Short-term Load Forecasting with an Improved Dynamic Decomposition-Reconstruction-Ensemble Approach

Dongchuan Yang[a], Ju-e Guo[a], Yanzhao Li[a], Shaolong Sun[a*], Shouyang Wang[b,c]

[a]School of Management, Xi'an Jiaotong University, Xi'an, 710049, China

[b]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[c]Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China

*Corresponding author. School of Management, Xi'an Jiaotong University, Xi'an 710049, China

Tel.: +86 15911056725; fax: +86 29 82665049.

E-mail address: sunshaolong@xjtu.edu.cn (S. L. Sun).

Abstract: Short-term load forecasting has evolved into an important aspect of power system in safe operation and rational dispatching. However, given the load series' instability and volatility, this is a challenging task. To this end, this study proposes a dynamic decomposition-reconstruction-ensemble approach by cleverly and dynamically combining two proven and effective strategies (i.e., the reconstruction strategy and the secondary decomposition strategy). In fact, by introducing the decomposition-reconstruction process based on the dynamic classification, filtering, and giving the criteria for determining the components that need to be decomposed again, our proposed model improves the decomposition-ensemble forecasting framework. Our proposed model makes full use of decomposition techniques, complexity analysis, reconstruction strategies, secondary decomposition strategies, and a neural network optimized by an automatic hyperparameter optimization algorithm. Besides, we compared our proposed model with state-of-the-art models including five models with reconstruction strategy and two models with secondary decomposition strategy. The experiment results demonstrate the superiority of our proposed dynamic decomposition-reconstruction strategy in terms of forecasting accuracy, precise direction, equality, stability, correlation, comprehensive accuracy, and statistical tests. To conclude, our proposed model has the potential to be a useful tool for short-term load forecasting.

Keywords: Short-term load forecasting; Time series modeling; Dynamic decomposition-reconstruction strategy; Neural networks

1 Introduction

Maintenance scheduling, cogeneration, hydro-thermal apportionment, creating price plans, unit commitment, security assessment, and reserve extractions [1-5] have all become more difficult without accurate and comprehensive short-term load forecasting (STLF). Accurate STLF is not only necessary for the power grid's steady and safe functioning, but it also provides significant economic benefits to power corporations. Hobbs et al. [6] discovered that if the prediction error is reduced by 1%, a 10 GW power station may save $1.6 million per year. As a result, establishing an accurate and robust STLF model for a power system is both required and beneficial.

Generally, there are three categories of STLF models, namely, statistical methods, machine learning (ML) models, and hybrid models. As for statistical methods, exponential smoothing (ES), auto-regressive integrated moving average (ARIMA) [7], linear regression (LR) [8], and gray model (GM) [9] have been widely used for STLF. Although the statistical methods are mature technologies and convenient to use, yet they are typically requiring strict statistical hypotheses like stationarity and linearity. Furthermore, the mapping ability of the statistical methods to nonlinear is week. To overcome those limitations, ML models are more popular in recent years. Yang et al. [10] found that support vector machine and neural network are the most common methods in the literature on electric load forecasting during the last 20 years. In Australia, Wang et al. [11] confirmed the accuracy of the LSTM-based approach for forecasting short-term loads. Zhang et al. [12] combined SVM, cuckoo search algorithms, and singular spectrum analysis for STLF. Çevik and Çunkaş [13] adopted an adaptive neuro-fuzzy inference system (ANFIS) to forecast short-term load in Turkey.

However, single ML models still suffer from hyperparameter optimization, overfitting, and low convergence [14]. To reduce the negative and inherent effects of single ML models, multitudinous hybrid models are proposed. These hybrid models typically combine statistical methods, ML models, and decomposition algorithms. By combining the advantages of multiple techniques, hybrid methods achieve better forecasting accuracy, efficiency, and robustness [15]. Decomposition- ensemble hybrid approaches are frequently utilized for energy forecasting, owing to the strong performance of decomposition techniques on time series data processing[16]. Specifically, decomposing the time series into smoother, regular, clearly recognized subseries, then forecasting each decomposed component, and finally merging the forecasts is a common decomposition-ensemble approach [17]. Decomposition techniques are applied to simplify complex data and extract data characteristics in a decomposition-ensemble approach. Decomposition techniques mainly include empirical mode decomposition (EMD), variational mode decomposition (VMD), wavelet decomposition (WT), singular spectrum analysis (SSA), and other decomposition approaches. To explore the relationships between the load data in Australia, Fan et al. [18] applied EMD and deep belief network (DBN) for STLF, and optimized the parameters in DBN using a multi-objective evolutionary algorithm. Additionally, improved decomposition techniques based on EMD, such as ensemble empirical mode decomposition (EEMD) [19, 20], complementary ensemble empirical mode decomposition (CEEMD) [21, 22], complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [23, 24], and improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) [25], are also widely used in decomposition-ensemble approaches. Bento et al. [26] extracted data characteristics by correlation analysis and WT for STLF. He et al. [27] developed a hybrid approach with VMD and LSTM to forecast power load in Hubei Province, China, and the results revealed that VMD can successfully increase forecasting accuracy. Sulandari et al. [29] combined singular spectrum analysis (SSA),

linear recurrent formula, neural networks, and fuzzy systems to forecast load in Indonesian and this model can significantly reduce the forecasting error. Kong et al. [30] used dynamic mode decomposition (DMD) to identify the underlying spatio-temporal dynamics of error sequences, and the effectiveness of DMD for STLF was confirmed by a detailed experimental analysis.

