Short-term Load Forecasting of an Integrated Energy System Based on STL-CPLE with Multitask Learning

Suxun Zhu, Hengrui Ma, Laijun Chen, Bo Wang, Hongxia Wang, Xiaozhu Li, and Wenzhong Gao, *Fellow*, *IEEE*

***Abstract*—Multienergy loads in integrated energy systems (IESs) exhibit strong volatility and randomness, and existing multitask sharing methods often encounter negative migration and seesaw problems when addressing complexity and competition among loads. In line with these considerations, a short-term multienergy load joint prediction method based on seasonal-trend decomposition using LOESS (STL) and convolutional progressive layered extraction (CPLE) is proposed, called STL-CPLE. First, STL is applied to model regular and uncertain load information into interpretable trend, seasonal, and residual components. Then, joint modeling is performed for the same type of components of multienergy loads. A one-dimensional convolutional neural network (1DCNN) is constructed to extract deeper feature information. This approach works in concert with the progressive layered extraction sharing method, and convolutional shared and task-specific experts are developed to acquire common and distinctive representations of multienergy loads, respectively. Task-specific parameters are gradually separated through progressive routing. Finally, a subtask network is built to learn temporal dependencies using long short-term memory (LSTM). Simulation validation is performed on the IES dataset at the Tempe campus of Arizona State University, and the experiments show that the STL-CPLE method exhibits higher prediction accuracy than do the other methods.**

***Index Terms*—Integrated energy system, multienergy load forecasting, convolutional progressive layer extraction network, seasonal-trend decomposition.**

I. Introduction

As the issues of energy shortages and environmental degradation become more serious, promoting the energy revolution has become a national strategy [1]. The traditional mode of supplying a single type of energy can no longer accommodate the diversified forms of energy demand and usage. As a crucial physical carrier facilitating the transition from a traditional power system to a new type of power system, an integrated energy system (IES) enables the coupling of various types of energy across different links, such as sources, grids, and loads. It also enhances the penetration rate of renewable energy in the power system [2]. Accurate and efficient multienergy load forecasting is essential for IESs to formulate planning schemes and optimize operation strategies. However, due to the diverse load types in IESs and the complex coupling relationships between energy subsystems, research on multienergy load forecasting methods in IESs is required [3], [4].

Currently, the majority of load forecasting methods primarily focus on the independent forecasting of individual loads and can be broadly categorized into two groups: statistical methods and machine learning methods. Statistical methods, such as regression analysis [5], exponential smoothing [6], and Kalman filtering [7], work well with linear data but are limited when dealing with nonlinear data. While traditional machine

Received: December 29, 2023
Accepted: August 27, 2024
Published Online: November 1, 2024

Suxun Zhu is with the College of Energy and Electrical Engineering, Qinghai University, Xining 810016, China (e-mail: zhuxun456@163.com).

Hengrui Ma (corresponding author) is with the College of Energy and Electrical Engineering, Qinghai University, Xining 810016, China and with the School of Electrical and Automation, Wuhan University, Wuhan 430072, China (e-mail: henry3764@foxmail.com).

Laijun Chen is with the College of Energy and Electrical Engineering, Qinghai University, Xining 810016, China (e-mail: chenlaijun@tsinghua.edu.cn).

Bo Wang is with the School of Electrical and Automation, Wuhan University, Wuhan 430072, China (e-mail: whwdwb@whu.edu.cn).

Hongxia Wang is with the Department of Electrical and Computer Engineering, University of Denver, Denver CO 80208, USA and with the School of Electrical and Automation, Wuhan University, Wuhan 430072, China (e-mail: Hongxia.Wang@du.edu).

Xiaozhu Li is with Engineering Research Center for Renewable Energy Generation and Grid Connection Technology, Ministry of Education, Xinjiang University, Urumchi 830046, China (e-mail: lixiaozhu@xju.edu.cn).

Wenzhong Gao is with the Department of Electrical and Computer Engineering, University of Denver, Denver CO 80208, USA (e-mail: David.Gao@du.edu).

DOI: 10.23919/PCMP.2023.000101

learning methods such as support vector machines [8], random forests [9], and extreme learning machines [10] have strong nonlinear fitting abilities, they cannot automatically learn the temporal dependence within sequence data. In recent years, cutting-edge machine learning methods centered on deep learning have been increasingly applied to load forecasting due to their strong learning ability and nonlinear fitting capability. Considering the time dependence of load series, reference [11] proposed a residential load forecasting model based on long short-term memory (LSTM). References [12] and [13] combined a convolutional neural network (CNN) and LSTM to better exploit the spatiotemporal information contained in historical loads, meteorological data, and other data. Based on the deep learning model, references [14] and [15] applied a modal decomposition algorithm to decompose the loads and extract implicit regularity information to predict them separately, thereby reducing the difficulty of modeling.

Compared to single-energy load forecasting, the application scenario for multienergy load forecasting is more complex [16]. Access to renewable energy sources and the existence of energy conversion devices not only create coupling relationships among different energies but also make load fluctuations more complex and variable. As a result, accurate load forecasting becomes more challenging. Feature engineering and prediction modeling are important factors that influence the accuracy of IES multienergy load forecasting. In terms of feature engineering, many studies decompose the multienergy loads before predicting them. The decomposition algorithm decomposes the load sequence into relatively smooth components, constructs a prediction model for each component, and combines the component prediction results to generate the final predicted values. Empirical mode decomposition (EMD) is used to decompose the load into intrinsic modal functions (IMFs) with varying amplitude-frequency characteristics [17], [18]. The nonlinear characteristics of the load are effectively weakened. However, the reliability of feature processing can be affected by mode aliasing in EMD. Twice decomposition is a viable method for mitigating the modal aliasing problem [19]. The disadvantage of this method is that it produces more modal components, which significantly increases the computational burden when processing the set of components [20]. In addition, the predicted values of many components are more likely to lead to the accumulation of estimation errors during the superposition reconstruction process, which adversely affects the final results. Some researchers have also attempted to use decomposition algorithms that suppress modal aliasing. Load sequences are processed using variational mode decomposition (VMD), and multiple models are combined for ultrashort-term forecasting [21], [22]. Reference [23] used complementary integrated empirical mode decomposition (CEEMD) supported by the addition of positive and negative white noise to reduce modal aliasing. Such methods primarily focus on decomposing high- and low-frequency components of sequential data and lack interpretability.

The forecasting models can be divided into two categories based on the number of models. The first category is the single prediction of multienergy loads, where multiple types of loads, such as cooling, heating, and electricity, are modeled and predicted separately, with independent models for them. For example, the CNN-LSTM hybrid model is employed for short-term load forecasting in IESs [24], [25]. Reference [26] constructed a correction prediction framework for IES loads based on a temporal convolutional network (TCN). Reference [27] combined dual attention and LSTM within an encoder-decoder architecture to learn the spatiotemporal characteristics of the input data. The above methods are simple extensions of single-load forecasting [28]. It is difficult to use the correlation information between the predicted loads and other loads for assisted prediction.

The second category is the joint forecasting of multienergy loads. These methods are usually combined with multitask learning frameworks, which consider multienergy loads as holistic forecasting targets and generate forecasting results for multiple loads, including electricity, heating, and cooling. References [29] and [30] developed a short-term joint load forecasting method by utilizing the temporal memory capability of LSTM and the shared information learning capability of multitask learning. A CNN-BiGRU multienergy load forecasting model based on the optimization of the attention mechanism was proposed [31]. References [32] and [33] applied a shared model based on transformers for load forecasting. The above study shows that considering the coupling relationships among energy loads can improve the prediction accuracy. Nevertheless, in the traditional shared-bottom structure, all the information must be shared, and the differences in correlations between subtasks cannot be considered [34]. That is, if the coupling relationship of some tasks is weak, the overall performance may be degraded due to the misdirection of other tasks [35].

To this end, reference [34] introduced multiple parallel expert subnetworks in the shared-bottom stage and assigned a gating unit to each task to control the weight of each expert. The established multigate mixture-of-experts (MMoE) forecasting model achieves a reasonable allocation of shared information and improves the overall performance of the model. However, the MMoE still has several limitations. Its underlying expertise is still shared, and there may be a “seesaw” phenomenon [36]: an increase in performance on one task is accompanied by a decrease in performance on another task. Moreover, the absence of differentiation

and interaction among expert subnetworks may compromise the efficacy of optimization.

Considering the load characteristics of IESs and the coupling relationship between different loads and considering the seesaw phenomenon of joint forecasting, a multienergy load short-term forecasting method based on seasonal-trend decomposition and convolutional progressive layered extraction is proposed, named STL-CPLE. The main contributions of this work are as follows:

1) A feature processing method based on STL timing decomposition is proposed. The STL algorithm is utilized to decouple the complex information within the load and break it down into trend, seasonal, and residual components. The interpretability of the load sequence is improved, and the influence of randomness on other components is avoided.

2) A progressive separation extraction multitask learning method based on a prior knowledge network is proposed. The expert network is clearly divided into shared experts and task-specific experts. This division alleviates the harmful parameter interference between different expert modules and enables customized knowledge sharing. At the same time, a progressive information routing mechanism is adopted, where information is absorbed from shallow experts and progressively distributed to specific towers after obtaining deeper levels of shared knowledge. More efficient and flexible knowledge transfer and shared representation learning are achieved.

3) A CPLE model based on progressive layered extraction (PLE) is proposed for the joint prediction of short-term electricity, cooling, and heating loads. The 1DCNN serves as a low-level expert for extracting high-dimensional features, while the LSTM functions as a subtask layer for modeling temporal relationships. The hybrid network is combined with the PLE method to further optimize the performance and efficiency of collaborative forecasting.

4) To validate the effectiveness of our model, we compare it with single-task models (LSTM, CNN-LSTM, TCN, DA-Seq2Seq) and multitask models (CNN-BiGRU-Attention, transformer, LSTNet-Skip, MMoE). In addition, comparisons are made with other decomposition algorithms (EMD, CEEMD, and STL based on additive forms). The MAPE, RMSE, and WMA of different models are calculated to evaluate their performances. The results demonstrate a significant improvement in the forecasting accuracy of the proposed model in cooling, heating, and electricity load forecasting tasks.

The remainder of this paper is structured as follows. Section II analyzes the load characteristics. Section III describes the details of the proposed model. Section IV reports the experimental results. Section V provides the conclusion.

II. Analysis of the Load Characteristics of the IES

A. Coupling Analysis of Multienergy Loads

The IES at the Tempe campus of Arizona State University was used for a study to investigate the coupling between different loads in the IES. The campus encompasses 288 buildings, accommodates 50 246 students, and boasts a total of 73 190 solar photovoltaic panels. Figure 1 illustrates the energy flow among the fundamental components of the IES.

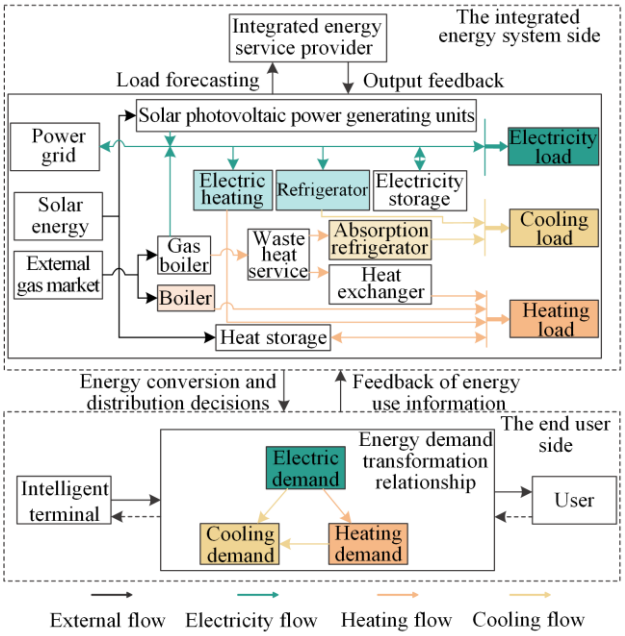


Fig. 1. Interaction structure of the IES and end-user sides.

The power supply for this system is derived from both an external grid and rooftop photovoltaic sources. Equipment such as combined cooling heating and power (CCHP) units, boilers, gas boilers, and chillers serve as energy conversion devices. The system consists of electrical, cooling, and heating subsystems that couple various energy sources by integrating multiple energy conversion and storage devices to fulfill diverse energy requirements. The energy service provider plays the role of coordinator and is responsible for planning the transformation, storage, distribution, and consumption of all types of energy within the IES. The user side employs smart terminals to collect diverse energy demands and transmit them back to the energy service provider. The energy service provider aggregates and analyzes this energy usage information and then makes energy conversion and distribution decisions with the help of accurate multienergy load forecasting.

The changes in electricity, heating, and cooling loads for the IES of the Tempe Campus from 2020 to 2021 are depicted in Fig. 2. The electrical and cooling loads are positively correlated and fluctuate at approximately the same frequency. After dimensionless quantization, it is

evident that the heat demand is negatively correlated with the electricity and cooling demand. This indicates that there is a dependent relationship between the energy demands and that the trend of load changes hides information about the pattern of other loads. The heating demand is significantly lower than the electricity and cooling loads. The reason for this difference is that Tempe has a tropical desert climate with long hours of sunshine and low rainfall, making it one of the cities with the highest average annual temperatures in the United States.

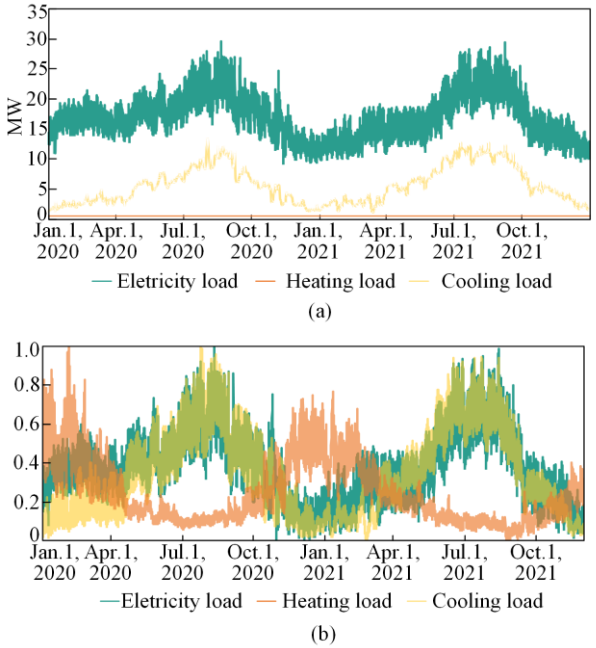


Fig. 2. Load curves for 2020–2021. (a) Original loads. (b) Dimensionless loads.

