} \tag{8}$$

where ξ_n indicates the correlation coefficient. Based on the results of ξ_n in the training set, feature data exceeding the feature threshold D are retained in the initial feature subset. Subsequently, the most appropriate feature subset is obtained by the feature recursive elimination (FRE) method. FRE selects the optimal subset of features via iterative elimination and shifting, effectively screening out a k-dimensional feature subset after repeated iterations.

Equations (9)–(10) are used to help the model better handle non-linear relationships in the data or to reduce noise.

$$\alpha = \begin{cases} 1 & |o_t - \overline{o}| > W \\ 0 & |o_t - \overline{o}| < W \end{cases} \tag{9}$$

$$u_r(\cdot) = [1, \alpha_1, \cdots, \alpha_k] \tag{10}$$

where o_t is the value of the feature at t, and $\bar{o}$ is the average value of the corresponding feature. W is the threshold value of the corresponding feature, and W is set with distinct values for different features: 12 (temperature), 88 (pressure), 18 (dew point), 0.6 (humidity) and 4 (wind bearing). The final observed feature is determined through this process, enabling the use of a predictive model for application in universal environments.

Learning and prediction

The state-space model is used to establish the relationship between the actual electricity consumption values and the observed feature sequence for prediction. The crux of implementing forecasting within this model hinges on the derivation of two probability matrices: $p(y_t|y_{t-1})$ and $p(y_t|o_t)$. y_t is actual electricity consumption at t time, y_{t-1} is electricity consumption at the previous moment (i.e. at t-1 time), and o_t denote the observed features at t time (such as weather, holidays, electricity price, and other external features that affect electricity consumption). The matrix $p(y_t|y_{t-1})$ reflects the probability distribution of the electricity consumption y_t given the previous electricity consumption y_{t-1}, it describes the conditional probability matrix for state transitions. $p(y_t|o_t)$ reflects the probability distribution of the electricity consumption y_t given the observed features o_t, it describes the conditional probability matrix for output relationships. When $m = m(t)$, the relationships are expressed as follows:

$$p(y_t|y_{t-1}) = N(y_t; u_c^T \eta_{c,m}, \sigma_{c,m}) \tag{11}$$

$$p(o_t|y_t) = N(y_t; u_r^T \eta_{r,m}, \sigma_{r,m}) \tag{12}$$

where the feature vectors $u_c^{\mathrm{T}} = [1, y_{t-1}]^{\mathrm{T}}$ and $\boldsymbol{u_r} = u_r(\boldsymbol{o_r}) \in R^R$, together form the mean with parameters $\eta_{c,m} \in R^2 \eta_{r,m} \in R^R$ $\sigma_{c,m} \in R$, and $\sigma_{r,m} \in R$, which are the standard deviations. We organize them into model initialization parameters Θ.

$$\Theta = \left\{ u_r^T \eta_{r,m}, \sigma_{r,m}, u_c^T \eta_{c,m}, \sigma_{c,m} \right\} \tag{13}$$

The model executes recursive calculations to determine the standard deviation and mean and obtains the probability matrix of the transformation relationships using a Gaussian distribution to achieve online prediction. Figure 5 shows the online learning process of the AECF-UC model. Specifically, after performing data cleansing and feature selection, the input data are acquired, including the feature data $X_t[o_t, o_{t+1}, \ldots, o_{t+L}]$, whose shape is $(k \times L)$, and the electricity consumption data $Y_t[y_t, y_{t+1}, \ldots, y_{t+L}]$. The forecasted electricity consumption $\hat{y}_{t+i}$ and standard error $\hat{e}_{t+i}$ at time $t+i$ can be obtained by

$$\hat{y}_{t+i} = \frac{\hat{u}_c^T \eta_{c,m} \sigma_{r,m}^2 + \hat{u}_r^T \eta_{r,m} \left(\sigma_{c,m}^2 + \left(v^T \eta_{c,m} \right)^2 \hat{e}_{t+i-1}^2 \right)}{\sigma_{r,m}^2 + \sigma_{c,m}^2 + \left(v^T \eta_{c,m} \right)^2 \hat{e}_{t+i-1}^2} \tag{14}$$

$$\hat{e}_{t+i} = \sqrt{\frac{\sigma_{r,m}^2 \left(\sigma_{c,m}^2 + \left(v^T \eta_{c,m} \right)^2 \hat{e}_{t+i-1}^2 \right)}{\sigma_{r,m}^2 + \sigma_{c,m}^2 + \left(v^T \eta_{c,m} \right)^2 \hat{e}_{t+i-1}^2}} \tag{15}$$