Unfortunately, as the decomposition-ensemble approaches have become more widely used, they have exposed a slew of issues, such as the pattern aliasing caused by decomposition techniques [31], unstable signals in decomposed components [32], complexity adding and accumulation of estimation errors [16, 33], etc. Li et al. [34] pointed out that the single decomposition strategy cannot completely eliminate the randomness and irregularity in time series, and the generated partial components have dynamic complexity and irregular frequency range, which makes forecasting models difficult. Li et al. [32] found that some decomposed components are unstable, which adversely affects the forecasting accuracy. They argued that the single decomposition strategy is not able to adequately mined the information in the complex time series. Numerous researchers have found that the first intrinsic mode function (IMF_1) decomposed by EMD techniques is random and irregular, which makes forecasting more difficult [31, 35, 36].

In response to these problems, two strategies of improvement have emerged in the literature: (1) reconstruction (adding a components reconstruction process between data decomposition and separate forecasting), (2) secondary decomposition (partial components after the first decomposition is decomposed again). As for reconstruction, it is usually to reconstruct similar components by analyzing the characteristics of decomposed components, such as complexity, frequency, and mean. Yu et al. [16] noted that the complexity of the model and the accumulation of estimation errors can be effectively reduced by reconstructing similar components. Therefore, the decomposition-ensemble approaches with reconstruction have been successfully applied in various fields. For example, Liu et al. [36] reconstructed the low-frequency components decomposed by EEMD into a new component to avoid the pattern aliasing for STLF. Sun and Wang [37] successfully implemented the reconstruction based on sample entropy into a decomposition-ensemble approach for wind speed forecasting. Paramesh Kumar et al. [38] reconstructed components by the fine-to-coarse (FTC) for biofuel production forecasting. Wang et al. [39] applied the run-length-judgment (RLJ) approach to reconstruct components into four parts: trend sequences, low frequency, medium frequency, and high frequency. The experimental analysis indicated that the reconstruction strategy can obtain more accurate and stable oil price forecasting results. Yu et al. [16] introduced the data characteristic driven (DCD) which including multiple characteristics as the reconstruction approach to forecast oil price. As for secondary decomposition, there are usually two ways, one is to perform a secondary decomposition on the most complex and irregular component [35, 40, 41], and the other is to perform a secondary decomposition on multiple complex and irregular components [42-44]. Generally, the former way usually adopted EMD and its improved techniques for the first decomposition, and the latter way applied wavelet packet decomposition (WPD) for the first decomposition. **Table 1** lists the summery of some state-of-the-art decomposition-ensemble approaches with reconstruction, and **Table 2** lists the summery of some state-of-the-art decomposition-ensemble approaches with secondary decomposition.

Table 1 The summary of some state-of-the-art decomposition-ensemble approaches with reconstruction.

References	Decomposition	Reconstruction method	Focus	Research objects

	method			
[45]	EEMD	Fine-to-coarse	Mean	Oil price
Zhu et al. [46]	EEMD	Fine-to-coarse	Mean	Energy prices
Paramesh Kumar et al. [38]	EMD	Fine-to-coarse	Mean	Biofuel production
Wang et al. [39]	EMD	Run-length-judgment	Frequency	Oil price
Liu et al. [36]	EEMD	Hilbert-huang transform	Frequency	STLF
Sun and Wang [37]	FEEMD	Sample entropy	Complexity	Wind speed
Xie et al. [47]	CEEMDAN	Multiscale permutation entropy	Complexity	Tourism
Fu et al. [48]	TVF-EMD	Fuzzy entropy	Complexity	Wind speed
Ruiz-Aguilar et al. [49]	EMD	Permutation entropy	Complexity	Wind speed
Duan et al. [50]	VMD	Sample entropy	Complexity	Wind power
Sun et al. [51]	EEMD	K-means	Clustering	Solar radiation
Zhu et al. [52]	EEMD	Evolutionary clustering	Clustering	Oil price
Wang et al. [53]	BEMD	K-means	Clustering	$PM_{2.5}$
Yu et al. [16]	EEMD	Data-characteristic-driven	Hybrid	Oil price

Note: FEEMD (fast ensemble empirical model decomposition), TVF-EMD (time varying filter based empirical mode decomposition).

Table 2 The summary of some state-of-the-art decomposition-ensemble approaches with secondary decomposition.