B. Decomposition of Multienergy Loads

From Fig. 2, the loads exhibit nonstationarity. Different loads have different trends and fluctuations in different seasons, with heating loads experiencing particularly dramatic changes in winter. Additionally, energy conversion between subsystems and diversified user demands increases the stochasticity and volatility of multienergy loads. However, the applicability of the model is limited, and direct modeling of IES loads with complex information is prone to prediction errors. The STL decomposition algorithm can isolate the multiple laws contained in loads [37]. Each decomposition term has a clear physical meaning [38], which helps the prediction model to better understand the inherent laws and characteristics of the sequence.

STL is a decomposition algorithm based on locally weighted regression (LOESS). The basic idea is to decompose the original time series into three components: trend T_t, seasonal/periodicity S_t, and residual R_t. The trend term represents the trend and direction of change, the seasonal term reflects the periodic changes at different time scales; and the residual term is the remaining part of the original series after subtracting the trend and seasonal terms. It represents the random factors in the series. The decomposition of the STL can be expressed as:

$$Y_t = T_t + S_t + R_t \tag{1}$$

where n is the length of the time series, $t = 1, 2, \cdots, n$.

The computational process of STL is divided into inner and outer loops. In the inner loop, the trend and seasonal terms are decomposed via local weighted regression smoothing and other operations. The outer loop then adjusts the robust weights based on the residual results to minimize the influence of outliers on the next inner loop. Let $T_t^{(k)}$ and $S_t^{(k)}$ be the results of the kth iteration, and let m_p be the period window, i.e., the number of samples in a cycle. Then the $(k+1)$ th iteration of the inner loop is as follows:

Step 1: Detrend. $Y_t - T_t^{(k)}$ is calculated for detrending.

Step 2: Smooth the periodic subsequence. The sequence obtained in step 1 is divided, and the divided subsequences are recombined after applying Loess to obtain the temporary periodic sequence $C_t^{(k+1)}$.

Step 3: Low-pass filtering. The low flux $L_t^{(k+1)}$ is obtained by applying a low-pass filter to $C_t^{(k+1)}$. The moving average of the length m_p, m_p, 3 is performed in order, and then perform a Loess to get the low-pass $L_t^{(k+1)}$.

Step 4: Calculate the seasonal component. The seasonal component of the $(k+1)$ th iteration is $S_t^{(k+1)} = C_t^{(k+1)} - L_t^{(k+1)}$. Among them, the subtraction of $L_t^{(k+1)}$ by $L_t^{(k+1)}$ is to prevent the low-frequency power from entering the seasonal component.

Step 5: Calculate the trend component $Y_t - S_t^{(k+1)}$ is computed to remove the seasonal component, and this difference series is then smoothed using Loess to obtain the trend component $T_t^{(k+1)}$ at the $(k+1)$ th iteration.

The main task of the outer loop is to find the robust weight ρ_t used in the inner loop to perform LOESS. It is assigned to each time point according to the size of the residual component. The seasonal and trend components obtained in the inner loop are used to compute the residual component R_t. A larger R_t corresponds to a smaller ρ_t, which reflects the extreme degree of R_t. It is calculated using the following formula:

$$h = 6\text{median}(|R_t|) \tag{2}$$

$$B(u) = \begin{cases} 0, u \geqslant 1 \\ (1-u^2), 0 \leqslant u < 1 \end{cases} \tag{3}$$

$$\rho_t = B(|R_t|/h) \tag{4}$$

where median is the median function; B is the double square weight function.

The STL model is an additive model, that is ideal for sequences with relatively consistent time scales. Most time series data, however, exhibit a pattern of changing cycles with trends. In such cases, the multiplicative model is more appropriate. The multiplicative form of the STL can be achieved through the logarithmic function, as shown in (5). From the analysis in Fig. 2, it is clear that the load decomposition should be conducted using the multiplication-based form of the STL.

$$\log Y_t = \log T_t + \log S_t + \log R_t \tag{5}$$

Combining (5) and the STL calculation process to decompose the electricity load, heating load, and cooling load sequences, the set of trend components, the set of seasonal components, and the set of residual components for the three types of loads are obtained. The load components mentioned below are all the same kind of component sets of multienergy loads.

C. Analysis of the Influencing Factors of Multi-energy Load Components

To filter out irrelevant features and reduce the computational burden, as well as to provide more refined data support for load decomposition forecasting, MIC correlation analysis is performed on the load components and meteorological features. The MIC is a measurement method based on mutual information that evenly covers various relationships, both linear and nonlinear. It is commonly used for feature selection in deep learning due to its universality and low computational complexity. The basic idea of MIC is to measure the relationship between two variables using a two-dimensional scatter plot while estimating the probability density distribution of variables in the grid of the scatter plot. Normalization is subsequently performed to obtain the MIC value. The range of MIC values is [0,1], with a higher value indicating a deeper correlation. The MIC calculation process can be summarized as follows:

$$f_{\mathrm{MIC}}(x,y) = \max_{a*b<n^{0.6}} \frac{I(x,y)}{\log_2 \min(a,b)} \tag{6}$$

$$I(x,y) = \sum p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{7}$$

where a and b respectively represent the number of grids in the x and y directions in the two-dimensional coordinate system; n is the sample size; $I(x,y)$ is the mutual information between variables x and y; $p(x)$ and $p(y)$ denote the probability density distribution of x and y; $p(x,y)$ represents the joint probability density distribution of x and y.

Based on (6) and (7), the MIC values between the load components and meteorological factors are calculated, and features with strong correlations are retained. Moreover, hourly and holiday features are constructed considering the characteristics of the time series. The hourly feature corresponds to encoded values of 0–23, and the holiday feature is set to 1 for a working day and 0 for a nonworking day.

III. Methodology

In the proposed model, the CNN network is coupled with the PLE sharing method to build a convolutional bottom-level expert network. There are two types of experts: shared experts who focus on understanding the coupling relationship between loads, and task-specific experts who focus on capturing load-exclusive information. A progressive routing strategy is applied to separate the specific parameters of different load forecasting tasks. The feature information obtained from various load forecasting tasks is then fed into the LSTM network to capture time dependencies separately.

A. CNN Method

Convolutional neural networks (CNN) are widely employed for feature extraction from load data [24]. The powerful information extraction abilities of CNNs allow for the comprehensive extraction of hidden spatial coupling information from multiple energy loads and meteorological factors [25]. The core of CNNs is to establish multiple convolutional kernels to extract local features and share the weight parameters in each convolutional kernel. Different local sensory fields are synthesized to obtain global information. The principle is depicted in Fig. 3, where the convolution kernel size is 3 and the stride size is 1, moving along the time direction.

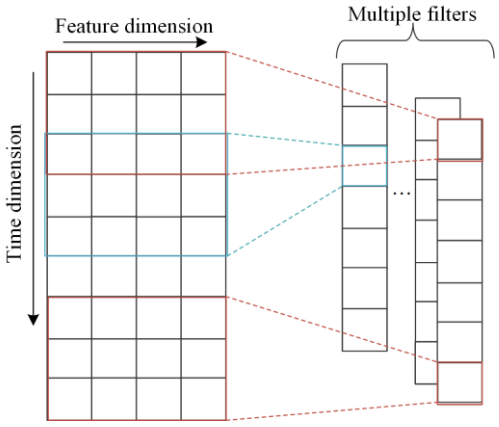


Fig. 3. One-dimensional convolution principle.

B. LSTM Method

To better learn the temporal features of multienergy load data, an LSTM is utilized to analyze the feature

vectors extracted by the CNN. The LSTM is an enhanced version of a recurrent neural network (RNN) that has demonstrated remarkable performance in modeling and predicting time series data. The gate mechanisms integrated into the LSTM regulate the flow and retention of features, amplifying its ability to manage long short-term temporal dependencies. Further information regarding the fundamental principles of LSTMs can be obtained from reference [39].

C. Multitask Method Based on Progressive Layered Extraction

Currently, most of the multitask methods applied in the field of joint forecasting of multiple energy loads involve traditional hard parameter sharing, i.e., using the same hidden layer in all tasks and reserving different output layers for specific tasks. This sharing approach is too simple because it fails to consider differences in the level of coupling among different loads and is susceptible to harmful interference caused by weak correlations between subtasks. In addition, the complex correlation between loads often results in a seesaw phenomenon, which makes it difficult to achieve significant improvements in all tasks compared to the corresponding single-task model. Therefore, to reduce the effects of loose tasks and improve the efficiency of joint learning, the PLE multitasking strategy is adopted.

PLE [36] improves upon MMoE [40]. The experts are explicitly divided into task-sharing experts and task-specific experts. Sharing experts focus on learning shared patterns, while task-specific experts work on learning task-specific patterns. This division helps to avoid parameter conflicts and negative migration between different tasks for customized sharing. On the information interaction side, unlike fully connected routing in MMoE, PLE involves more flexible and efficient progressive separation routing. In MMoE, there is no distinction or collaboration among experts, and all tasks use the same expert output. In contrast, PLE uses all lower-level experts as feature extractors and applies progressive routing to extract higher-order shared knowledge and gradually separate task-specific parameters at higher levels. Specifically, PLE customizes gating units for each type of task and selects the required information from shared and exclusive experts for the task. The customized gating network eliminates the connection between the subtask layer and other task-exclusive experts, allowing task-specific experts to focus on their own knowledge learning. The gating unit corresponding to the shared expert draws information from the network of all experts. This implies that shared experts are influenced by all tasks, while the parameters of task-specific experts are influenced only by their own tasks. Figure 4 shows a comparison of the structures of the three models.

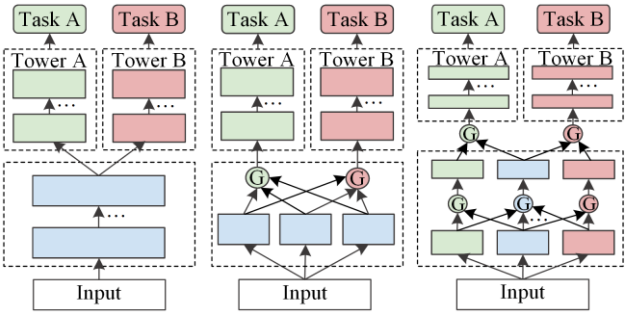


Fig. 4 Framework of different multitask methods. (a) Hard parameter sharing. (b) MMoE. (c) PLE.

PLE consists of expert modules, gating units, and tower networks. The expert modules and gating units are located at the bottom and can form a multilayer feature extraction network. Likewise, a tower network also consists of a multilayer network placed on top of the model. Each expert module contains multiple subnetworks to support feature learning.

In PLE, gating units selectively fuse different expert modules. As shown in Fig. 5, each expert module has a gate unit, which is composed of a single-layer fully connected network with a softmax activation function. Here, the input acts as a selector to assign weights to the selected vectors, resulting in the output of the gating unit. If expressed by a formula, the output of the gating network is:

$$g^{k,j}(x) = w^{k,j}(x) S^{k,j}(x) \tag{8}$$

$$w^{k,j}(x) = \text{softmax}\left[\boldsymbol{W}_g^k (g^{k,j-1}(x))\right] \tag{9}$$

where x is the input; the output of the ith subtask at the jth layer network is given by $g^{k,j}(\boldsymbol{x})$; the underlying network has a total of N layers; the weight parameter of the task k at layer j, denoted as $w^{k,j}(\boldsymbol{x})$ is calculated by the linear matrix $\boldsymbol{W}_g^k \in \mathbf{R}^{(m_k+m_s)\times d}$ and softmax function; m_s and m_k respectively denote the number of shared and task-specific experts; d represents the input dimension; the selection matrix of the task k at layer j is $S^{k,j}(\boldsymbol{x})$, composed of the shared module $\boldsymbol{E}_s$ and the task-specific module $\boldsymbol{E}_k$, as expressed in (10).

$$S^{k,j}(\boldsymbol{x}) = [\boldsymbol{E}_{(k,1),j}^{\mathrm{T}}, \boldsymbol{E}_{(k,2),j}^{\mathrm{T}}, \ldots, \boldsymbol{E}_{(k,m_k),j}^{\mathrm{T}}, \boldsymbol{E}_{(s,1),j}^{\mathrm{T}}, \boldsymbol{E}_{(s,2),j}^{\mathrm{T}}, \ldots, \boldsymbol{E}_{(s,m_s),j}^{\mathrm{T}}]^{\mathrm{T}} \tag{10}$$

Unlike the selection matrix of a task-specific module, the selection matrix of the shared module is composed of all experts within that particular layer.