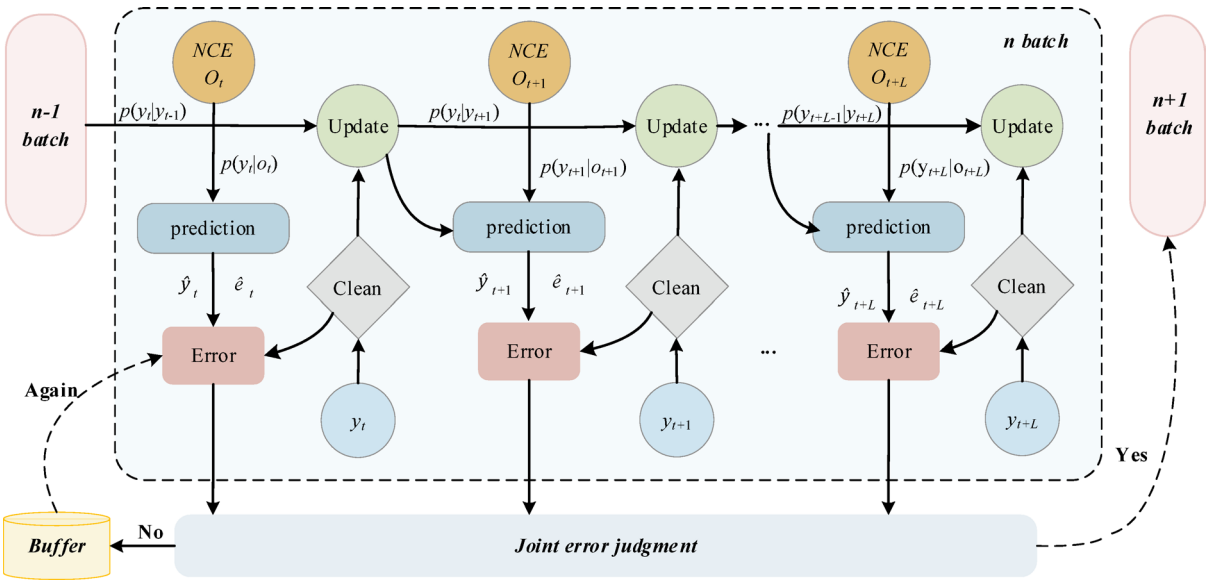


Fig. 5. AECF-UC model for online learning processes.

Input X_t observed feature data, Y_t electricity consumption data, Θ_0 initial model parameters, Γ_0 initial state variables

output $\hat{y}_t$ electricity consumption forecast, $\hat{e}_t$ standard error

Initialization:

$\Gamma \leftarrow \Gamma_0$, $\Theta \leftarrow \Theta_0$

$\hat{y} = 0$, $\hat{e} = 0$

while data are available **do**

$m = m(t)$

$D\ (y_m, o_m) \leftarrow$ *Data Windows* (X_t, Y_t)

$B_f\ (y_m\ o_m) \leftarrow$ buffer data

$B\ (y_m, o_m) \leftarrow B_f \cup D$

for $i = 1, 2, 3, \ldots, L$ **do**

$u_r \leftarrow ur(o_i)$

$u_c \leftarrow [1, y_{i-1}]$

$y_i \leftarrow$ *Date Clean*(c_i)

$\hat{y}_i, \hat{e}_i \leftarrow$ *forecast* $(u_c, u_r, \Theta, \Gamma)$

$\Theta, \Gamma \leftarrow$ *update* $(u_c, u_r, \Theta, \Gamma, y_i)$

See Equations (13-14) and (15-18) for the recursions.

$l_p, l_s \leftarrow$ *error calculation* $(\hat{y}_t, \hat{e}_t, y_t)$

loss $\leftarrow$ *Batch error* (l_p, l_s) Joint error calculation

if *loss > Buffer threshold* **then**

Store the current batch of data into the buffer

Algorithm 1. Online learning step for AECFUC.

where $v = [0, 1]^{\mathrm{T}}$, $\hat{y}_t = 0$, $\hat{e}_t = 0$, and $i = 1, 2, \ldots, L$. Furthermore, the latest prediction model is updated by combining the previous parameter and the actual electricity consumption y_t.

$$\eta_i = \eta_{i-1} + \frac{P_{i-1}u_{ti}}{\lambda + u_{ti}^T P_{i-1} u_{ti}} \left(c_{ti} - u_{ti}^T \eta_{i-1}\right) \tag{16}$$

$$\sigma_i = \sqrt{\sigma_{i-1}^2 - \frac{1}{\gamma_i}\left(\sigma_{i-1}^2 - \frac{\lambda (c_{ti} - u_{ti}^T \eta_{i-1})^2}{\lambda + u_{ti}^T P_{i-1} u_{ti}}\right)} \tag{17}$$

$$P_i = \frac{1}{\lambda}\left(P_{i-1} - \frac{P_{i-1}u_{ti}u_{ti}^T P_{i-1}}{\lambda + u_{ti}^T P_{i-1} u_{ti}}\right) \tag{18}$$

$$\gamma_i = 1 + \lambda\gamma_{i-1} \tag{19}$$

The initial value η_0 ($i=1$) is set as a zero vector with a length of L; P_0 ($i=1$) denotes the $L\times L$ identity matrix; $g_0=0$; and $\sigma_0=0$. Γ, consisting of four state variables $P_{c,m}$, $P_{r,m}$, $\gamma_{c,m}$, and $\gamma_{r,m}$, can be written as follows:

$$\Gamma = \{P_{c,m}, P_{r,m}, \gamma_{c,m}, \gamma_{r,m}\} \tag{20}$$

Then, forecasted electricity consumption is used as important data to quantify the batch loss in the buffer module.

Joint buffer design

The buffer module provides a dynamic updating mechanism to enable online prediction models to learn incrementally to adapt to changes in the distribution of observed data and to maintain data representations at multiple time scales. The core mechanism is to mitigate the problems of "catastrophic forgetting" and "overfitting to the new distribution" of the model due to changes in the data distribution by dynamically balancing the learning of historical knowledge and new data. The playback buffer stores historical data and mixes it proportionally with new data during training, forcing the model to optimize the loss of both current and historical distributions. The total loss function is calculated as follows:

$$L_{total} = \alpha L_{current}(X_{new}, y_{new}) + \beta L_{historical}(X_{buffer}, y_{buffer}) \tag{21}$$

where α and β control the weights of the old and new data, preventing the model from forgetting the old patterns due to the dominance of the new data. To ensure the accuracy of the model's probabilistic prediction, a joint loss function is proposed that combines the features of point forecasting and probabilistic forecasting. This function serves as the foundational mechanism of the buffer module.

$$\text{loss} = {l_s}^k + \frac{1}{L}\sum_{i=t}^{t+L} l_p\,(i)^{1-k} \tag{22}$$

$$l_s = \sqrt{(\hat{c}_t - c_t)^2 + (\hat{c}_{t+1} - c_{t+1})^2 + \cdots + (\hat{c}_{t+L} - c_{t+L})^2} \tag{23}$$

$$l_p\,(i) = \begin{cases} q\left(c_i - \hat{c}_i^{(q)}\right) & c_i \geqslant \hat{c}_i^{(q)} \\ 1 - q\left(\hat{c}_i^{(q)} - c_i\right) & c_i < \hat{c}_i^{(q)} \end{cases} \tag{24}$$

where $\hat{c}_i$ is the predicted value, c_i is the actual value, q is the quantile, l_s is the point forecasting error, and l_p is the probability forecasting error. The replay buffer retains historical data and introduces the historical loss gradient ($\nabla L_{\text{Historical}}$) during each parameter update, effectively imposing a soft constraint on the direction of parameter updates to prevent them from deviating entirely from historically optimal regions.