References	Decomposition Method 1	Decomposition Method 2	Research objects
Yin et al. [35]	EMD	WPD	Wind power
Liu et al. [42]	WPD	FEEMD	Wind speed
Gan et al. [43]	WPD	CEEMD	$PM_{2.5}$
Li et al. [32]	CEEMD	WPD	Container throughput
Sun and Huang [31]	EMD	VMD	Carbon price
Xiang et al. [54]	SSA	VMD	Wind speed
Sun and Huang [55]	EMD	VMD	Carbon price
Li et al. [34]	CEEMD	VMD	Carbon price
Yang et al. [40]	ICEEMDAN	VMD	Electricity price

Sun et al. [44]	WT	VMD	Wind power

However, although both strategies listed above reduce the forecasting difficulty to some extent, there is still room for improvement. On the one hand, the reconstruction strategy is often based on a specific criterion, which makes it difficult to effectively capture the internal characteristics for reconstruction [16]. On the other hand, the components that need to be decomposed again are often determined by experience rather than a specific criterion, and components could still be unstable and high-complexity after secondary decomposition. Moreover, both strategies determine the modeling details of reconstruction or secondary decomposition on the basis of a specific problem but cannot be dynamically adjusted, which will lead to the poor migration of these strategies. Therefore, it is essential to propose a dynamic strategy that can adequately simplify complex data and adjust adaptively according to data characteristics.

To address the abovementioned limitations, this study combines both strategies and builds a dynamic decomposition-reconstruction strategy to overcome the shortcomings of previous decomposition and reconstruction methods. Our proposed approach includes two parts, i.e., dynamic decomposition-reconstruction and ensemble forecasting. Especially, dynamic decomposition-reconstruction is made up of data decomposition, characteristics analysis and performance evaluation of validation set, classification and reconstruction, and filter. Firstly, the data is decomposed using ICEEMDAN, and the decomposed components are characterized and reconstructed according to the reconstruction strategy to reduce the accumulation of estimation errors. Next, the components are decomposed again based on the secondary decomposition strategy to further simplify the data and reduce its complexity. Inspired by the embedded method in feature selection, this study determines whether the decomposed component needs to be further decomposed to reduce its complexity and irregularity according to its performance evaluation on the validation set. Meanwhile, to fulfill both high accuracy and stability, the BPNN optimized by an automatic Bayesian optimization algorithm (BOA) based on the tree-structured parzen estimator (TPE) is developed to forecast the components decomposed, severally. Finally, generating the forecasting values by ensemble the forecasting results of each component separately. The following are primary contributions of this study:

(1) This study proposes a dynamic decomposition-reconstruction strategy, which incorporates the reconstruction strategy and the secondary decomposition strategy into one framework. To our best knowledge, both reconstruction and secondary decomposition are incorporated into the decomposition-ensemble framework for the first time, filling the gap in this research area.
(2) By verifying whether the final component is more easily simulated by the model relative to the original data, this study gives the specific criteria for determining the components that need to be decomposed again.
(3) The dynamic decomposition-reconstruction framework is constructed by dynamic adjusting the reconstruction and decomposition modeling details on the basis of the data characteristics and the performance evaluation of the validation set, which improves the existing decomposition-ensemble forecasting framework.
(4) In order to verify the effectiveness of our proposed strategy, a thorough experimental analysis is conducted. We compare the state-of-the-art strategies with our proposed dynamic decomposition-reconstruction strategy, and the load from three regions are utilized to verify the advantages of our proposed strategy. Moreover, this study measures all models in terms of

forecasting accuracy, precise direction, equality, stability, correlation, comprehensive accuracy, and statistical tests.

The remainder of this study is as follows: **Section 2** presents the relevant methods and our proposed forecasting framework. The experiments and analysis are introduced in **Section 3**. **Section 4** provides conclusions and suggestions for future work.

2 The proposed dynamic decomposition-reconstruction-ensemble approach

2.1 Dynamic decomposition-reconstruction strategy

A dynamic decomposition-reconstruction model, which combines the advantages of reconstruction and secondary decomposition, is presented in this subsection in detail. In particular, the rules of the decomposition-reconstruction strategy are described firstly, and then the relevant methods including ICEEMDAN and permutation entropy (PE) are introduced. The forecasting model employed in the performance evaluation on the validation set will be described in **Section 2.2**.

2.1.1 Dynamic decomposition-reconstruction rules

To reduce the accumulation of estimation errors, and further simplify the complexity of data, this study incorporates the reconstruction strategy and the secondary decomposition strategy into a framework. In particular, it is an important issue that the criteria for determining whether a component needs to be further decomposed, which need to be addressed in our proposed framework. Inspired by the embedded method in feature selection, a method similar to feature selection is applied to dynamically determine the components to be decomposed again. The embedded method evaluates the performance of a subset of features according to a machine learning scoring function, where the model training and feature selection are performed simultaneously [56]. This process will be repeated until the stop condition is met. Appropriate feature selection is critical for improving forecasting performance [57]. Similarly, in our proposed decomposition-reconstruction strategy, model training is performed simultaneously with the selection of the components to be decomposed. The forecasting accuracy of the component in the training model is used to determine whether it needs to be further decomposed. Meanwhile, the selected components are decomposed to produce new components, which will face a new round of selection. The above process is continuously repeated until the stopping condition is reached. In this way, the original data is decomposed several times, which can fully explore its internal information and overcome the problem that the components could still be unstable and high-complexity after secondary decomposition. As for the part of the reconstruction, it's reasonable to reconstruct low-complexity components to reduce estimation errors accumulation and computational complexity. This is because high-complexity components are difficult to forecast [48], but low-complexity components are easy. Reconfiguration strategies by analyzing the complexity of the components have been reported in a wide variety of domains, such as wind speed forecasting [49], tourism demand forecasting [47], crude oil price forecasting [20], etc. Therefore, this study reconstructs the components according to the complexity analysis of components with permutation entropy (PE), which is powerful and successfully used in various researches. Specifically, if the PE value of a component is greater than a given threshold