After the above calculations, the predicted result of task k is obtained via (11). Here, t^k denotes the tower network for task k.

$$y^k(x) = t^k(g^{k,N}(\boldsymbol{x})) \tag{11}$$

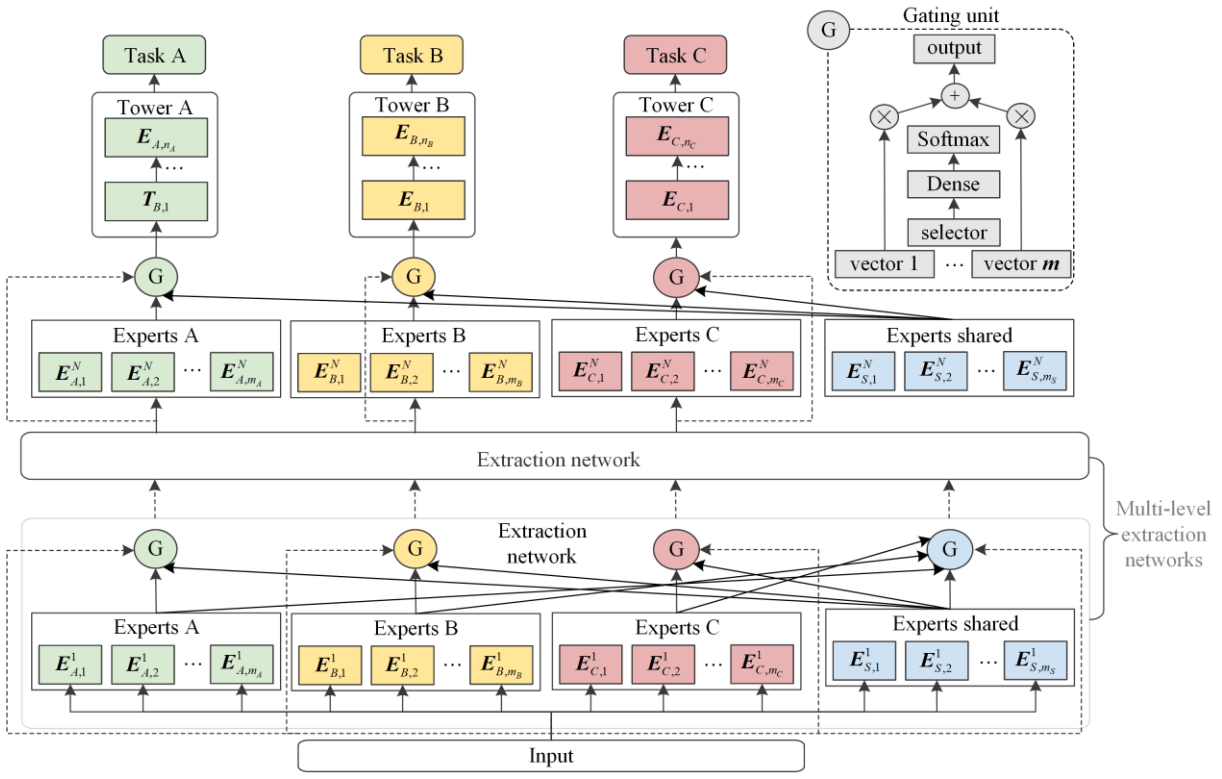


Fig. 5. Structural details of PLE.

D. Model Framework

In light of the above methods, on the basis of PLE, we propose a collaborative load forecasting model called CPLE, as depicted in Fig. 6. We utilize a 1DCNN instead of FC as the feature extraction unit to extract information from input sequences. Compared with the FC-based PLE, using a convolutional network with powerful information extraction abilities can better extract hidden spatial coupling information from multiple energy loads and meteorological factors. Concurrently, because of the remarkable performance of LSTM in predicting time series, a tower network is built using LSTM to learn temporal relations from multienergy data.

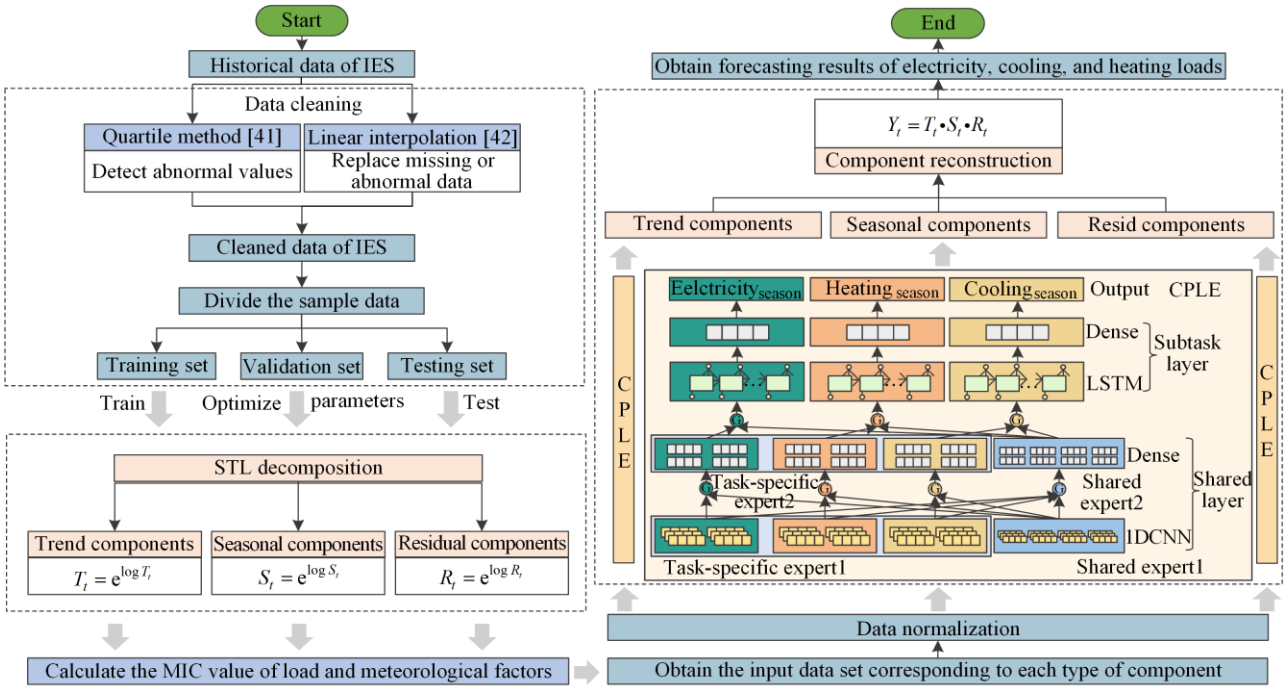


Fig. 6. Flowchart of proposed forecasting approach.

The forecasting method consists of two main stages: 1) Load decomposition and analysis: The raw data are first cleaned to remove missing or abnormal values [41], [42]. The cleaned electricity, cooling, and heating loads

are decomposed using STL to obtain the corresponding trend, seasonal, and residual components. Then, the MIC values between the components and the meteorological factors are calculated to screen for strongly correlated meteorological features. Finally, the input feature sets are constructed separately for different load components by combining the time features.

2) Component modeling and prediction: The input feature sets obtained in stage (1) are normalized, and corresponding CPLE prediction models are built for each type of load component. The obtained seasonal, trend, and residual component predictions are reorganized to derive the predicted values of electricity, cooling, and heating loads. The optimal parameters of the proposed model are determined via a grid search [43], as shown in Table I.

TABLE I
PROPOSED MODEL PARAMETERS

Parameter type	Parameter name	Network type	Number	Value
Architecture	Shared expert1	1DCNN	4	16
	Task-specific expert1	1DCNN	2/2/2	16
	Kernel size			3
	Shared expert2	Dense	8	8
	Task-specific expert2	Dense	4/4/4	8
	Subtask layer	LSTM	1/1/1	16
		Dense	1/1/1	1
Optimization	Loss function		MAE	
	Epoch		300	
	Optimizer		Adam	
	Batch size		64	

IV. CASE STUDY

A. Data Sources and Experimental Environment

The dataset used in this study was collected from the CAMPUS METABOLISM project at Arizona State University [44]. It specifically comes from the Tempe campus and includes electricity, cooling, and heating loads. The time period selected is from 00:00 on January 1, 2020, to 23:00 on November 30, 2021, all with a sampling interval of 1 hour. Since the IES load data have inconsistent units, Equation (12) [45] uniformly converts different types of load units into MWs.

$$1\ \text{MW} = 284\ \text{Ton} = 3.4\ \text{mmBtu/h} \tag{12}$$

The corresponding meteorological data were crawled from the Global Weather Accurate Forecasting Network, specifically from the weather station closest to the Tempe campus (Phoenix Sky Harbor International Airport Station) [46]. The data include 7 categories of meteorological data, namely, temperature, dew point, humidity, wind speed, wind gust, pressure, and precipitation. The software platform is implemented using the TensorFlow and Keras frameworks. The hardware configuration includes two Xeon 4215R CPUs and two NVIDIA RTX 3090 GPUs.

To ensure that the impact of each variable is considered equal, the input data need to be normalized using the following formula:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values for the input feature, respectively.

B. Evaluation Metrics

To assess the performance of the proposed model, the following three evaluation metrics are adopted: root mean square error (RMSE), mean absolute percentage error (MAPE), and weighted mean accuracy (WMA). The specific calculation is given by:

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^{n}(y_{\text{act}}(i) - y_{\text{pred}}(i))^2}{y_{\text{act}}(i)}} \tag{14}$$

$$E_{\text{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{\text{act}}(i) - y_{\text{pred}}(i)}{y_{\text{act}}(i)}\right| \times 100\% \tag{15}$$

where $y_{\text{act}(i)}$ is the true load value at time i; $y_{\text{pred}(i)}$ is the predicted load value at time i; n is the number of samples in the training/validation/testing set.

Considering the varying importance of each load in the IES, the weight values for electricity, cooling, and heating loads in the WMA are set at 0.4, 0.4, and 0.2, respectively [47]. The WMA can be expressed as:

$$E_{\text{MA}} = 1 - E_{\text{MAPE}} \tag{16}$$

$$E_{\text{WMA}} = \alpha_{\text{elec}} E_{\text{MA}_{\text{elec}}} + \alpha_{\text{cool}} E_{\text{MA}_{\text{cool}}} + \alpha_{\text{heat}} E_{\text{MA}_{\text{heat}}} \tag{17}$$

C. Dataset Partitioning and Input Feature Construction

To take into account the seasonal differences in load, the data are divided into the seasons of spring, summer, autumn, and winter. For each season, the data are divided into a training set, a validation set, and a testing set at a ratio of 8:1:1, after which they are combined to form the total training set, a validation set, and a testing set [48]. For ease of later expression, the symbols EL, CL, and HL are used to represent the electricity, cooling, and heating loads, respectively. The symbolic representation of the remaining features is presented in Table A1.

According to the concept of dynamic decomposition and prediction [38], the raw load data undergo STL decomposition. The decomposition details of the training set data for the spring, summer, autumn, and winter seasons are presented in Fig. A1. The three types of loads are decomposed into trend, seasonal, and residual terms. This approach separates regular and stochastic load components, which not only clearly represents the trend and cycle change patterns but also avoids noise interference in predicting trend and seasonal components. This decomposition and forecasting approach can assist the model in capturing the changing characteristics of load data from multiple perspectives.

In an attempt to make the prediction model more concise and efficient, the correlation between multivariate loads and meteorological factors is analyzed using the MIC method, and the features with a strong influence on the target variables are screened out. An MIC greater than 0.3 [49] indicates a strong correlation, and an MIC less than 0.1 indicates a weak correlation. The MIC heatmap in Fig. 7 reveals that there is a strong correlation between the multiple energy loads. Loads are strongly correlated with T and DP and moderately correlated with H and P. The

correlations between loads and WS, WG, and Precip are close to zero, so these three meteorological factors are directly excluded from the subsequent analysis.

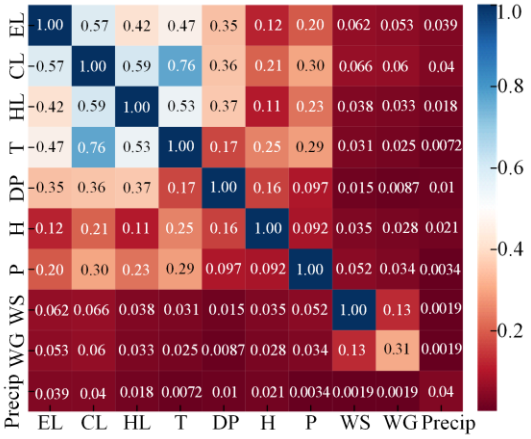


Fig. 7. Heatmap of correlation analysis for multi-energy loads.

Load decomposition breaks down the contained pattern of change into different components that represent different characteristics of the load. The factors affecting the components may vary depending on the type or characteristics of the load they represent. Seasonal variability may also lead to changes in the factors affected [49]. Hence, it is necessary to select appropriate input characteristics for forecasting different load components during different seasons. The MIC values between the load components and weather data during the different seasons are calculated, as shown in Table A2. From the table, it can be observed that the weather factor has a strong influence on the set of load trend components and less influence on the set of seasonal and residual components. In summer, the multivariate load trend term strongly correlated with DP and H, and the mean MIC value was greater than 0.3. In winter, there was a strong correlation between T and DP, and the mean MIC value was greater than 0.3. Based on the results of the correlation analysis, input features are constructed for the three types of load components under different seasons. The specific input datasets are shown in Table A3.

D. Experimental Framework

In this paper, we primarily validate the efficacy of the STL-CPLE model through two sets of experiments, as illustrated in Fig. 8. For load forecasting models, we compare single-task models such as LSTM, CNN-LSTM, TCN [26], and DA-Seq2Seq based on LSTM [27]; among them, CNN-LSTM serves as the corresponding single-task model for CPLE. Moreover, we compare multitask models, including the CNN-BiGRU-Attention [31], transformer, LSTNet-Skip [47], and MMoE [34]. Furthermore, ablation experiments are conducted for CPLE. Regarding the load decomposition methods, the CPLE is considered the baseline. No decomposition, EMD decomposition, CEEMD decomposition, or STL decomposition based on additive forms are used for comparison. The parameters of the compared models employed in our study are presented in Table II. The corresponding input characteristics can be found in Appendix Table A4. All models have a forecast window of 12 hours and a forecast step of 1 hour.

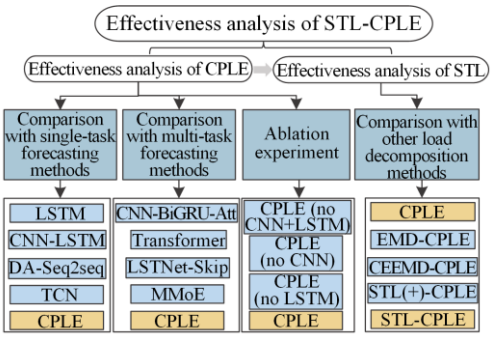


Fig. 8. Experimental framework.