Then, the least-squares method[42] is employed to determine the optimal buffer threshold and size. When the loss value reaches this threshold, the data are worthy of further investigation. The batch of data is deposited into the buffer to prepare for the dynamic adjustment of the learnable weights. As the buffer reaches its maximum capacity, the oldest batch of data is replaced by the most recent batch, thereby maintaining a current and relevant data buffer for model training.

To illustrate the whole online learning process more intuitively and clearly, the specific pseudocode is given in Algorithm 1.

Experiments and results

Experimental environment and parameter settings

The configurations employed for the experiments are as follows: CPU/Intel(R) Core (TM) i7-10700, Memory/24 GB, Graphics Card/NVIDIA GeForce GTX 3080, and Programming Language/Python 3.10.

The values of the parameters, especially the consumption feature forgetting factor and feature thresholds are very important for the results. We conduct more than 1,000 sets of exhaustive traversal experiments, with the corresponding accuracy results represented by varying colors; darker hues denote smaller errors. The results are optimal when the values of λ_d, λ_r, and D are 0.2, 0.7, and 0.14, respectively. The optimization results obtained for these three parameters are shown in Fig. 6 and the hyperparameter settings for all models are demonstrated in Table 3.

Evaluation indicators

The root means square error (RMSE) is used to quantify the overall prediction performance, which can reflect the accuracy of the forecasting interval and the corresponding point forecasting ability of the model. These evaluation indicators can be mathematically expressed by

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(c_t - \hat{c}_t)^2} \tag{23}$$

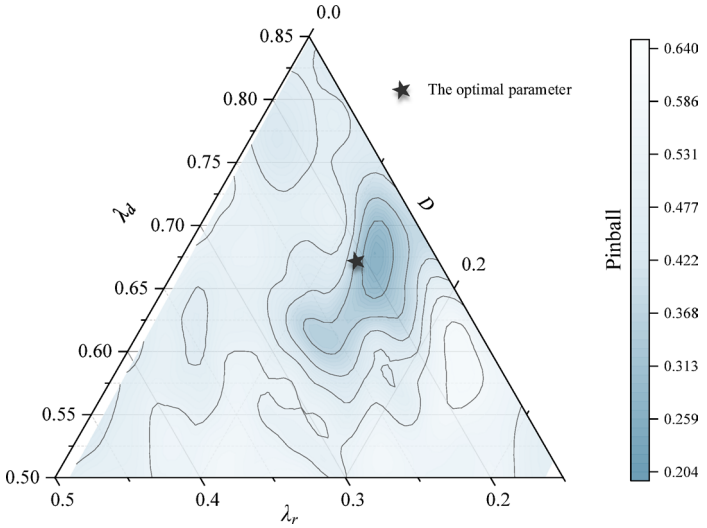


Fig. 6. Optimization results of parameter λd, λr, and D.

Model	Parameters
OARNN	Hidden layer sizes: 64, Number of Layers: 2
GNN	Number of Layers: 4; Hidden Units 64
QLSTM	LSTM unit: 16; FC unit: 16; layer: 3
FPSeq2Q	N:100, Latnet size: 64, Head size: 4, Num layer:4
BNN	FC-1 unit: 512 Layer:1; FC-2 unit: 256 layers: 4
BNNDeT	Encoder layer: 4; Decoder layer: 4; Decoder multi-head: 8; model dimension: 24
BSDeAuto	Encoder layer: 4; Decoder layer: 4; Decoder multi-head: 8; model dimension: 24
APLF	Forgetting factors 1: 0.2; Forgetting factors 2: 0.7

Table 3. Parameter settings for all models.

where N is the number of predicted data points and c_t and $\hat{c}_t$ are the actual and predicted electricity consumption, respectively. The pinball loss is used as an indicator to evaluate the probabilistic predictions. It can quantify the uncertainty of the predictions.

$$pinball\_loss(q, c_t) = \begin{cases} q\left(c_t - \hat{c}_t^{(q)}\right) & c_t \geqslant \hat{c}_t^{(q)} \\ 1 - q\left(\hat{c}_t^{(q)} - c_t\right) & c_t < \hat{c}_t^{(q)} \end{cases} \tag{24}$$

where q is the quantile.

Comparison experiment

In this section, the proposed model is compared with existing load forecasting models, namely, the online adaptive recurrent neural network (OARNN)[14], GNN[15], TFT[16], QLSTM[19], MQR[20], DeepAR[21], fully parameterized sequence-to-quantile regression (Fpseq2q)[21], BNNDeT[22] and APLF[29]. The experimental results are recorded in Table 4 in groups, and the individual probabilistic forecasting results of these models are shown in Fig. 7.

Table 4 presents the overall prediction errors induced by the proposed model and the existing load forecasting models on a multi-environment dataset, as assessed by the RMSE, pinball loss and CRPS. The proposed model exhibits superior overall performance in universal environments, with average RMSE, pinball loss and CRPS values of 0.3041, 0.0567 and 0.1683, respectively. Specifically, In the Individual dataset, the models (the proposed model, GNN, and BNNDeT) that take heterogeneity into account have better performance. Compared to BNNDeT, the proposed model has better RMSE and CRPS, with an average RMSE and CRPS of 0.1364 and 0.1073, respectively. This result indicates that the proposed model can better adapt to the variability and volatility of individual data than the baselines and outperforms the other adaptive models, such as the OARNN, MQR, BSAuto and APLF. Due to the proposed method considers the volatility and even concept drift of data affected by the environment. The proposed model is more stable than other models as it fully considers the differences among individual users, demonstrating its effectiveness in a generalized environment.