θ, then the component appears as a high-complexity feature, and conversely, the component appears as a low-complexity feature. Furthermore, refer to [16], the given threshold θ is set to 0.5 in this study.

According to the above description, the dynamic decomposition-reconstruction strategy can be developed, and the process is shown in **Fig. 1.** To present this strategy more clearly, an example for the processes of a round of classification, reconstruction, and decomposition is shown in **Fig.2.** The red points represent the components decomposed from the original data, where the red points pointed by the arrows represent the new components reconstructed from the components, and the orange points represent the new components obtained from the components decomposed again. Specifically, the dynamic decomposition-reconstruction strategy includes the following four steps:

Step 1: Data decomposition

An advanced decomposition technique, ICEEMDAN, is applied to decompose the load time series.

Step 2: Characteristics analysis and performance evaluation of validation set

The decomposed components are deeply characterized to obtain the internal features with complexity testing, namely permutation entropy (PE), which is used to determine whether it needs to be further reconstructed. This study training the model with each component separately to obtain the forecasting accuracy of the corresponding component, which is used to determine whether it needs to be further decomposed. In particular, based on the naïve idea that the decomposed components are more easily simulated by the model relative to the original data. Hence, the forecasting accuracy of each component that needs to be decomposed again should be lower than the integrated forecasting accuracy of all components obtained from the first decomposition. It is seen as the threshold for determining whether the components need to be decomposed further. More specifically, the integrated forecasting accuracy of all components obtained from the first decomposition here refers to how well the original data is fitted with the sum of the forecasting results of all components obtained from the first decomposition and that all evaluations are performed on the validation set.

Step 3: Classification and reconstruction

According to the PE value and forecasting accuracy of each component obtained in step 2, the components are divided into four categories: high-complexity with high accuracy, high-complexity with low-accuracy, low-complexity with high accuracy, and low-complexity with low accuracy. Among them, components that appear low-complexity with high accuracy are reconstructed into a new component, and components that appear low-complexity with low accuracy are also reconstructed into another new component.

Step 4: Filter

Low-accuracy components, including high-complexity with low-accuracy components and a new component reconstructed from low-complexity with high-accuracy components, are "filtered" and wait to be decomposed again, while high-accuracy components, including high-complexity with high-accuracy components and a new component reconstructed from low-complexity with the high accuracy component, are “saved”. Each “filtered” component is treated as new original data and the above four steps are repeated. Since the noise in the original data is difficult to eliminate, there will still be components with low accuracy after multiple decomposition and reconstruction processes. To obtain the output of all components of our proposed strategy in a limited time, this study would stop the above loop when there were only a very few low accuracy components. In particular, when

the sum of the variances of the filtered components is less than 5% of the variance of the original data, the loop is stopped and all components are output.

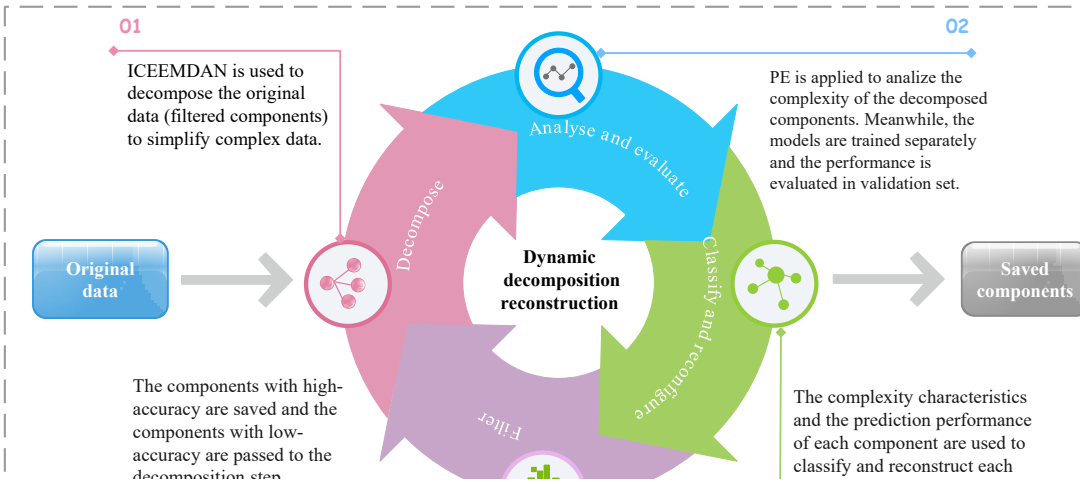

Fig.1. The dynamic decomposition-reconstruction strategy.