TABLE II
MAIN PARAMETERS OF THE COMPARISON MODELS

Model type	Model name	Hyperparameter	Value
Single task	LSTM	Number of LSTM hidden units	16
	CNN-LSTM	Number of CNN hidden units	16
		Kernel size	3
		Number of dense hidden units	8
		Number of LSTM hidden units	16
	TCN [26]	Number of TCN hidden units	32
		Kernel size	4
		Expansion factor	[1,2,4]
		Dropout	0.2
	DA-Seq2Seq (LSTM) [27]	Encoder number of hidden units	128
		Decoder number of hidden units	128
		Number of attention layers	2
		Number of attention hidden units	[timesteps, features]
Multitask	CNN-BiGRU-Attention [31]	Number of CNN hidden units	[4,8]
		Kernel size	[3,3]
		Number of BiGRU hidden units	10
		Number of attention layers	3
		Number of attention hidden units	64
	Transformer	Number of transformer blocks	2
		Number of heads	4
		Head size	16
		Number of dense hidden units	16
	LSTNet-Skip [47]	Number of CNN hidden units	100
		Number of RNN hidden units	100
		Skip length of skip-RNN layer	4
		Kernel size	6
		Dropout	0.2
	MMoE [34]	Number of experts	4/4/4/4
		Number of expert hidden units	32/32/32/32
		Number of sub-task-layer (LSTM) hidden units	[16,8]/ [16,8]/ [16,8]
		Dropout	0.3

E. Comparison with Single-task Forecasting Models

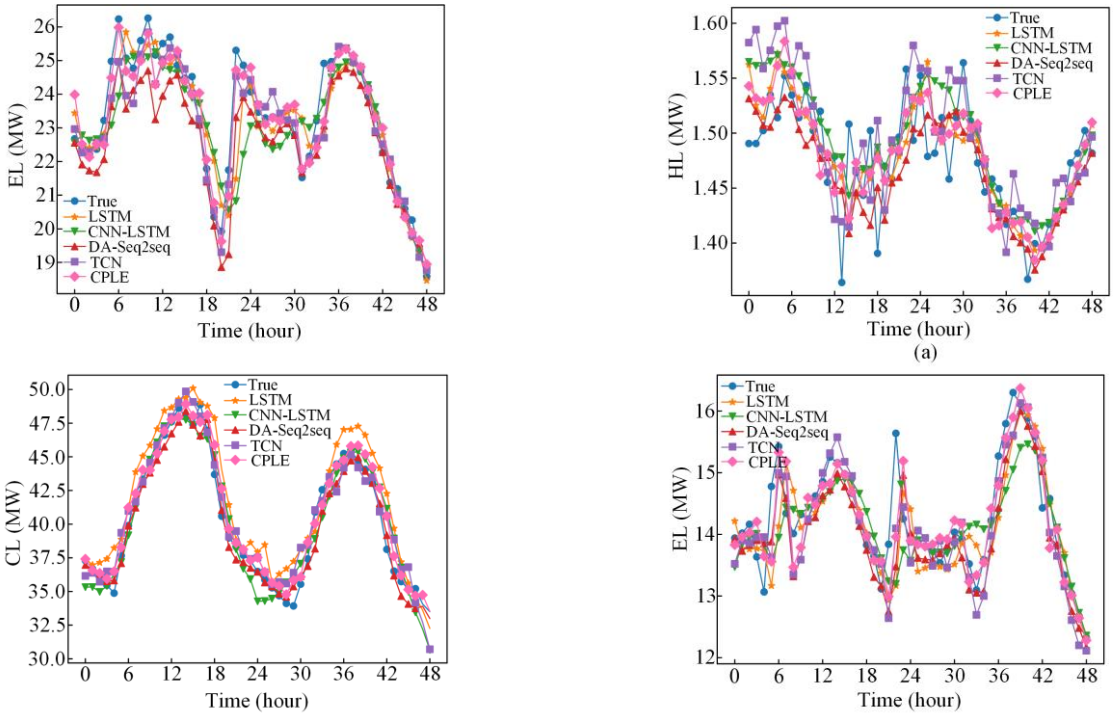
To verify the superiority of the proposed model in multienergy load forecasting, it is first compared to four models in single-task mode. The detailed prediction results of the model are presented in Table III. The overall performance comparison of the proposed model with other models is shown in Table IV. From the test sets of summer and winter, two consecutive days are chosen to compare the predicted and true values, as displayed in Fig. 9. The specific time ranges are shown in Table A5.

TABLE III
MAPE, RMSE, AND WMA EVALUATION METRICS FOR SINGLE-TASK PREDICTION RESULTS

Season	Model	RMSE (MW)			MAPE (%)			WMA (%)
		EL	CL	HL	EL	CL	HL	
Spring	LSTM	0.97	0.89	0.07	4.37	2.99	3.11	96.43
	CNN-LSTM	0.89	0.90	0.06	3.25	3.07	2.62	96.94
	TCN	0.92	0.93	0.07	4.23	2.95	3.08	96.51
	DA-Seq2Seq	0.72	1.06	0.06	2.82	3.71	2.64	96.86
	CPLE	0.75	0.75	0.05	2.96	2.14	2.43	97.47
Summer	LSTM	1.22	1.61	0.05	3.80	2.98	2.97	96.69
	CNN-LSTM	1.01	1.81	0.04	2.92	3.43	2.55	96.95
	TCN	0.85	1.40	0.05	2.61	2.50	3.02	97.35
	DA-Seq2Seq	1.20	1.15	0.04	3.97	2.11	2.56	97.05
	CPLE	0.70	1.00	0.04	2.16	1.71	2.55	97.94
Autumn	LSTM	0.99	1.64	0.09	4.80	8.27	4.60	93.85
	CNN-LSTM	0.85	0.82	0.08	3.68	3.83	3.95	96.20
	TCN	0.82	1.26	0.08	3.87	5.75	3.83	95.39
	DA-Seq2Seq	0.81	0.86	0.06	3.84	4.49	2.84	96.10
	CPLE	0.75	0.51	0.06	3.27	2.47	2.97	97.11
Winter	LSTM	0.75	0.68	0.12	3.72	4.95	3.26	95.88
	CNN-LSTM	0.60	0.61	0.11	2.78	4.48	2.99	96.50
	TCN	0.62	0.57	0.11	2.86	3.88	2.99	96.71
	DA-Seq2Seq	0.57	0.48	0.11	2.59	3.46	3.25	96.93
	CPLE	0.47	0.38	0.08	1.99	2.53	2.28	97.74

TABLE IV
COMPARISON OF THE OVERALL ACCURACY IMPROVEMENT OF THE CPLE MODEL WITH THAT OF OTHER SINGLE-TASK MODELS

Model	RMSE (%)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CPLE vs. LSTM	31.78	45.19	25.58	37.79	53.83	26.67	1.93
CPLE vs. CNN-LSTM	20.33	36.09	16.88	17.81	40.16	15.64	0.95
CPLE vs. TCN	14.49	31.20	42.30	21.00	40.81	39.30	1.27
CPLE vs. DA-Seq2Seq	19.03	25.65	10.80	21.44	35.59	9.38	0.85



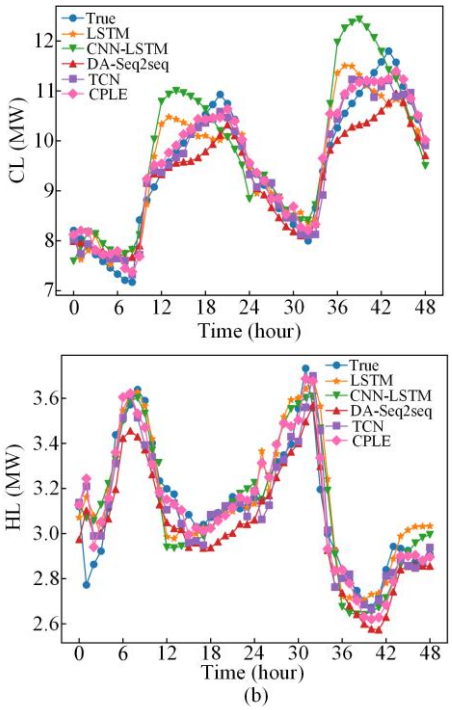


Fig. 9. Comparison of the EL, HL, and HL curves for single-task models. (a) Summer. (b) Winter.

The prediction curves of the model in this paper are more consistent with the real load curves. The MAPE and RMSE of the CPLE are smaller than those of the single-load prediction methods, and the improvement in CL forecasting is more significant. The proposed model reduces the MAPE of EL, CL, and HL by 17.81%, 40.16%, and 15.64%, respectively, compared with the corresponding single-task model called CNN-LSTM. The MAPE of the three types of loads is still reduced by 21.44%, 35.59%, and 9.38%, respectively, compared to the best prediction achieved by the single-task model, namely, DA-Seq2Seq. Even with the inclusion of other loads as auxiliary features, the single-task approach still struggles to effectively exploit the correlation and complementarity among IES loads.

F. Comparison with Multitask Forecasting Models

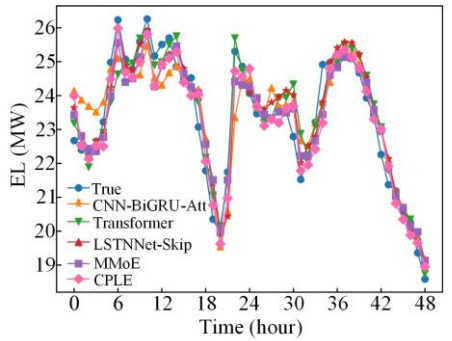
Next, comparisons are made with other multitask forecasting models, namely, the CNN-BiGRU-Attention, transformer, LSTNet-Skip, and MMoE. The first three models employ traditional hard sharing methods; the MMoE adopts a sharing method involving multiple gates and multiple experts, while the CPLE utilizes a sharing method with progressive hierarchical extraction. Table II presents the parameters of the compared models, Table V displays the prediction errors of these models on four test sets (spring, summer, autumn, and winter), and Table VI shows the overall performance of the CPLE in comparison to other models. Figure 10 presents the true and predicted curves for 48 consecutive hours in summer and winter, and the time ranges are listed in Table A5. A detailed analysis of the prediction results follows.

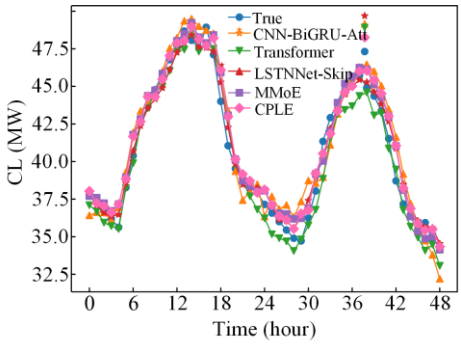
TABLE V
MAPE, RMSE, AND WMA EVALUATION METRICS FOR MULTITASK PREDICTION RESULTS

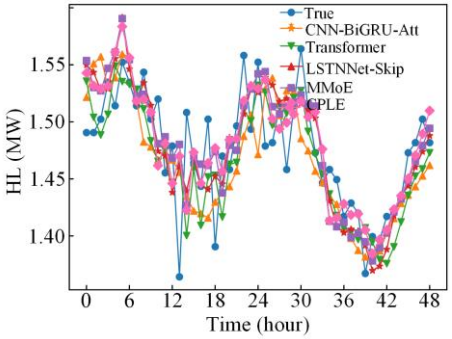
Season	Model	RMSE (MW)			MAPE (%)			WMA (%)
		EL	CL	HL	EL	CL	HL	
Spring	CNN-BiGRU-Att	0.81	0.73	0.07	3.30	2.13	3.17	97.19
	Transformer	0.81	0.60	0.07	3.16	1.79	2.95	97.43
	LSTNet-Skip	0.87	0.68	0.06	3.95	2.15	2.57	97.05
	MMoE	0.84	0.66	0.08	3.36	2.03	3.52	97.14
	CPLE	0.76	0.75	0.06	2.96	2.15	2.43	97.47
Summer	CNN-BiGRU-Att	0.95	1.44	0.05	2.92	2.63	2.77	97.23
	Transformer	0.71	1.21	0.05	2.22	2.23	2.89	97.64
	LSTNet-Skip	0.75	1.18	0.05	2.30	2.11	2.62	97.71
	MMoE	0.67	1.08	0.04	2.07	1.94	2.38	97.92
	CPLE	0.71	1.01	0.05	2.16	1.72	2.55	97.94
Autumn	CNN-BiGRU-Att	0.82	0.56	0.09	4.05	2.74	3.88	96.51
	Transformer	0.79	0.54	0.09	3.94	2.59	4.14	96.56
	LSTNet-Skip	0.78	1.10	0.07	3.61	5.58	3.23	95.68
	MMoE	0.72	0.60	0.07	3.04	2.89	3.32	96.96
	CPLE	0.75	0.51	0.07	3.27	2.47	2.97	97.11
Winter	CNN-BiGRU-Att	0.58	0.63	0.15	2.63	4.64	4.28	96.23
	Transformer	0.60	0.86	0.11	2.63	6.06	3.03	95.92
	LSTNet-Skip	0.51	0.39	0.09	2.34	2.69	2.40	97.51
	MMoE	0.49	0.40	0.09	2.14	2.90	2.53	97.48
	CPLE	0.47	0.39	0.09	1.99	2.53	2.28	97.74

TABLE VI
COMPARISON OF THE OVERALL ACCURACY IMPROVEMENT OF THE CPLE MODEL WITH OTHER MULTI-TASK MODELS

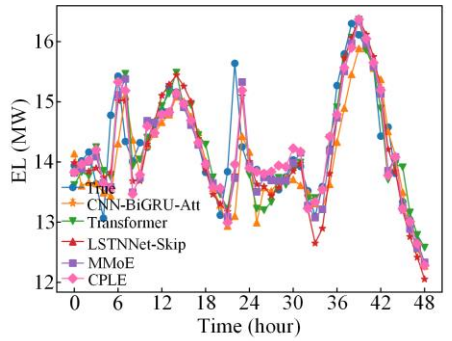
Model	RMSE (%)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CPLE vs. CNN-BiGRU-Att	12.74	13.18	17.68	16.01	8.09	17.32	0.40
CPLE vs. Transformer	7.38	17.37	19.40	13.09	29.99	21.43	0.70
CPLE vs. LSTNet-Skip	7.57	20.77	5.54	14.82	29.22	5.51	0.59
CPLE vs. MMoE	1.36	3.21	12.03	2.15	9.08	13.05	0.19

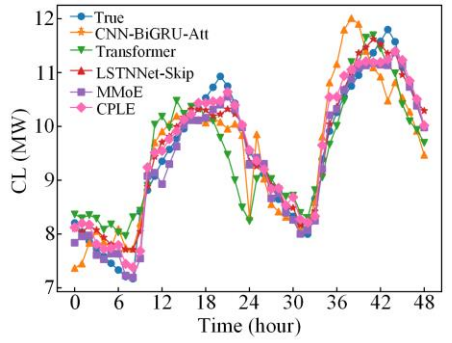


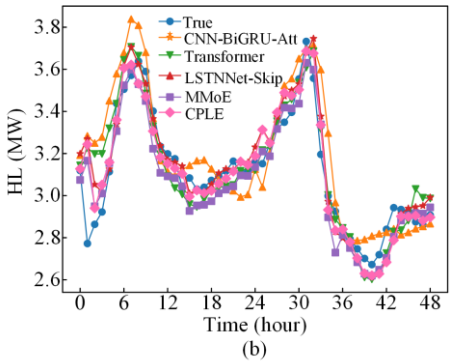




(a)







(b)

Fig. 10. Comparison of the EL, HL, and HL curves for multi-task models. (a) Summer. (b) Winter.