Figure 7 demonstrates the load forecasting curves of different methods, with the probabilistic forecasting curves explicitly displaying 60% and 90% confidence intervals. Comparing the OARNN with other prediction

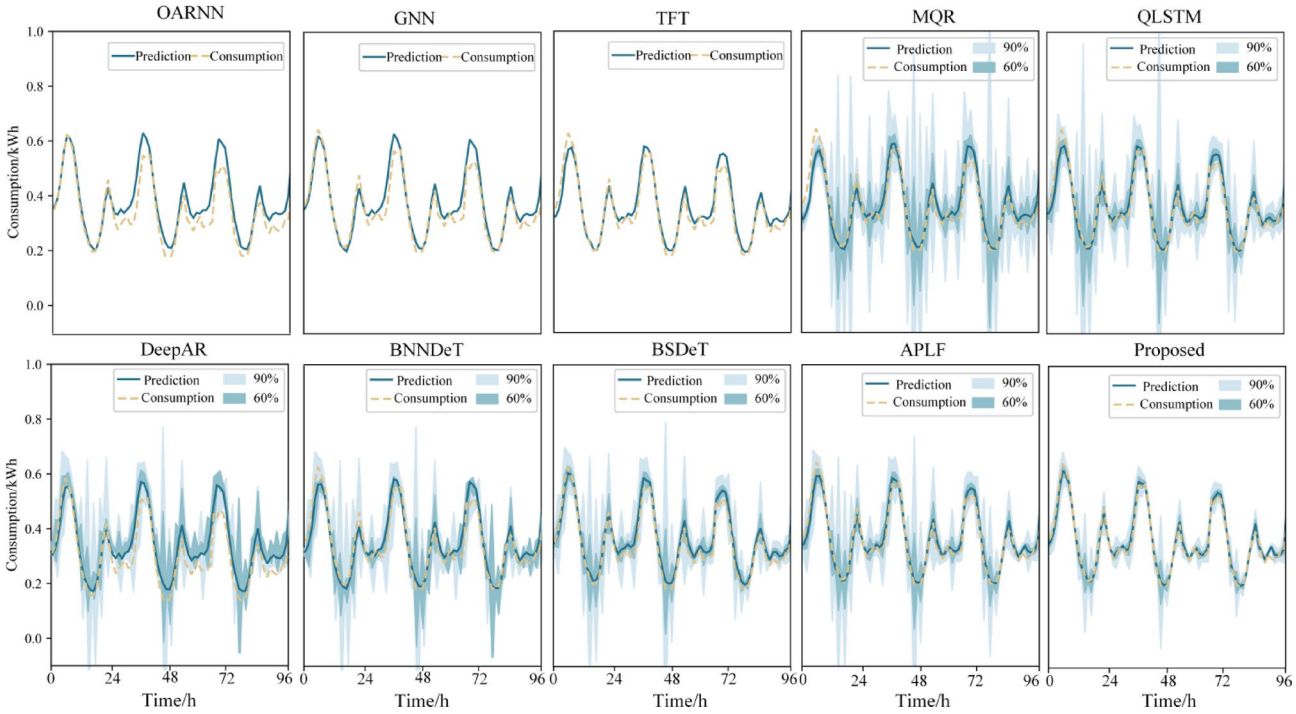
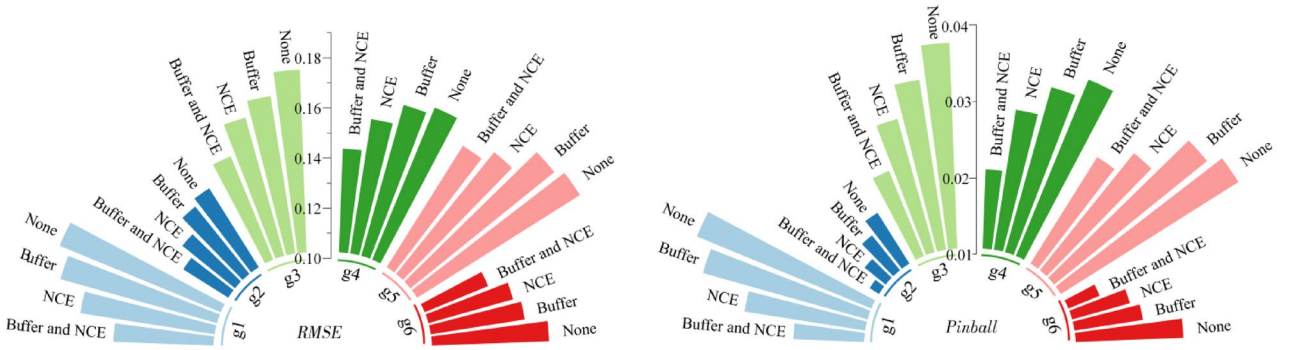


Fig. 7. The result of an individual electricity consumption probabilistic forecasting, 90% and 60% represent the confidence interval.



(a) The effect of Buffer and NCE modules on prediction result: RMSE

(b) The effect of Buffer and NCE modules on prediction result: Pinball

Fig. 8. The effects of different modules on the predicted results. g1: Individual Std, g2: Individual Tou, g3: Campus Downtown, g4: Campus Polytechnic, g5: Region Dayton, g6: Dayton New England.

models, it is observed that the results of probabilistic forecasting, which include prediction intervals, are more intuitive than those of point forecasting. Some of the probabilistic predictions fall outside the 60% confidence interval, but most of them fall within the 90% confidence interval. The proposed model ensures that most of the prediction points fall within the 90% confidence interval, and its prediction range is more accurate, which indicates that the proposed model has good probabilistic forecasting performance.