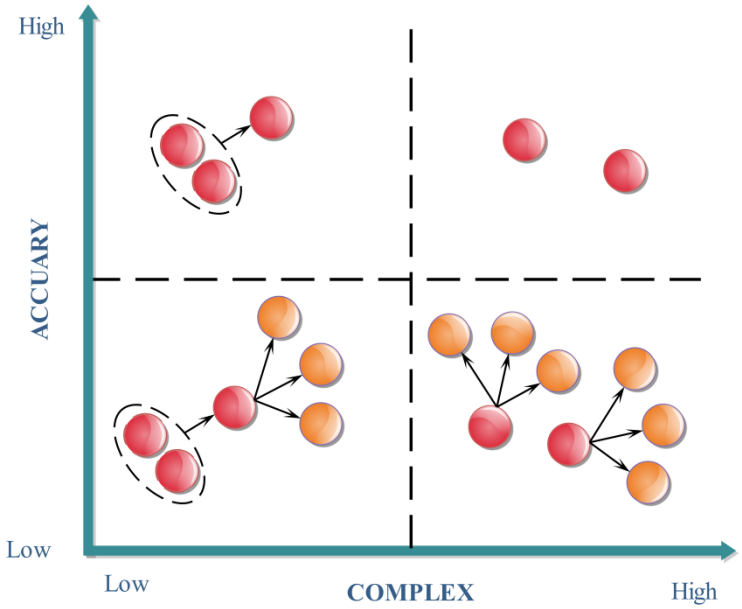

Fig.2. An example for the processes of classification, reconstruction, and decomposition.

2.1.2 Data decomposition

In this study, the ICEEMDAN is employed to decompose and capture the internal characteristics of the load series. ICEEMDAN, proposed by Colominas et al. [58], adds a special kind of white noise in the decomposition process, which is the IMF of Gaussian white noise after EMD decomposition. Each component calculates the local mean value of signal with noise and defines the decomposed IMF as the difference between local mean and residual signal. The ICEEMDAN technique significantly reduces the noise in IMF and improves the deficiency of traditional decomposition techniques in generating false component and mode aliasing. $E_k(\cdot)$ represents the k_{th} component after decomposition with EMD, $M(\cdot)$ represents the signal's

local mean, then $E_1(x) = x - M(x)$. $w^{(i)}$ represents a kind of white Gaussian noise, whose unit variance and mean are zero, and $x^{(i)} = x + w^{(i)}$. $\langle\ \rangle$ is the operator that produces the mean value. The specific decomposition process is as follows:

Step 1: The white noise $E_1(w^{(i)})$ is added to the original signal x, then $x^{(i)} = x + \beta_0 E_1(w^{(i)})$, and the first decomposition residual is $r_1 = \langle M(x^{(i)}) \rangle$.

Step 2: Calculate the first component when $k = 1$: $\tilde{d}_1 = x - r_1$.

Step 3: The second residual is estimated as the mean of $r_1 + \beta_1 E_2(w^{(i)})$. Accordingly, the second component is defined as： $\tilde{d}_2 = r_1 - r_2 = r_1 - \langle M(r_1 + \beta_1 E_2(w^{(i)})) \rangle$.

Step 4: For $k = 3,...,k$, calculates the k-th residual: $r_k = \langle M(r_{k-1} + \beta_{k-1} E_k(w^{(i)})) \rangle$.

Step 5: Compute the k-th component: $\tilde{d}_k = r_{k-1} - r_k$.

Step 6: Go back to step 4 for $k+1$.

2.1.3 Characteristic analysis with permutation entropy (PE)

A popular method for complexity testing is permutation entropy (PE), which has significant advantages in simplicity, efficiency, and robustness over other techniques (e.g. sample entropy and fuzzy entropy). Therefore, numerous researchers such as Yu et al. [16], Ruiz-Aguilar et al. [49] had applied PE to measure the complexity characteristics of the data and reconstructed the components accordingly. PE is a kind of information entropy calculation method proposed by Bandt and Pompe [59], which measures the complexity by mapping a time series to a symbolic series. The calculation process is as follows:

By applying phase space reconstruction to the load series, $\{x_t,\ t = 1,2,...,N\}$, the matrix X can be obtained.

$$X = \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(d-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(d-1)\tau \\ x(j) & x(j+\tau) & \cdots & x(j+(d-1)\tau \\ \vdots & \vdots & & \vdots \\ x(k) & x(k+\tau) & \cdots & x(k+(d-1)\tau \end{bmatrix},\ j = 1,2,...,k, \tag{2}$$

where τ represents a time-lag, d represents embedding dimension, k represents the number of subsequences reconstructed.