From the results, the CPLE has the highest accuracy, followed by the MMoE, CNN-BiGRU-Attention, LSTNet-Skip, and the transformer. Compared with other models, CPLE significantly improves the accuracy of HL and CL. This indicates that the proposed model fully utilizes the complementarity between HL and CL. The model based on traditional sharing is the least effective. This is attributed to the architecture's complete sharing of underlying parameters, thereby neglecting differences in coupling levels among multiple loads. The two models, LSTNet-Skip and the transformer, are less effective than CNN-BiGRU-Attention, probably because of the limitations of the sliding window. They are more suitable for dealing with the prediction of long time series.

Compared to CNN-BiGRU-Attention, which is the most effective of the traditional hard-sharing methods, MMoE achieves an average reduction of 0.12% in the MAPE and 9.21% in the RMSE. The WMA improves by 0.21%. This indicates that the MMoE improves the number of prediction points where large errors exist. As depicted in Fig. 10, the MMoE notably diminishes prediction errors at peak and trough values, especially for HL in summer and CL in winter. These findings validate that the integrated approach of multiple experts and multiple gates enables the shared network to comprehensively learn load patterns while optimizing task information allocation. According to Table V, the WMA of MMoE is lower than that of CNN-BiGRU-Attention in spring. Although the performance of CL is good, the accuracy of HL and CL is sacrificed, resulting in a decrease in the overall performance of joint prediction. This confirms the limitations of the MMoE sharing approach. It still shares all experts across all tasks, and relying only on gated networks may make it difficult to model more complex relationships between loads.

CPLE has better performance than CNN-BiGRU-Attention, both in the overall evaluation and in the seasonal evaluation. Compared with the MMoE, the CPLE performs better on the overall test set. In the season test sets, the WMA indicators are better, and the MAPE and RMSE indicators lead in most cases.

Despite the superiority of the CPLE over these two models, the influence of other factors, such as network type or parameters, cannot be excluded. Therefore, to verify that the "customized expert and gating + progressive separation routing" sharing method of the proposed model is better than the traditional hard sharing method and the "undifferentiated expert and gating + fully connected routing" sharing method of MMoE, we set only shared methods differently, reconstructing CNN-LSTM as well as MMoE for comparison. The reconstructed CNN-LSTM is denoted as CNN-LSTM*, and the MMoE is denoted as MMoE*. Table VII provides a performance comparison with the PLE method, and the detailed results are found in Table A6 and Fig. A2. The time ranges of the figure remain unchanged. The following analysis is performed in conjunction with Tables III and IV.

TABLE VII
COMPARISON OF THE OVERALL ACCURACY IMPROVEMENT OF THE PLE SHARING METHOD THAT OF OTHER SHARING METHODS

Model	RMSE (%)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CPLE vs. CNN-LSTM*	13.39	31.22	47.72	15.49	40.64	46.56	1.29
CPLE vs. MMoE*	2.72	4.91	0.39	0.59	7.37	0.42	0.09

Among the three shared methods, the PLE method is the most effective, and the traditional method is the least effective. The prediction accuracy of CNN-LSTM* is even lower than that of the corresponding single-task method, with serious negative migration problems. The effect of MMoE* is closer to that of CPLE than that of the original MMoE, but there is still a certain gap. CPLE outperforms MMoE* on all the evaluation metrics for the overall test set. In the season test sets, the WMA of CPLE also outperforms that of MMoE*, but the MAPE and RMSE are still not fully greater than those of MMoE*. These findings suggest that although the PLE method can alleviate the seesaw phenomenon, it cannot entirely eliminate it. Overall, PLE can improve the performances of the EL, HL, and CL forecasting tasks simultaneously. This contributes to the efficiency of joint learning.

G. Ablation Study

To further validate the necessity of each component of the CPLE model, ablation experiments are conducted for analysis. The contribution of each component to the model performance is evaluated by reducing the number of components. Three scenarios are included: 1) The CNN module is the only module removed (C). 2) The LSTM module is the only module removed (L). 3) The LSTM and CNN modules are simultaneously removed (C+L).

The results are summarized in Table VIII and Fig. A3. The time ranges for the comparison graphs are the same as those for the experiments above, as shown in Table A5. After removing the CNN module and the LSTM module separately, the performance of the model decreases, but it is better than that of the model with both the CNN module and the LSTM module removed. This suggests that the proposed method for enhancing the shared layer and subtask layer is efficient, and their amalgamation can further enhance the prediction accuracy. This is because, unlike a general FC, the 1DCNN acting as the feature extraction layer can explore the invisible correlations between various energies efficiently. It also provides coupled feature details that preserve temporal characteristics for the following phase. The distinct structure of an LSTM network based on an RNN allows it to learn longer-term feature patterns and capture more complete temporal information. A CNN alone cannot capture long-term temporal dependencies, whereas an LSTM alone fails to identify spatially coupled features effectively. Both methods have limitations. Hence, the proposed model combines CNN and LSTM networks to exploit both spatial and temporal relationships between various energy loads while resolving task conflicts through PLE based on a progressive routing mechanism, which significantly enhances the expressive power of the model.

TABLE VIII
ABLATION STUDY OF THE CPLE

Model	RMSE (MW)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CPLE	0.67	0.66	0.06	2.60	2.22	2.56	97.56
–C	0.70	0.65	0.07	2.66	2.27	2.78	97.47
–L	0.71	0.66	0.07	2.66	2.31	2.79	97.45
– (C+L)	0.72	0.70	0.07	2.74	2.44	2.88	97.35

H. Comparison Under Different Decomposition Models

To estimate the effectiveness of the proposed STL method, i.e., the multiplicative form of STL, we compare no decomposition, EMD, CEEMD, and STL (+), i.e., the additive form of STL, based on the CPLE model. The period window of the STL is set to coincide with the time step of the input features, i.e., 12 hours. The

parameters of STL (+) were the same as those of STL. The parameter configurations of EEMD and CEEMD are shown in Table A7. In addition, to optimize the IMF features obtained via EEMD and CEEMD, the features are processed via frequency division based on the sample entropy and the rate of over zeroing. Table A4 displays the input features corresponding to the four categories of decomposition methods. The evaluation metric values for the different decomposition methods can be found in Table IX. Table A8 shows the test set results for the four seasons. Figure 11 illustrates the forecast results for a nonworking day (2021/8/21 00:00–23:00) in summer and a working day (2021/2/19 00:00–23:00) in winter.

TABLE IX
MAPE, RMSE, AND WMA EVALUATION METRICS FOR CPLE MODELS WITH DIFFERENT LOAD DECOMPOSITION METHODS

Model	RMSE (MW)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CPLE	0.67	0.66	0.06	2.60	2.22	2.56	97.56
EMD-CPLE	0.61	0.67	0.06	2.37	2.97	2.55	97.35
CEEMD-CPLE	0.60	0.68	0.06	2.38	2.41	2.51	97.58
STL (+)-CPLE	0.40	0.39	0.04	1.70	1.30	1.70	98.46
STL-CPLE	0.38	0.41	0.04	1.60	1.38	1.64	98.48

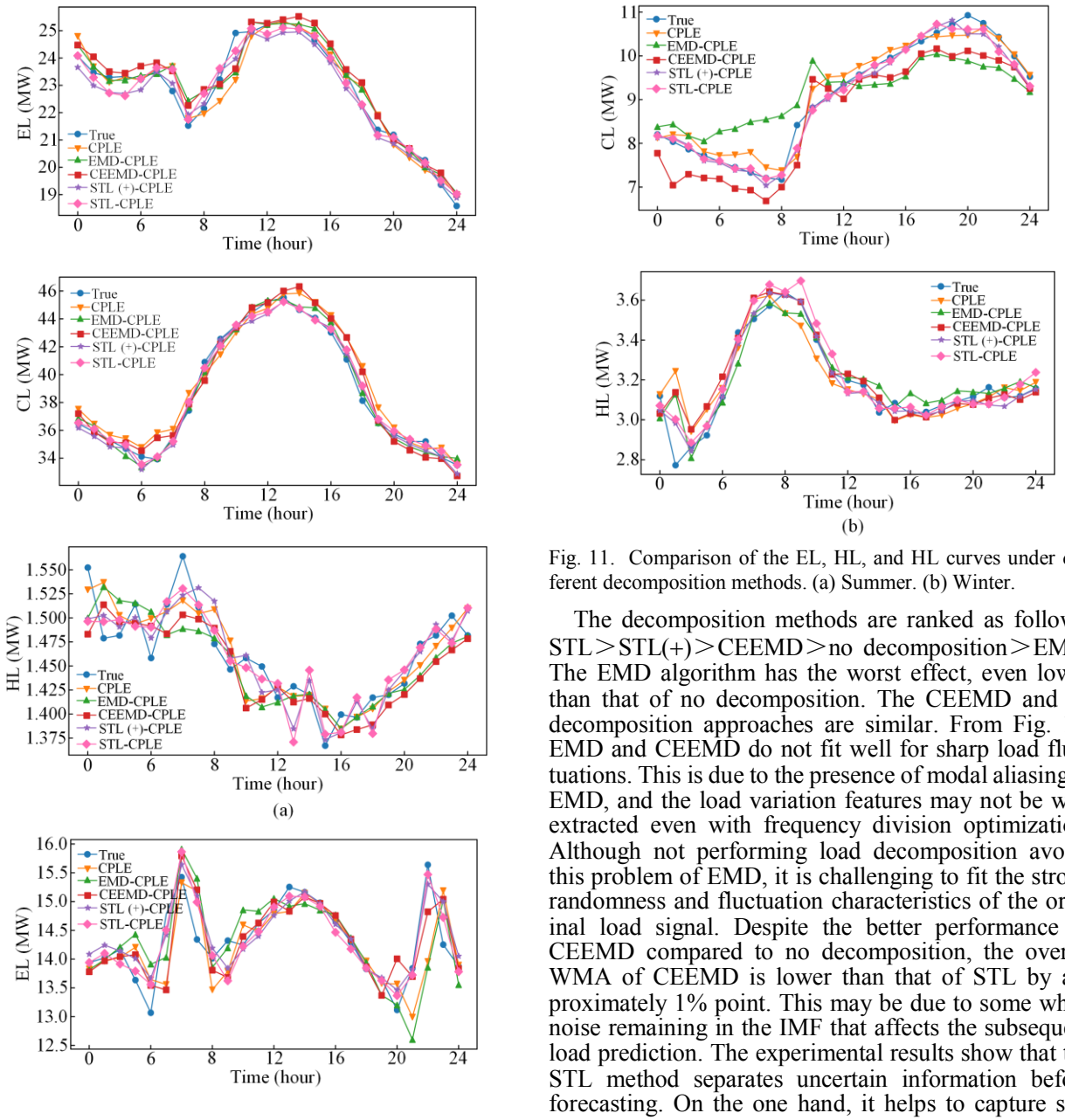


Fig. 11. Comparison of the EL, HL, and HL curves under different decomposition methods. (a) Summer. (b) Winter.

The decomposition methods are ranked as follows: STL＞STL(+)＞CEEMD＞no decomposition＞EMD. The EMD algorithm has the worst effect, even lower than that of no decomposition. The CEEMD and no decomposition approaches are similar. From Fig. 11, EMD and CEEMD do not fit well for sharp load fluctuations. This is due to the presence of modal aliasing in EMD, and the load variation features may not be well extracted even with frequency division optimization. Although not performing load decomposition avoids this problem of EMD, it is challenging to fit the strong randomness and fluctuation characteristics of the original load signal. Despite the better performance of CEEMD compared to no decomposition, the overall WMA of CEEMD is lower than that of STL by approximately 1% point. This may be due to some white noise remaining in the IMF that affects the subsequent load prediction. The experimental results show that the STL method separates uncertain information before forecasting. On the one hand, it helps to capture sto-

chastic fluctuations. On the other hand, it avoids the interference of regular information and can better model the trend and cycle characteristics of the load. The accuracy of STL is greater than that of STL (+), which verifies that the multiplicative form is more suitable than the additive form for addressing the situation in which the load cycle is constantly changing.

V. Conclusion

A short-term collaborative load forecasting method called STL-CPLE is proposed to decouple loads and improve joint learning performance. It has been validated and analyzed using the IES of Arizona State University Tempe Campus, USA. The main conclusions are as follows:

1) The STL decomposition method effectively segregates the distinct characteristics inherent in the load, thereby simplifying its complexity while enhancing the clarity of its physical implications. This approach is conducive to subsequent modeling and forecasting. The MAPEs of EL, CL, and HL are 1.60%, 1.37%, and 1.63%, respectively, and the WMA is 98.48%, achieving fewer errors and greater accuracy than other decomposition methods.