Ablation experiments

To verify the effectiveness of the feature-adaptive module and the buffer module, ablation experiments are conducted, as shown in Fig. 8. The predictive performance is evaluated using the RMSE and pinball metrics. The experimental group employing the feature-adaptive module, and the buffer module exhibits the best forecasting performance. In contrast, the removal of these two modules results in the largest prediction error. These modules are crucial for addressing the problem of concept drift and user heterogeneity and are effective at enhancing forecasting performance. The contribution of the buffer module is greater due to the higher forgetting factor of the observed feature data compared to that of the historical electricity consumption data.

Index / Selection	RMSE						Average Error
C_6^1	0.1611 (A)	0.1609 (B)	0.1592 (C)	0.1597 (D)	0.1604 (E)	0.1600 (F)	**0.1602**
C_6^2	0.1626 (AB)	0.1619 (AC)	0.1623 (AD)	**0.1599 (AE)**	0.1626 (BC)	-	**0.1621**
	0.1617 (BE)	0.1627 (CD)	0.1618 (CE)	0.1635 (DE)	0.1634 (BD)	-	
	0.1621 (AF)	0.1619 (BF)	0.1622 (CF)	0.1615 (DF)	0.1614 (EF)	-	
C_6^3	0.1652 (ABC)	0.1639 (ABD)	0.1635 (ABE)	0.1646 (ACD)	0.1642 (ACE)	-	**0.1646**
	0.1650 (BCD)	0.1627 (BCE)	0.1647 (BDE)	0.1647 (CDE)	0.1651(ADE)	-	
	0.1636 (ABF)	0.1639 (ACF)	0.1643 (ADF)	0.1649 (AEF)	0.1646 (BCF)	-	
	0.1647 (BEF)	0.1645 (CDF)	0.1648 (CEF)	**0.1602 (DEF)**	0.1639 (BDF)	-	
C_6^4	0.1656 (ABCD)	0.1657 (ABCE)	0.1641 (ABDE)	0.1661 (ACDE)	0.1653 (BCDE)	-	**0.1654**
	0.1652 (ABCF)	0.1663 (ABDF)	0.1655 (ABEF)	0.1656 (ACDF)	0.1652 (ACEF)	-	
	0.1650 (BCDF)	0.1660 (BCEF)	0.1647 (BDEF)	0.1657 (CDEF)	0.1651 (ADEF)	-	
C_6^5	0.1669 (ABCDE)	0.1653 (BCDEF)	0.1671 (ACDEF)	0.1663 (ABDEF)	0.1669 (ABCEF)	0.1674 (ABCDF)	**0.1666**
C_6^6	0.1684 (ABCDEF)	-	-	-	-	-	**0.1684**
NCE		-	-	-	-	-	**0.1577**

Note: A→ Temperature; B→ Pressure; C→ Dewpoint; D→ Humidity; E→ Wind bearing; F→ Electricity price

Table 5. Relationship between the forecasting result and the different feature selections.

To analyze the correlation results and the associations of the feature selection process, the related results are tested. One hundred individuals in the DTou group are used to exploring the relationships between the prediction results and the different selected features, and the corresponding results are shown in Table 5.

According to Table 5, the temperature, pressure, dewpoint, humidity, wind bearing, and electricity price are selected for the analysis of the prediction results. Overall, the more information contained in the features, the worse the prediction ability of the model is. However, some combinations of features are better than a single feature, such as "AE" in C_6^2 and "DEF" in C_6^3. Therefore, compared with a fixed feature input, the proposed adaptive feature selection method has the best performance, with an average error of 0.1577. This indicates the effectiveness of feature selection, and the impact of user heterogeneity is mitigated.

In Fig. 9a, we compare the predictive performances of three methods using different features: temperature as the sole fixed feature, six sets of data as fixed features, and the NCE method to adaptively selecting features. To clearly observe the performance differences among these methods, we normalize their prediction errors. It is observed that the NCE method performs better in the prediction task than do the fixed features. This is attributed to the use of inappropriate observed features for prediction, which may lead to poor estimates. The findings underline the importance of selecting correct features for enhancing the accuracy of model prediction and validate the feasibility of employing NCE in the feature selection. We further explore different feature selection methods for electricity consumption forecast, including the linear Pearson correlation method and the nonlinear Spearman correlation method, to assess the advantages of nonmonotonic correlation. Additionally, we compare correlation analysis and recursive elimination to ascertain the effectiveness of the NCE method in the feature selection process. The results are shown in Fig. 9b, and the prediction error is normalized to clearly observe the performance differences. We observe that the predictive performance attained by combining correlation and recursion is superior to that of correlation or recursion alone. The NCE method outperforms the feature subsets selected by the Pearson and Spearman correlation methods with recursive elimination. In summary, NCE is most suitable for feature selection in electricity consumption forecast tasks.