Next, X_j, which is the j-th row vector of the matrix X, is sorted in descending order, and

there will be a total of $d!$ possible ordinal patterns for an embedding dimension d. π_j is the ordinal patterns of X_j, and $f(\pi_j)$ is the frequency of ordinal patterns π_j occurrence in the time series. The relative frequency of π_j is:

$$p(\pi_j) = \frac{f(\pi_j)}{N-(d-1)\tau}, \tag{3}$$

and PE is represented as:

$$H_P = -\sum_{j=1}^{k} P(\pi_j) \ln\left(P(\pi_j)\right). \tag{4}$$

Lastly, the PE is normalized into:

$$H_P = \frac{H_P}{\ln(d!)}. \tag{5}$$

If the probabilities of all ordinal patterns are equal, the value of H_p is 1. Thus, the closer the value of H_p is to 1, the higher the complexity of the data.

2.2 Ensemble forecasting

2.2.1 Back propagation neural network (BPNN)

Appling error back propagation algorithm, the BPNN has a strong ability of nonlinear mapping and complex logic operation and becomes the most widely used neural network [60, 61]. The BPNN learns some hidden rules by its training without any prior given mapping relationship and can obtain the simulation output with the least deviation from the desired output value when the input values are given. During the learning process, the signal is forward transmitted, and when the desired output cannot be obtained, the error is back propagated, and the error is minimized by modifying the connection weights between each neuron until the desired output is obtained [62].

The structure of a BPNN consists of multiple layers, including one input layer, one or more hidden layers, and one output layer, with fully connected neurons between the layers. For a three-layer BPNN, there will be n input neurons, s hidden neurons, and m output neurons. Firstly, the input neurons receive the vector $x = (x_1, x_2, ..., x_n)$ from the database and pass them to all neurons in the hidden layer. Then, each hidden neuron first calculates the net input y_h and generates the output Y_h, which is calculated as follows:

$$y_h = \sum_{i=1}^{n} w_{ih} x_i, \tag{6}$$

$$Y_h = f\left(y_h\right) = f\left(\sum_{i=1}^{n} w_{ih} x_i\right), \tag{7}$$

where w_{ih} represents the weight between the i-th neuron in the input layer and the h-th neuron in the hidden layer, $f\left(\cdot\right)$ represents the activation function.

Finally, each output neuron j receives the outputs of the hidden neurons as its inputs and repeats the above operation until the stopping condition is met, the iterative process of output O_j is as follows:

$$O_j = f\left(\sum_{h=1}^{s} Y_h w_{hj}\right) = f\left[\sum_{h=1}^{s} f\left(\sum_{i=1}^{n} x_i w_{ih}\right) w_{hj}\right]. \tag{8}$$

It is worth mentioning that this study uses a BPNN containing two hidden layers as predictors for STLF, and it has one output neuron, multiple hidden neurons, and input neurons.

2.2.2 Hyperparameter optimization

Hyperparameters are parameters whose values are set before training the model, such as the learning rate, the number of hidden layer's neurons, etc., rather than the parameter obtained by training. In general, it is necessary to choose an optimal set of hyperparameters to improve the forecasting performance and effectiveness of the model. In order to objectively assess all strategies, the hyperparameters of all models are determined using an automated BOA called the tree-structured parzen estimator (TPE) [63]. Specifically, the hyperparameters of BPNN are as follows:

(1) The number of first (second) hidden layer's neurons.
(2) Dropout: Dropout refers to the process of training a neural network in which neurons are temporarily discarded from the network with a certain probability. It is a widely used regularization technique, which can effectively prevent overfitting and improve forecasting results.
(3) L2 regularization: It's also an effective regularization technique for preventing overfitting, with a cost function $\left(\lambda/2m\right)\|W\|^2$, where W represents network weights, λ represents the hyperparameter, and m represents the number of data in a batch.
(4) Activation function: It's the function responsible for mapping the inputs of the neurons to the outputs. For convenience, the same activation function is used between all layers.

2.3 Overview of our proposed approach

Since the original data could not completely simplify the original data and capture the intrinsic factors after primary decomposition, the reconstruction strategy and secondary decomposition strategy were proposed successively. However, both strategies determine the modeling details of reconstruction or secondary decomposition for specific problems and cannot be adjusted dynamically. Moreover, once reconstruction and decomposition process for the decomposed components improves the simple decomposition-ensemble approach, but there still has much room

for improvement. Therefore, the strategy proposed in this study combines the two strategies and dynamically applies the reconstruction and decomposition process to components multiple times for fully converting the original data into components that are easy to simulate by the model. Accordingly, the framework of the dynamic decomposition-reconstruction-ensemble approach is illustrated in **Fig. 3**.