2) The PLE sharing method optimizes the modeling of the coupling relationship between loads and effectively mitigates the negative migration and seesaw phenomenon in the multitask learning method. Compared with the corresponding single task, the MAPE of the three types of loads is reduced by 24.54% on average, and the WMA is improved by 0.95%. Compared with the corresponding traditional sharing method, the MAPE is reduced by 34.32% on average, and the WMA is improved by 1.29%.

3) CPLE is based on the PLE strategy and introduces CNN and LSTM networks for further improvement, which fully exploits the advantages of hybrid networks. Compared with the state-of-the-art approach in the single-task mode, the MAPE is reduced by 22.14% on average, and the WMA is improved by 0.85%. Compared to the state-of-the-art approach in multitask methods, the MAPE is reduced by an average of 8.05%, and the WMA is improved by 0.19%.

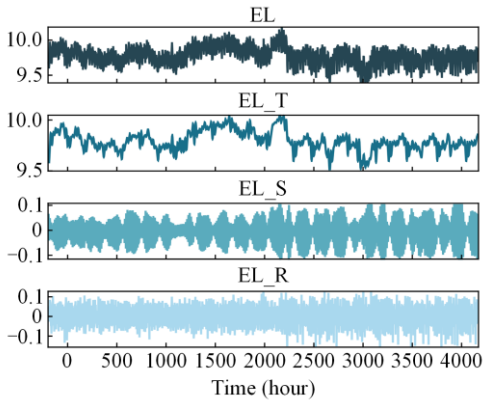
Appendix A

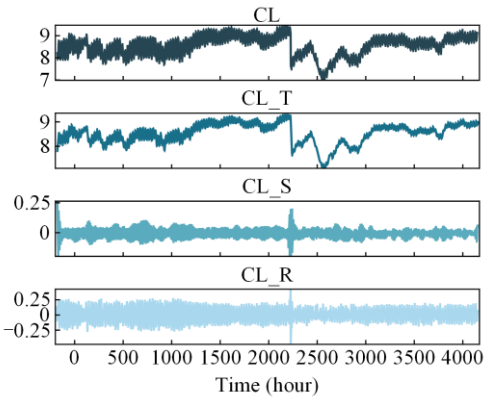
TABLE A1
Symbolic Labels Corresponding to the Feature Variables

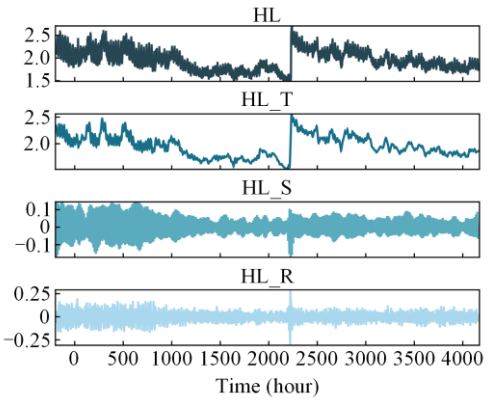
Feature variable	Symbol mark
Temperature	TEMP
Dew point	DP
Humidity	H
Pressure	P
Wind speed	WS
Wind gust	WG
Precipitation	Precip
Hour	hr
Holiday	HOL
Trend components of electricity load, heating load, and cooling load	EL_T, CL_T, HL_T
Seasonal components of electricity load, heating load, and cooling load	EL_S, CL_S, HL_S
Residual components of electricity load, heating load, and cooling load	EL_R, CL_R, HL_R
High-frequency components of electricity load, heating load, and cooling load	EL_H, CL_H, HL_H
Mid-frequency components of electricity load, heating load, and cooling load	EL_M, CL_M, HL_M
Low-frequency components of electricity load, heating load, and cooling load	EL_M, CL_M, HL_M

TABLE A2
MIC Correlation Coefficients of Load Components with Meteorological Factors for Each Season After STL

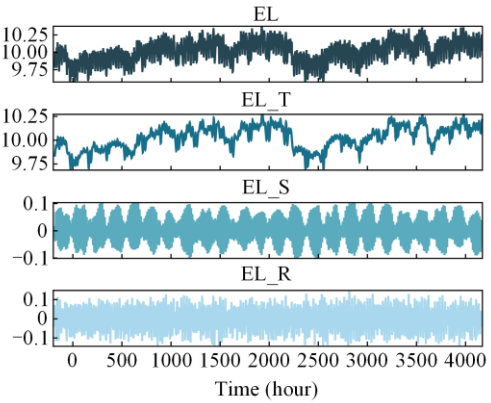
Season	Influencing factor	EL_T	CL_T	HL_T	EL_S	CL_S	HL_S	EL_R	CL_R	HL_R
Spring	T	0.27	0.57	0.43	0.09	0.08	0.20	0.28	0.35	0.31
	DP	0.17	0.13	0.18	0.10	0.09	0.12	0.09	0.11	0.10
	H	0.31	0.34	0.32	0.1	0.09	0.18	0.08	0.21	0.15
	P	0.32	0.25	0.26	0.08	0.08	0.11	0.07	0.12	0.19
Summer	T	0.09	0.10	0.08	0.09	0.09	0.08	0.08	0.10	0.09
	DP	0.46	0.46	0.37	0.08	0.09	0.10	0.10	0.08	0.08
	H	0.38	0.35	0.26	0.08	0.09	0.09	0.09	0.12	0.09
	P	0.12	0.21	0.10	0.13	0.08	0.09	0.09	0.08	0.08
Autumn	T	0.47	0.57	0.25	0.09	0.08	0.15	0.12	0.26	0.16
	DP	0.37	0.38	0.42	0.09	0.08	0.20	0.09	0.11	0.15
	H	0.14	0.15	0.14	0.09	0.08	0.08	0.23	0.32	0.28
	P	0.32	0.31	0.28	0.08	0.08	0.12	0.08	0.15	0.11
Winter	T	0.30	0.38	0.36	0.08	0.08	0.09	0.34	0.47	0.38
	DP	0.37	0.25	0.30	0.10	0.11	0.14	0.09	0.12	0.10
	H	0.14	0.16	0.14	0.09	0.10	0.11	0.22	0.37	0.23
	P	0.12	0.11	0.16	0.09	0.08	0.09	0.10	0.09	0.09

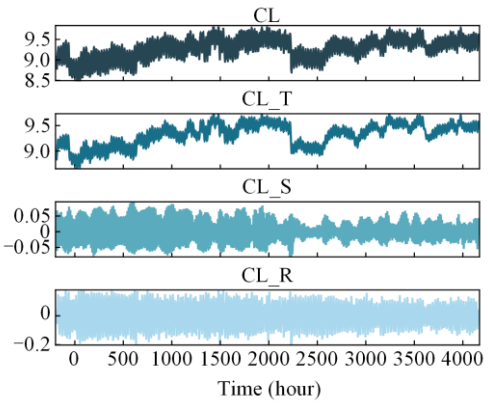

EL
EL_T
EL_S
EL_R
Time (hour)

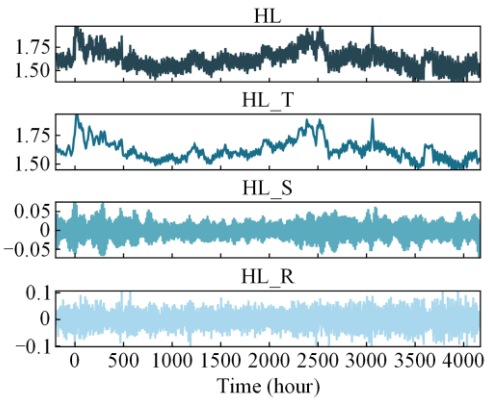

CL
CL_T
CL_S
CL_R
Time (hour)


HL
HL_T
HL_S
HL_R
Time (hour)

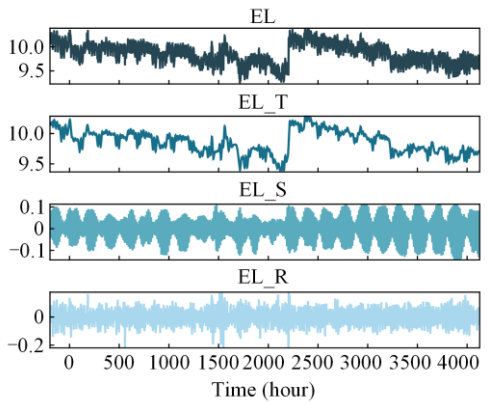
(a)

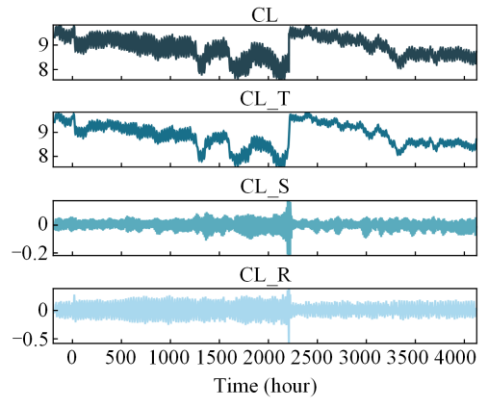

EL
EL_T
EL_S
EL_R
Time (hour)


CL
CL_T
CL_S
CL_R
Time (hour)


HL
HL_T
HL_S
HL_R
Time (hour)

(b)


EL
EL_T
EL_S
EL_R
Time (hour)


CL
CL_T
CL_S
CL_R
Time (hour)

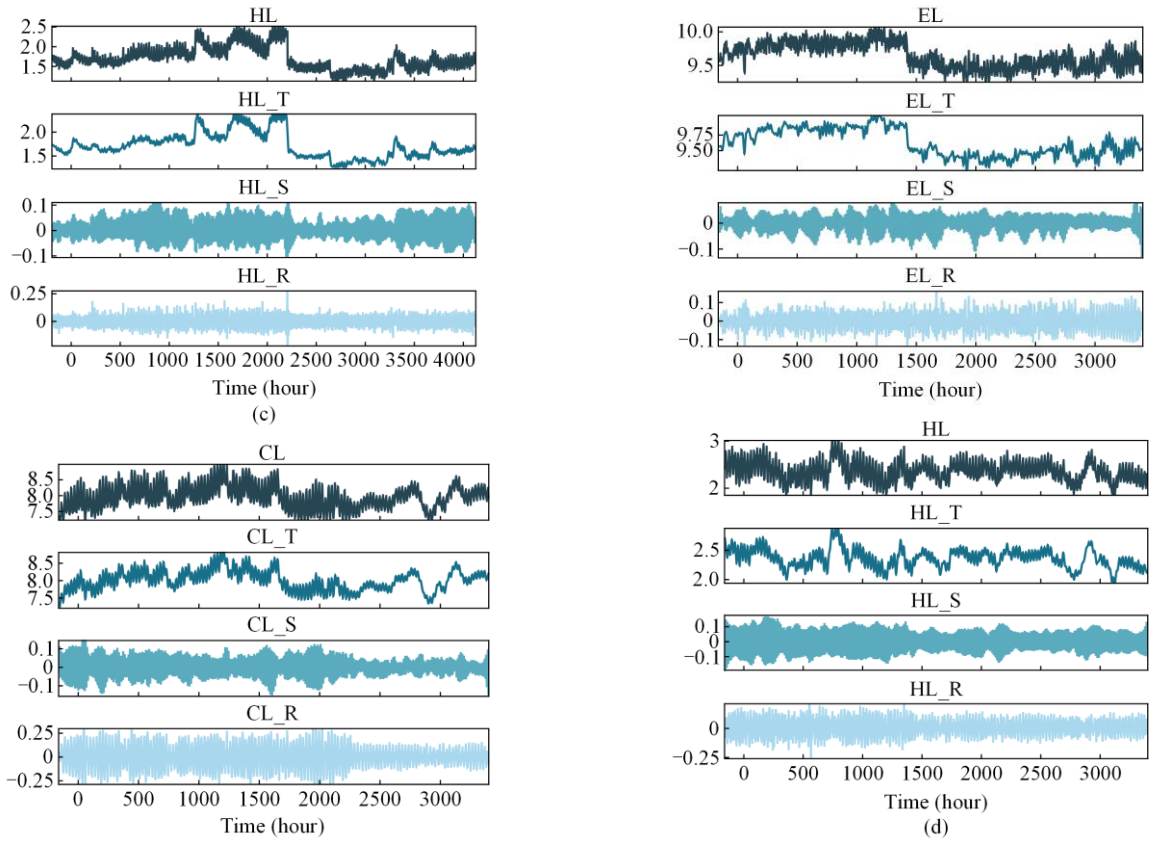


Fig. A1. Decomposition results of multi-energy loads based on STL. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

TABLE A3
INPUT CHARACTERISTICS CORRESPONDING TO DIFFERENT LOAD COMPONENTS

Model type	Spring	Summer	Autumn	Winter
Trend component prediction model	EL_T, CL_T, HL_T, TEMP, H, P, hr, HOL	EL_T, CL_T, HL_T, DP, H, hr, HOL	E_T L, CL_T, HL_T, TEMP, DP, P, hr, HOL	EL_T, CL_T, HL_T, TEMP, DP, hr, HOL
Seasonal component prediction model	EL_S, CL_S, HL_S, hr, HOL	EL_S, CL_S, HL_S, hr, HOL	EL_S, CL_S, HL_S, hr, HOL	EL_S, CL_S, HL_S, hr, HOL
Residual component prediction model	EL_R, CL_R, HL_R, TEMP, H, hr, HOL	EL_R, CL_R, HL_R, hr, HOL	EL_R, CL_R, HL_R, H, hr, HOL	EL_R, CL_R, HL_R, TEMP, H, hr, HOL

TABLE A4
INPUT FEATURES CORRESPONDING TO MODELS UNDER DIFFERENT DECOMPOSITION METHODS. (A) NO DECOMPOSITION ALGORITHM. (B) EMD. (C) CEEMD.