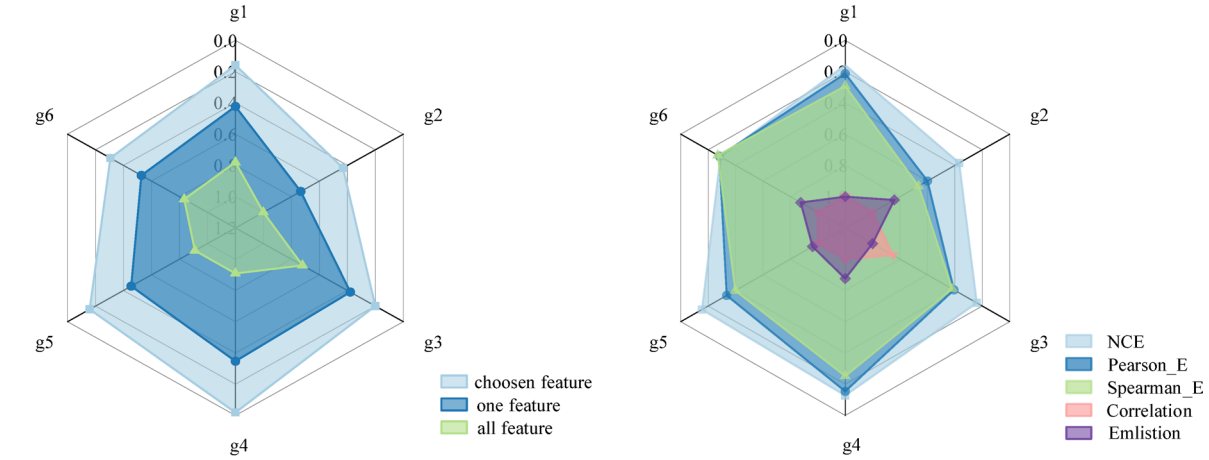
To evaluate the effectiveness of the joint loss buffer design, we conduct a comparison experiment with the single-loss buffer method. Experiments are carried out over various forecasting steps, including half a day (a step), one day (two steps), two days (four steps) and one week (fourteen step), and the experimental results are shown in Fig. 10. The prediction accuracy of the buffer design using the joint loss method improves by approximately 6% over that of the single-loss method. The joint loss function demonstrates superior performance across different datasets and prediction durations. The adoption of the joint loss method in electricity consumption forecast enables the model to better adapt to data changes, effectively mitigates the concept drift problem, and provides superiority in model adaptability. The prediction accuracy decreases as the prediction dimensionality increases, but the RMSEs remain within 0.18 (g1-g5) and 1.5 (g6). These results affirm the effectiveness of the model in short-term electricity consumption forecast scenarios.

Conclusion

In this paper, we propose an adaptive electricity consumption forecast method (AECF-UC) to address two key challenges: user heterogeneity and concept drift. Specifically, the nonmonotonic correlation recurrence elimination method can find suitable feature data in multiple environments, enabling the application of the model in universal environments. We introduce a buffer module based on the joint loss function. It can ensure that the data used for training and updating the model accurately reflects the current data distribution, thereby aiding the online learning model in adapting to concept drift. We conduct various experiments on multi-environment datasets to assess the accuracy of AECF-UC and compare it with other methods. These datasets represent challenging scenarios with varying scales, domains, and consumption patterns. The experimental results demonstrate that the proposed model attains the best overall performance in universal environments, with average RMSE, pinball loss and CRPS of 0.3041, 0.0567 and 0.1683 respectively.

	Individual						Campus						Region					
	Std			Tou			Downtown			Polytechnic			Dayton			New England		
Methods	RMSE	Pinball	CRPS	RMSE	Pinball	CRPS	RMSE	Pinball	CRPS	RMSE	Pinball	CRPS	RMSE	Pinball	CRPS	RMSE	Pinball	CRPS
OARNN	0.2296	–	–	0.1822	–	–	0.2193	–	–	0.2278	–		0.2872	–		1.1320	–	–
GNN	0.1489			0.1337			0.1519			0.1498			0.1648			1.1302		
FTF	0.1519	–	–	0.1297	–	–	0.1479	–	–	**0.1378**	–		0.1604	–		1.1298	–	–
QLSTM	0.2311	0.0277	0.1487	0.1942	0.0251	0.1539	0.2127	0.0273	0.1578	0.2447	0.0383	0.1507	0.1967	0.0256	0.1619	1.2991	0.2977	1.1924
MQR	0.3133	0.0294	0.1869	0.2917	0.0217	0.1927	0.2688	0.0242	0.1692	0.2044	0.0237	0.1415	0.2335	0.0313	0.1628	1.2336	0.3018	0.9310
DeepAR	**0.1513**	0.0262	0.1167	0.1293	0.0190	**0.1027**	0.1436	0.0265	0.1224	0.1425	0.0236	0.1308	0.1639	0.0250	**0.0136**	1.1343	0.2409	**0.7939**
FPSeq2Q	0.2444	0.0209	0.1541	0.1383	0.0196	0.1231	0.1922	0.0294	0.1519	0.1731	0.0294	0.1571	0.2089	0.0292	0.1692	1.1319	0.2487	1.2981
BNNDeT	0.1922	**0.0195**	**0.1129**	**0.1219**	**0.0189**	0.1046	0.1419	0.0215	0.1551	0.1587	**0.0191**	0.1618	0.1898	**0.0217**	0.1498	**1.1288**	**0.2344**	1.2150
BSAuto	**0.1529**	**0.0183**	0.1297	0.1303	0.0194	0.1309	**0.1402**	0.0221	0.1232	0.1440	0.0199	**0.1208**	**0.1618**	0.0251	0.1409	1.1305	0.2401	1.1072
APLF	0.1682	0.0341	0.1319	0.1417	0.0233	0.1324	0.1489	**0.0219**	**0.1217**	0.1447	0.0228	0.1289	0.1651	0.0379	0.1371	1.1521	0.2525	0.8310
proposed	**0.1476**	0.0201	**0.1108**	**0.1253**	**0.0181**	**0.1039**	**0.1351**	**0.0207**	**0.1208**	**0.1319**	**0.0197**	**0.1215**	**0.1574**	**0.0248**	**0.1331**	**1.1274**	**0.2371**	**0.8198**

Table 4. RMSEs and pinball loss induced by AECF-UC and some existing methods for six groups of forecasting data.



(a) The effect of different observed feature on predicting performance

(b) The effect of feature selection methods on predicting performance

Fig. 9. The effects of different observed features on the predicted results.