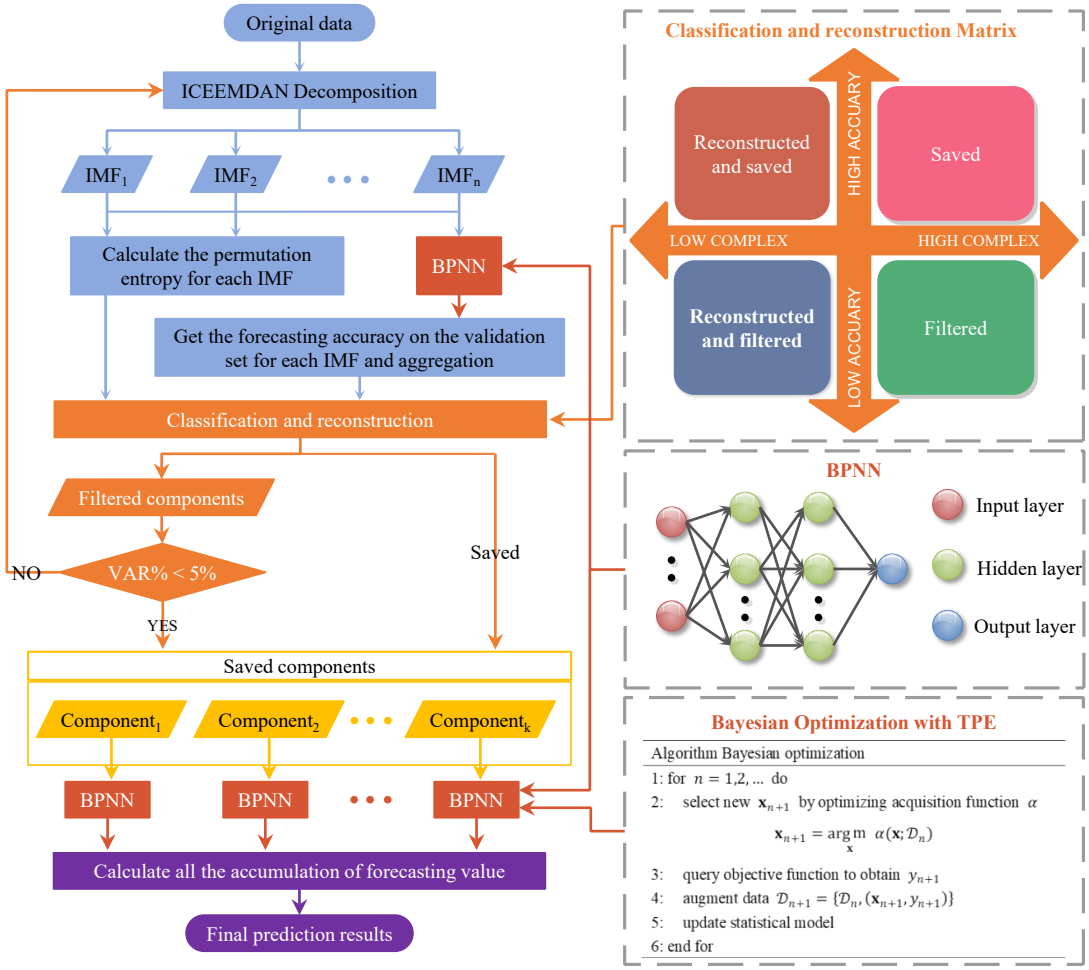

Fig. 3. The framework of our proposed dynamic decomposition-reconstruction-ensemble approach.

Step 1: Decompose the short-term load series into n IMFs adopting ICEEMADN.

Step 2: Calculate the FE value for each IMF, and train the BPNN model with each IMF separately to obtain its forecasting accuracy on the validation set.

Step 3: Classify and reconstruct IMFs according to the defined criteria.

Step 4: Filter components obtained in step 3 into 2 groups, i.e. filtered components and saved components. Each filtered component is treated as a new original data and repeats the above four steps until the stopping condition is reached. Then, the remaining filter components are exported to the group of saved components.

Step 5: Employ the BPNN to model each component, and apply the TPE algorithm to optimize its hyperparameters.

Step 6: Accumulate the forecasts of all components to obtain the final forecasting for the short-term load.

3 Experimental analysis

3.1 Experiment design

3.1.1 Data description

To verify the effectiveness of our proposed approach for load datasets with various time intervals from different regions, load data from Xi'an, China with an interval of 60 minutes, New South Wales, Australia with an interval of 30 minutes and Belgium with an interval of 15 min were collected in this study. The three cases collected in this study are named Xi'an, NSW, and BE. Specifically, the Xi'an covers periods from September 1 to December 31 in 2019, the NSW covers periods from June 1 to July 31 in 2019, and the BE covers periods from March 1 to March 31 in 2019. Additionally, the corresponding detailed information is shown in **Fig. 4** and **Table 3**. The load time series are non-linear and non-stationary, which may limit the performance of the predictor to some extent. Further, the original load is divided into three parts, i.e., the training set (60%), validation set (20%), and test set (20%). In particular, the training set is employed for model construction, the validation set is employed for components selection for further decomposition and hyperparameter selection, and the test set is employed for model validation.

Table 3 the descriptive statistical characteristics of the collected datasets (MW).

Dataset	Length	Max	Min	Mean	Range	Std.	Skewness	Kurtosis
Xi'an	2928	27576.98	12712.76	19227.72	14864.22	3290.631	0.54399	2.49956
NSW	2928	12112.84	6095.49	8597.57	6017.35	1176.176	0.23943	2.41598
BE	2976	12265.15	7398.15	9990.22	4867.00	1101.320	-0.21093	1.89823

Note: Std. refers to the standard deviation.