(A)

Spring	Summer	Autumn	Winter
EL, CL, HL, TEMP, H, hr, HOL	EL, CL, HL, DP, H, hr, HOL	EL, CL, HL, TEMP, DP, P, hr, HOL	EL, CL, HL, TEMP, hr, HOL

(B)

Model type	Spring	Summer	Autumn	Winter
High-frequency component prediction model	EL_H, CL_H, HL_H, TEMP, H, hr, HOL	EL_H, CL_H, HL_H, hr, HOL	EL_H, CL_H, HL_H, hr, HOL	EL_H, CL_H, HL_H, TEMP, H, hr, HOL
Mid-frequency component prediction model	EL_M, CL_M, HL_M, hr、HOL	EL_M, CL_M, HL_M, hr, HOL	EL_M, CL_M, HL_M, hr, HOL	EL_M, CL_M, HL_M, TEMP, H, hr, HOL
Low-frequency component prediction model	EL_L, CL_L, HL_L, TEMP, D, H, P, hr, HOL	EL_L, CL_L, HL_L, DP, H, hr, HOL	EL_L, CL_L, HL_L, TEMP, DP, P, hr, HOL	EL_L, CL_L, HL_L, TEMP, H, DP, P, hr, HOL

(C)

Model type	Spring	Summer	Autumn	Winter
High-frequency component prediction model	EL_H, CL_H, HL_H, hr, HOL	EL_H, CL_H, HL_H, hr, HOL	EL _H, CL_H, HL_H, hr, HOL	EL_H, CL_H, HL_H, TEMP, hr, HOL
Mid-frequency component prediction model	EL_M, CL_M, HL_M, hr, HOL	EL_M, CL_M, HL_M, hr, HOL	EL_M, CL_M, HL_M, hr, HOL	EL_M, CL_M, HL_M, TEMP, H, DP, P, hr, HOL
Low-frequency component prediction model	C, TEMP, H, DP, P, hr, HOL	EL_L, CL_L, HL_L, DP, H, hr, HOL	EL_L, CL_L, HL_L, TEMP, DP, P, hr, HOL	EL_L, CL_L, HL_L, TEMP, H, hr, HOL

TABLE A5
TIME RANGES OF THE LOAD CURVES

	Time range	
	Summer	Winter
Start and end time	2021/8/20 00:00–2021/8/21 23:00	2021/2/19 00:00–2021/2/20 23:00
Date type	2021/8/20 is a working day; 2021/8/21 is a non-working day	2021/2/19 is a working day; 2021/2/20 is a non-working day
Number of time points	48	48

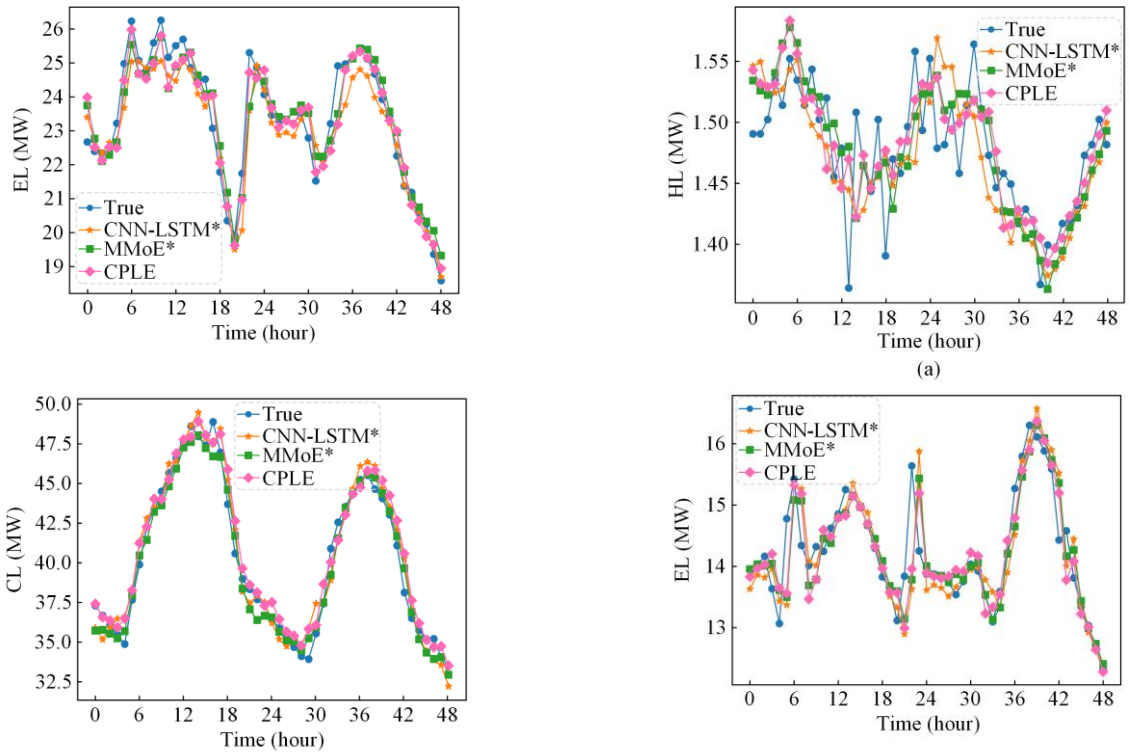
TABLE A6
MAPE, RMSE AND WMA EVALUATION METRICS FOR DIFFERENT MULTITASKING SHARING METHODS. (A) TEST SETS FOR FOUR SEASONS. (B) OVERALL TEST SET.

(A)

Season	Model	RMSE (MW)			MAPE (%)			WMA (%)
		EL	CL	HL	EL	CL	HL	
Spring	CNN-LSTM*	0.86	0.93	0.13	3.44	2.98	5.79	96.27
	MMoE*	0.79	0.68	0.06	2.73	2.17	2.50	97.53
	CPLE	0.75	0.75	0.05	2.96	2.14	2.42	97.47
Summer	CNN-LSTM*	0.91	1.20	0.05	2.71	2.23	3.09	97.39
	MMoE*	0.69	1.17	0.04	2.09	2.11	2.44	97.82
	CPLE	0.70	1.00	0.04	2.15	1.71	2.54	97.93
Autumn	CNN-LSTM*	0.80	1.18	0.08	3.90	5.85	3.85	95.32
	MMoE*	0.76	0.51	0.06	3.29	2.48	3.00	97.08
	CPLE	0.75	0.51	0.06	3.27	2.47	2.97	97.10
Winter	CNN-LSTM*	0.51	0.53	0.22	2.22	3.86	6.38	96.28
	MMoE*	0.50	0.41	0.08	2.32	2.80	2.31	97.48
	CPLE	0.47	0.38	0.08	1.99	2.52	2.27	97.73

(B)

Model	RMSE (MW)			MAPE (%)			WMA (%)
	EL	CL	HL	EL	CL	HL	
CNN-LSTM*	0.77	0.96	0.12	3.07	3.73	4.78	96.31
MMoE*	0.69	0.69	0.06	2.61	2.39	2.56	97.47
CPLE	0.67	0.66	0.06	2.59	2.21	2.55	97.56



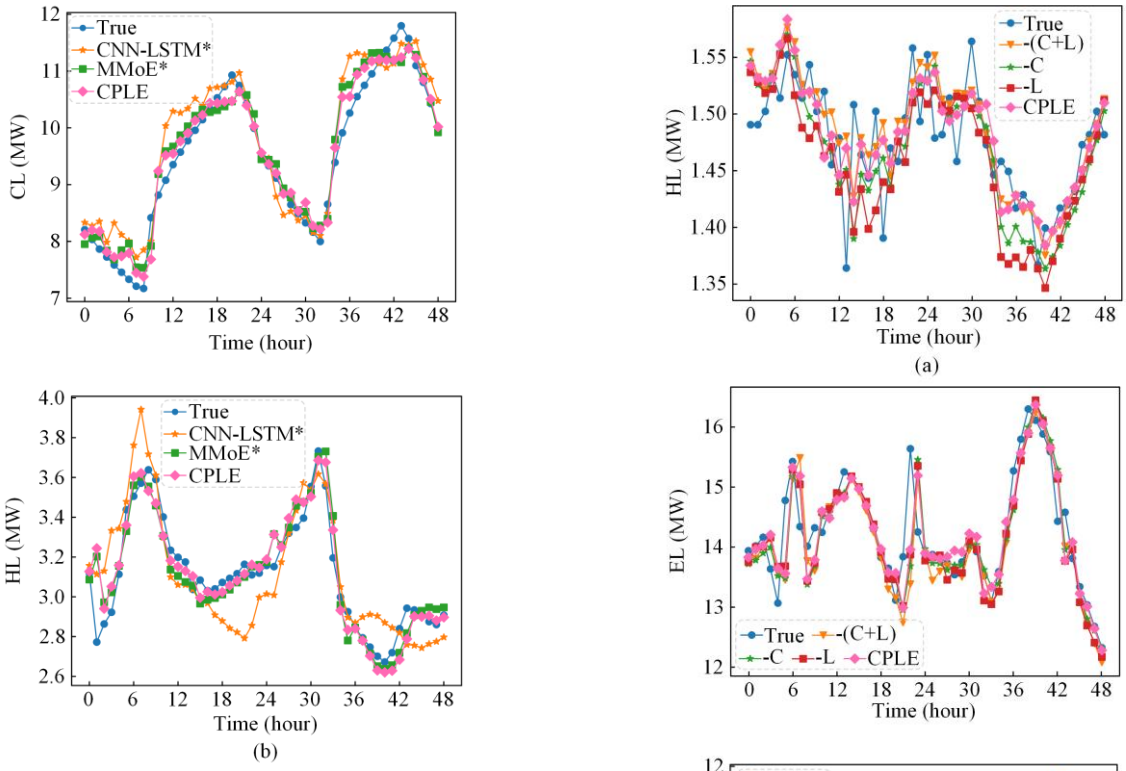


Fig. A2 Comparison of the EL, HL, and HL curves under different sharing methods. (a) Summer. (b) Winter.

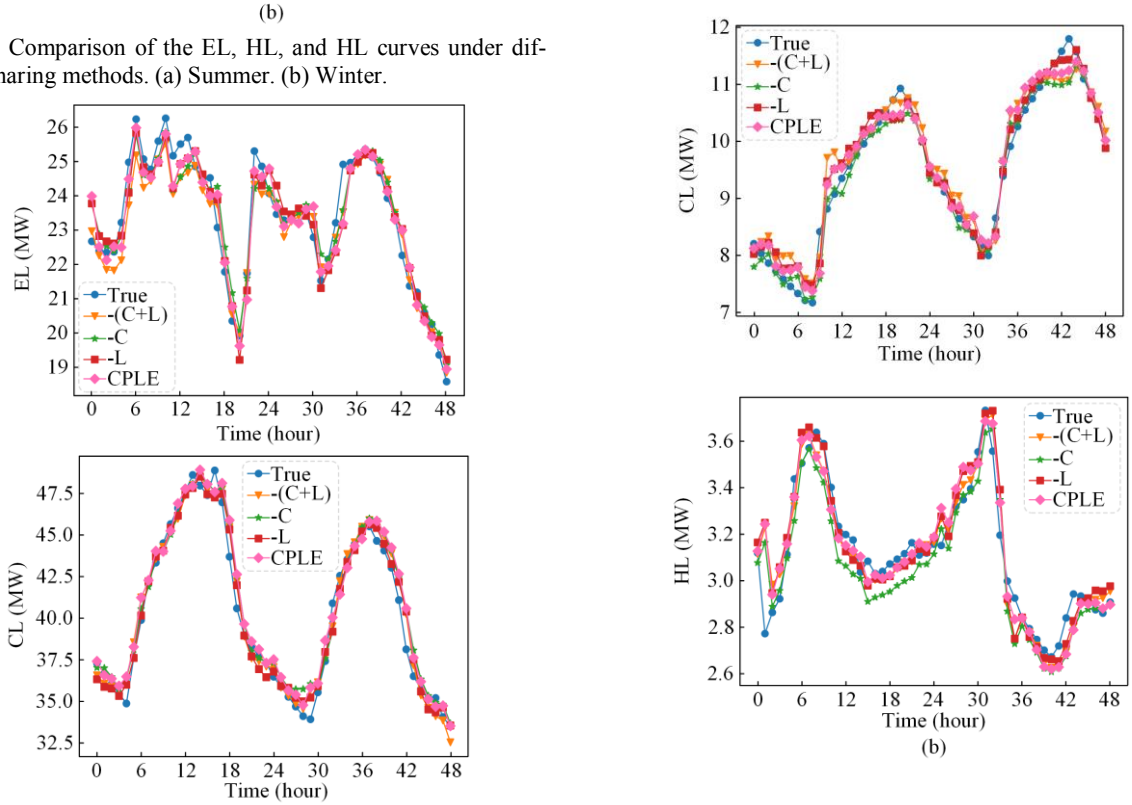


Fig. A3. Comparison of true and predicted values in ablation experiments. (a) Summer. (b) Winter.

TABLE A7
PARAMETER SETTINGS OF DIFFERENT LOAD DECOMPOSITION METHODS

Model name	Parameter	Value
EMD	Number of iterations	1000
	White noise standard deviation	0.2
CEEMD	Number of noise additions	100
	White noise standard deviation	0.2

TABLE A8
MAPE, RMSE AND WMA EVALUATION METRICS FOR DIFFERENT DECOMPOSITION METHODS

Season	Model	RMSE (MW)			MAPE (%)			WMA (%)
		EL	CL	HL	EL	CL	HL	
Spring	CPLE	0.75	0.75	0.05	2.96	2.14	2.42	97.47
	EMD-CPLE	0.64	0.63	0.05	2.49	2.02	2.25	97.74
	CEEMD-CPLE	0.61	0.78	0.05	2.47	2.46	2.24	97.57
	STL(+)-CPLE	0.40	0.31	0.04	1.74	0.95	1.72	98.57
	STL-CPLE	0.39	0.35	0.03	1.69	1.00	1.61	98.59
Summer	CPLE	0.70	1.00	0.04	2.15	1.71	2.54	97.93
	EMD-CPLE	0.67	0.64	0.04	1.92	1.15	2.43	98.27
	CEEMD-CPLE	0.73	1.00	0.04	2.10	1.69	2.43	97.99
	STL(+)-CPLE	0.51	0.67	0.02	1.62	1.17	1.60	98.55
	STL-CPLE	0.44	0.69	0.02	1.36	1.19	1.60	98.65
Autumn	CPLE	0.75	0.51	0.06	3.27	2.47	2.97	97.10
	EMD-CPLE	0.61	0.66	0.06	2.77	3.39	2.99	96.93
	CEEMD-CPLE	0.58	0.49	0.06	2.65	2.20	2.85	97.48
	STL(+)-CPLE	0.37	0.34	0.03	1.86	1.66	1.71	98.24
	STL-CPLE	0.38	0.40	0.03	1.83	1.97	1.58	98.16
Winter	CPLE	0.48	0.38	0.08	2.05	2.71	2.30	97.63
	EMD-CPLE	0.50	0.72	0.09	2.30	5.30	2.52	96.45
	CEEMD-CPLE	0.48	0.45	0.09	2.29	3.28	2.50	97.26
	STL(+)-CPLE	0.31	0.20	0.06	1.55	1.40	1.74	98.46
	STL-CPLE	0.30	0.20	0.06	1.51	1.35	1.73	98.50

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Suxun Zhu and Hengrui Ma: theoretical analysis and modeling of the process and performed simulation and experiment to verify the proposed method. Laijun Chen, Bo Wang, Hongxia Wang, Xiaozhu Li and Wenzhong Gao: offered help in theory and practice, read and put forward suggestions for the paper. All authors read and approved the final manuscript.