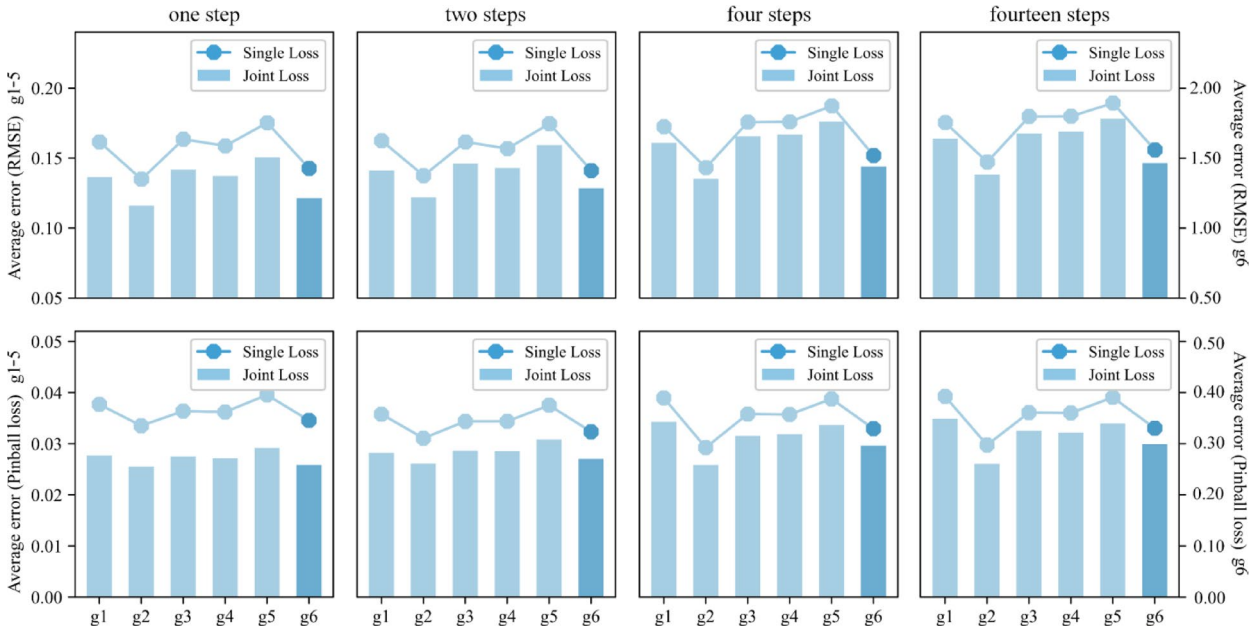


Fig. 10. The prediction errors are induced under various forecasting steps. g1: Individual Std, g2: Individual Tou, g3: Campus Downtown, g4: Campus Polytechnic, g5: Region Dayton, g6: Region New England. (one step represents 12 h ahead)

However, recursive feature selection with non-monotonic correlation elimination improves feature quality, but it increases the computational complexity of the model and lacks finer-grained modeling of user behavior. Efficient and comprehensive feature selection methods can be further explored in the future to cope with user heterogeneity.

Data availability

The datasets generated and analysed during the current study are available in the https://data.london.gov.uk/blog/electricity-consumption-in-a-sample-of-london-households/, https://cm.asu.edu, and https://www.iso-ne.com.