FUNDING

This work is supported by the National Natural Science Foundation of China Joint Fund Program (No. U22A20224).

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

DECLARATIONS

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

AUTHORS' INFORMATION

Suxun Zhu received the B.S. degree in Computer Science and Technology from Qinghai University, Xining, China, in 2021. She is currently pursuing the M.S. degree in School of Computer Technology, Qinghai University, Xining, China. Her research interests include load forecasting in integrated energy systems and power systems.

Hengrui Ma received the Ph.D. degree in power system and automation from the School of Electrical engineering, Wuhan University, Wuhan, China, in 2018. His research interests include integrated energy systems and energy storage systems.

Laijun Chen received the B.Eng. and Ph.D. degrees, both in electrical engineering, from Tsinghua University, Beijing, China in 2006 and 2011, respectively. He is currently an associate professor in the Department of Electrical Engineering, Tsinghua University. He is also the vice director of New Energy (Photovoltaic) Industry Research Center, Qinghai University. His research mainly concerns distribution generation, microgrids and energy storage.

Bo Wang received the Ph.D. degree in computer science from Wuhan University, Wuhan, China, in 2006. He did the Post-Doctoral Research with the School of Electrical Engineering, Wuhan University, from 2007 to 2009. He is currently a Professor with the School of Electrical Engineering, Wuhan University. His research interests include power system online assessment, big data, and integrated energy systems

Hongxia Wang received the B.S. degree in electrical engineering and automation from HeFei University of Technology, Hefei, China, in 2018, and the M.S. degree in School of Electrical Engineering and Automation, Wuhan University, Wuhan, China, in 2020. She is currently pursuing the Ph.D. degree in the School of Electrical Engineering and Automation, Wuhan University, Wuhan, China. Her research interests include big data and data fusion in power systems.

Xiaozhu Li received the Ph.D. degree in power system and automation from the School of Electrical Engineering, Xinjiang University, Xinjiang, China, in 2021. Her re-

search interests inclued planning and operation regulation of new power systems.

Wenzhong Gao received the M.S. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, USA, in 1999 and 2002, respectively. He is currently a professor with the Department of Electrical and Computer Engineering, University of Denver, Denver, USA. His current research interests include renewable energy and distributed generation.

REFERENCES

[1] Y. Wang, Y. Wang, and Y. Huang *et al*., "Planning and operation method of the regional integrated energy system considering economy and environment," *Energy*, vol. 171, pp. 731-750, Mar. 2019.

[2] G. Song, B. Cao, and L. Chang, "Review of grid-forming inverters in support of power system operation," *Chinese Journal of Electrical Engineering*, vol. 8, no. 1, pp. 1-15, Mar. 2022.

[3] B. Bai, J. Liu, and X. Wang *et al*., "Short-term multivariate load forecasting for urban energy based on minimum redundancy maximum relevance and dual-attention mechanism," *Automation of Electric Power Systems*, vol. 46, no. 17, pp. 44-55, Jun. 2022. (in Chinese)

[4] B. Ren, C. Huang, and L. Chen *et al*., "CLSTM-AR-based multi-dimensional feature fusion for multi-energy load forecasting," *Electronics*, vol. 11, no. 21, pp. 3481, Oct. 2022.

[5] C. M. Lee and C. N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902-5911, May 2011.

[6] J. F. Rendon-Sanchez and L. M. de Menezes, "Structural combination of seasonal exponential smoothing forecasts applied to load forecasting," *European Journal of Operational Research*, vol. 275, no. 3, pp. 916-924, Jun. 2019.

[7] N. Liu, Q. Tang, and J. Zhang *et al*., "A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids," *Applied Energy*, vol. 129, pp. 336-345, Sept. 2014.

[8] Y. Li, Z. Wen, and Y. Cao *et al*., "A combined forecasting approach with model self-adjustment for renewable generations and energy loads in smart community," *Energy*, vol. 129, pp. 216-227, Jun. 2017.

[9] H. Tang and L. Wu, "Short-term load forecasting based on improved stochastic forest," in *2020 International Conference on Virtual Reality and Intelligent Systems* (*ICVRIS*), Zhangjiajie, China, Jul. 2020, pp. 827-830.

[10] J. Li and L. Ma, "Short-term wind power prediction based on soft margin multiple kernel learning method," *Chinese Journal of Electrical Engineering*, vol. 8, no. 1, pp. 70-80, Mar. 2022.

[11] W. Kong, Z. Dong, and Y. Jia *et al*., "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.

[12] J. Lu, Q. Zhang, and Z. Yang *et al*., "A hybrid model based on convolutional neural network and long short-term memory for short-term load forecasting," in *2019 IEEE Power & Energy Society General Meeting* (*PESGM*), Atlanta, GA, USA, Aug. 2019, pp. 1-5.

[13] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180544-180557, Oct. 2020.

[14] M. H. D. M. Ribeiro, R. G. da Silva, and G. T. Ribeiro *et al*., "Cooperative ensemble learning model improves electric short-term load forecasting," *Chaos*, *Solitons & Fractals*, vol. 166, Jan. 2023.

[15] S. Zhou, Y. Li, and Y. Guo *et al*., "A load forecasting framework considering hybrid ensemble deep learning with two-stage load decomposition," *IEEE Transactions on Industry Applications*, vol. 60, no. 3, pp. 4568-4582, May-Jun. 2024.

[16] J. Wang, H. Zhong, and Z. Ma *et al*., "Review and prospect of integrated demand response in the multi-energy system," *Applied Energy*, vol. 202, pp. 772-782, Sept. 2017.

[17] A. T. Eseye and M. Lehtonen, "Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7743-7755, Feb. 2020.

[18] H. Li, L. Dou, and S. Li *et al*., "Abnormal state detection of OLTC based on improved fuzzy C-means clustering," *Chinese Journal of Electrical Engineering*, vol. 9, no. 1, pp. 129-141, Mar. 2023.

[19] J. Chen, Z. Hu, and W. Chen *et al*., "Load forecasting of integrated energy systems using combined DBiLSTM-MLR with secondary mode de-composition," *Automation of Electric Power Systems*, vol. 45, no. 13, pp. 85-94, Mar. 2021. (in Chinese)

[20] Z. Shi, Q. Ran, and F. Xu, "Short-term load forecasting based on aggregated second-order mode decomposition and Informer," *Power System Technology*, vol. 47, no. 10, pp. 1-15, Oct. 2023. (in Chinese)

[21] F. Jiang, Z. Lin, and W. Wang *et al*., "Optimal bagging ensemble ultra-short-term multi-load forecasting considering minimum average envelope entropy load decomposition," *Proceedings of the CSEE*, vol. 43, no. 5, pp. 1-17, Mar. 2023. (in Chinese)

[22] J. Ye, J. Cao, and L. Yang *et al*., "Ultra-short-term load forecasting of user-level integrated energy systems based on variational mode decomposition and multi-model fusion," *Power System Technology*, vol. 46, no. 2, pp. 2610-2622, Feb. 2022. (in Chinese)

[23] J. Shi and J. Teh, "Load forecasting for regional integrated energy system based on complementary ensemble empirical mode decomposition and multi-model fusion," *Applied Energy*, vol. 353, Jan, 2024.

[24] K. Wu, J. Wu, and L. Feng *et al*., "An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system," *International Transactions on Electrical Energy Systems*, vol. 31, no. 1, Sept. 2020.

[25] A. Wan, Q. Chang, and A. B. Khalil *et al*., "Short-term power load forecasting for combined heat and power

using CNN-LSTM enhanced by attention mechanism," *Energy*, vol. 282, Nov. 2023.

[26] W. Li, P. Zhang, and Q. Shi *et al*., "Load correction forecasting of integrated energy systems based on aggregated hybrid mode decomposition and temporal convolutional neural network," *Power System Technology*, vol. 46, no. 6, pp. 3345-3357, Jun. 2022. (in Chinese)

[27] B. Bai, J. Liu, and X. Wang *et al*., "Short-term prediction of urban energy multi-load based on MRMR and dual attention mechanism," *Automation of Electric Power Systems*, vol. 46, no. 11, pp. 44-55, Jun. 2022. (in Chinese)

[28] K. Li, Y. Mu, and F. Yang *et al*., "A novel short-term multi-energy load forecasting method for integrated energy system based on feature separation-fusion technology and improved CNN," *Applied Energy*, vol. 351, Dec. 2023.

[29] K. Wu, J. Gu, and L. Meng *et al*., "An explainable framework for load forecasting of a regional integrated energy system based on coupled features and multi-task learning," *Protection and Control of Modern Power Systems*, vol. 7, no. 1, pp. 1-14, Jan. 2022.

[30] Z. Lü, J. Gu, and L. Meng *et al*., "Short-term load forecasting of integrated energy systems based on coupled features and multi-task learning," *Automation of Electric Power Systems*, vol. 46, no. 3, pp. 58-66, Feb. 2022. (in Chinese)

[31] D. Niu, M. Yu, and L. Sun *et al*., "Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism," *Applied Energy*, vol. 313, May 2022.

[32] Y. Li and Z. Zhang, "Short-term load forecasting of integrated energy systems considering comprehensive demand response using Trans-GNN," *Transactions of China Electrotechnical Society*, vol. 38, no. 24, pp. 1-11, Dec. 2023. (in Chinese)

[33] Y. Li and Z. Zhang, "Short-term load forecasting of integrated energy systems based on GRU-TGTransformer," *Power System Protection and Control*, vol. 51, no. 15, pp. 33-41, Aug. 2023. (in Chinese)

[34] C. Wu, J. Yao, and G. Xue *et al*., "Load forecasting of integrated energy systems based on MMoE multi-task learning and long short-term memory network," *Electric Power Automation Equipment*, vol. 42, no. 5, pp. 33-39, May 2022. (in Chinese)

[35] Z. Yang, Q. Xu, and X. Cao *et al*., "Task-feature collaborative learning with application to personalized attribute prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4094-4110, Apr. 2020.

[36] H. Tang, J. Liu, and M. Zhao *et al*., "Progressive layered extraction (PLE): a novel multi-task learning (MTL) model for personalized recommendations," in *Proceedings of the 14th ACM Conference on Recommender Systems*, New York, NY, USA, Sept. 2020, pp. 269-278.

[37] R. B. Cleveland, W. S. Cleveland, and J. E. McRae *et al*., "STL: a seasonal-trend decomposition," *Journal of Offcial Statistics*, vol. 6, no. 1, pp. 3-73, Mar. 1990.

[38] K. Zhang, S. Cai, and T. Zhang *et al*., "Industry medium- to long-term load forecasting method considering multidimensional time-domain features," *Automation of Electric Power Systems*, vol. 47, no. 15, pp. 104-114, Aug. 2023. (in Chinese)

[39] A. Graves and A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*, pp. 37-45, Jan. 2012.

[40] J. Ma, Z. Zhao and X. Yi *et al*., "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA, Jul. 2018, pp. 1930-1939.

[41] B. S. Ardhan, M. Khedkar, and I. Srivastava *et al*., "A comparative analysis of hyperparameter tuned stochastic short term load forecasting for power system operator," *Energies*, vol. 16, no. 3, Jan. 2023.

[42] D. Kiruthiga and V. Manikandan, "Intraday time series load forecasting using Bayesian deep learning method—a new approach," *Electrical Engineering*, vol. 104, no. 3, pp. 1697-1709, Nov. 2021.

[43] Y. Hong, Y. Zhou, and Q. Li *et al*., "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785-55797, Mar. 2020.

[44] Campus Metabolism. Arizona State University's energy monitoring system. (2020, January 1 - 2021 November 31). Tempe Campus, Load data. [Online]. Available: http://cm.asu.edu/

[45] C. Wang, Y. Wang, and T. Zheng *et al*., "Comprehensive energy system multi-load prediction based on ResNet-LSTM network and attention mechanism," *Journal of Building Engineering*, vol. 07, pp. 1789-1799, Dec. 2021.

[46] WunderGround. Weather Report Data Platform. (2020, January 1 - 2021 November 31). Phoenix, AZ, Weather data. [Online]. Available: https://www.wunderground.com/weather/us/az/phoenix

[47] B. Lu, Z. Huo, and M. Yu, "Short-term multi-load forecasting of integrated energy systems based on LSTNet-Skip," *Proceedings of the CSEE*, vol. 42, no. 7, pp. 2273-2283, Apr. 2022. (in Chinese)

[48] W. Sun, S. Zhu, and J. Yang *et al*., "Topology identification of microgrid based on graph convolutional network," *Automation of Electric Power Systems*, vol. 45, no. 23, pp. 71-81, Dec. 2021. (in Chinese)

[49] S. Cui and X. Wang, "Multivariate load forecasting in integrated energy system based on maximal information coefficient and multi-objective Stacking ensemble learning," *Electric Power Automation Equipment*, vol. 42, no. 2, pp. 32-39, Feb. 2022. (in Chinese)