Received: 18 March 2025; Accepted: 2 July 2025
Published online: 08 July 2025

References

1. Rho, S., Chae, M. & Won, D. Forecast-based optimal operation of EV charging station with PV considering charging demand and distributed system. In *2024 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 1–5 (2024).
2. Ji, X., Dong, Z., Zhou, G., Lai, C. S. & Qi, D. MLG-NCS: Multimodal local–global neuromorphic computing system for affective video content analysis. In *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2024).
3. Xie, Y., Zeng, P. & Chen, J. Hybrid attention-based improved temporal convolutional BiGRU approach for short-term load forecasting. *J. Phys. Conf. Ser. IOP Publ.* **2703**, 1 (2024).
4. Ge, Q. et al. Industrial power load forecasting method based on reinforcement learning and PSO-LSSVM. *IEEE Trans. Cybern.*. **52** (2), 1112–1124 (2022).
5. Peplinski, M. et al. A machine learning framework to estimate residential electricity demand based on smart meter electricity, climate, Building characteristics, and socioeconomic datasets. *Appl. Energy*. **357**, 122413 (2024).
6. Li, Q., Xu, Y., Chew, B. S. H., Ding, H. & Zhao, G. An integrated missing-data tolerant model for probabilistic PV power generation forecasting. *IEEE Trans. Power Syst.* **37** (6), 447–4459 (2022).
7. Ji, X. et al. Time-frequency hybrid neuromorphic computing architecture development for battery state-of-health Estimation. *IEEE Internet Things J.* (2024).
8. Yang, W. et al. A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. *Appl. Energy*. **307**, 118197 (2022).
9. Liu, T., Chen, S., Li, K., Gan, S. & Harris, C. J. Adaptive multioutput gradient RBF tracker for nonlinear and nonstationary regression. *IEEE Trans. Cybern.* 1–14 (2023).
10. Ge, Y., Zhang, W., Liu, G., Li, Z. & Li, H. Adaptive feature selection for probabilistic Multi-Energy load forecasting. *IEEE Trans. Ind. Appl.* (2023).
11. Wang, Y., Gao, N. & Hug, G. Personalized federated learning for individual consumer load forecasting. *CSEE J. Power Energy Syst.* **9** (1), 326–330 (2023).
12. Jalali, S. M. J. et al. A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting. *IEEE Trans. Ind. Inf.* **17** (12), 8243–8253 (2021).
13. Dong, Z. et al. Periodic segmentation transformer-based internal short circuit detection method for battery packs. *IEEE Trans. Transp. Electrif.* (2024).
14. Lai, C. S. et al. Multi-view neural network ensemble for short and mid-term load forecasting. *IEEE Trans. Power Syst.* **36** (4), 2992–3003 (2020).
15. Lv, Y., Wang, L., Long, D., Hu, Q. & Hu, Z. Multi-area short-term load forecasting based on spatiotemporal graph neural network. *Eng. Appl. Artif. Intell.* **138**, 109398 (2024).
16. Lim, B. et al. Temporal fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **37** (4), 1748–1764 (2021).
17. Zhang, X. et al. Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model. *IEEE Trans. Cybern.* **51** (6), 3157–3170 (2020).
18. Lemos-Vinasco, J., Bacher, P. & Møller, J. K. Probabilistic load forecasting considering Temporal correlation: online models for the prediction of households' electrical load. *Appl. Energy*. **303**, 117594 (2021).
19. Lin, J., Ma, J., Zhu, J. & Cui, Y. Short-term load forecasting based on LSTM networks considering attention mechanism. *Int. J. Electr. Power Energy Syst.* **137**, 107818 (2022).
20. Bracale, A., Caramia, P., De Falco, P. & Hong, T. Multivariate quantile regression for short-term probabilistic load forecasting. *IEEE Trans. Power Syst.* **35** (1), 628–638 (2019).
21. Salinas, D. et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* **36** (3), 1181–1191 (2020).
22. Faustine, A. & Pereira, L. FPSeq2Q: fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates. *IEEE Trans. Smart Grid*. **13** (3), 2440–2451 (2022).
23. Wang, C., Wang, Y., Ding, Z. & Zhang, K. Probabilistic multi-energy load forecasting for integrated energy system based on bayesian transformer network. *IEEE Trans. Smart Grid*. **15** (2), 1495–1508 (2023).
24. Zhou, S., Wang, J., LAI, C. S., Yuan, Y. & Dong, Z. Multi-view adaptive probabilistic load forecasting combing bayesian autoformer network. *J. Electron. Inform. Technol.* **46** (12), 4432–4440 (2024).
25. Wang, Y. et al. Short-term industrial load forecasting based on ensemble hidden Markov model. *IEEE Access.* **8**, 160858–160870 (2020).
26. Luan, W., Yang, F., Zhao, B. & Liu, B. Industrial load disaggregation based on hidden Markov models. *Electr. Power Syst. Res.* **210**, 108086 (2022).
27. Yin, B. et al. Non-intrusive load monitoring algorithm based on household electricity use habits. *Neural Comput. Appl.* **34** (18), 15273–15291 (2022).
28. Silva, P. C. D. L. et al. Forecasting in non-stationary environments with fuzzy time series. *Appl. Soft Comput.* **97**, 106825 (2020).
29. Munkhammar, J., van der Meer, D. & Widén, J. Very short term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model. *Appl. Energy*. **282**, 116180 (2021).
30. Álvarez, V., Mazuelas, S. & Lozano, J. A. Probabilistic load forecasting based on adaptive online learning. *IEEE Trans. Power Syst.* **36** (4), 3668–3680 (2021).
31. Hosseinpour, Z., Arefi, M. M., Mozafari, N., Luo, H. & Yin, S. An ensemble-based fuzzy rough active learning approach for broken rotor bar detection in nonstationary environment. *IEEE Trans. Instrum. Meas.* **71**, 1–8 (2022).
32. Electricity Consumption in a Sample of London Households. https://data.london.gov.uk/blog/electricity-consumption-in-a-sample-of-london-households/.
33. Campus Metabolism. https://cm.asu.edu.
34. PJM connection. https://www.pjm.com/.
35. Iso new england. zonal information. https://www.iso-ne.com.
36. RB, C. STL: A seasonal-trend decomposition procedure based on loess. *J. Off Stat.* **6**, 3–73 (1990).
37. St, L. & Wold, S. Analysis of variance (ANOVA). *Chemometr. Intell. Lab. Syst.* **6**(4), 259–272 (1989).
38. Verma, P., Chodkowska-Miszczuk, J. & Raghubanshi, A. S. Are cities ready for low-carbon inclusive strategies? Household energy management under heterogeneous socioeconomic conditions. *Sustain. Dev.* **32** (5), 4518–4534 (2024).
39. Angelopoulos, A. et.al. Impact of classifiers to drift detection method: a comparison. In *International Conference on Engineering Applications of Neural Networks* (Springer International Publishing, 2021).
40. Ji, Y., Geng, G. & Jiang, Q. Enhancing model adaptability using concept drift detection for short-term load forecast. In *2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, 464–469 (2021).
41. Charusheela, C. & Lalitha, L. A comparative analysis of algorithm for detecting concept-drift. *AIP Conf. Proc.* **2742**, 1 (2024).
42. Chatterjee, S. A new coefficient of correlation. *J. Am. Stat. Assoc.* **116** (536), 2009–2022 (2021).
43. Sáez, P. B. & Rittmann, B. E. Model-parameter estimation using least squares. *Water Res.* **26** (6), 789–796 (1992).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants (62206062

and 62401326), the China Postdoctoral Science Foundation under Grants 2024T170463 and 2024M751676, and the Shuimu Tsinghua Scholar program under Grant 2023SM035.

Author contributions

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.We understand that the Corresponding Author is the sole contact for the Editorial process. She is responsible for communicating with the other authors about progress submissions of revisions and final approval of proofs.All authors as follows: Author 1 (First Author Z.S.): Conceptualization, Data Curation, Methodology, Formal Analysis, Writing-Original Draft, Writing-Review and Editing; Author 2 (N.S.): Methodology, Writing-Review & Editing, Funding Acquisition; Author 3 (H.Y.): Data Curation, Validation; Author 4 (Corresponding D.Z.): Conceptualization, Writing-Review & Editing, Funding Acquisition; Author 5 (L.C.): Investigation, Funding Acquisition.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.