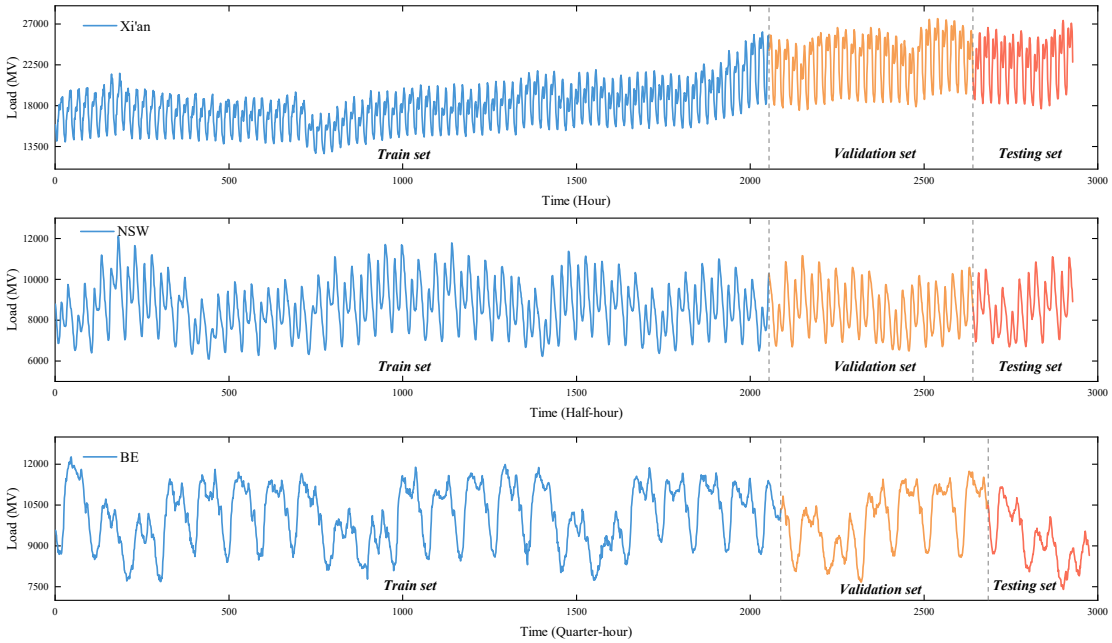

Fig.4. Short-term load datasets were collected from Xi'an, New South Wales, and Belgium.

3.1.2 Evaluation criteria

To validate the performance of our proposed strategy, various metrics such as average error (AE), fractional bias (FB), mean absolute error (MAE), square root of mean square error (RMSE), mean absolute percentage error (MAPE), Theil U statistic 1 (U_1), Theil U statistic 2 (U_2), directional accuracy (DS), and coefficient of determination (R^2) are adopted in this study. Their definition and equations are illustrated in **Table 4**.

Since each series has different dimensions, different methods are also measured differently with the same validity. Therefore, the forecasting validity degree (FVD) is used to evaluate the forecasting performance [61]. For a time series x and q forecasting method, x_{it} represents the forecasting value of i_{th} method at time t, where $i = 1,2,...,q$ and $t = 1,2,...,n$.

Definition 1. e_{it} represents the relative error of the i_{th} method at time t, and E represents the matrix of the relative error.

$$\mathbf{e}_{it} = \begin{cases} -1, \dfrac{x_t - x_{it}}{x_t} < -1 \\ \dfrac{x_t - x_{it}}{x_t}, -1 < \dfrac{x_t - x_{it}}{x_t} < 1. \\ 1, \dfrac{x_t - x_{it}}{x_t} > 1 \end{cases} \tag{9}$$

Definition 2. $FA_{it} = 1 - |e_{it}|$ represents the forecasting accuracy of the i_{th} method at time t. The FVD of i_{th} forecasting method is $FVD_{it} = E(FA_{it})[1 - \sigma(FA_{it})]$, where $E(.)$ represents the operator of mathematical expectation and $\sigma(.)$ represents the operator of standard deviation [64].

Table 4 Evaluation metrics.

Metric	Definition	Equation
AE	Average error	$AE = \frac{1}{N}\sum_{i=1}^{N}(F_i - A_i)$
FB	Fractional bias	$FB = 2 \times (\bar{\mathbf{A}} - \bar{\mathbf{F}}) / (\bar{\mathbf{A}} + \bar{\mathbf{F}})$
MAE	Mean absolute error	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert F_i - A_i \rvert$
RMSE	Square root of the mean square error	$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N}(\mathrm{F}_i - \mathrm{A}_i)^2}$
MAPE	Mean absolute percentage error	$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left\lvert \frac{\mathrm{A}_i - \mathrm{F}_i}{\mathrm{A}_i} \right\rvert \times 100\%$
U_1	Theil U statistic 1	$\mathrm{U1} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - F_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i)^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i)^2}}$

U_2	Theil U statistic 2	$\mathrm{U2}=\dfrac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{A_{i+1}-F_{i+1}}{A_i}\right)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{A_{i+1}-F_i}{A_i}\right)^2}}$
DS	Direction accuracy	$DS=\begin{cases}1, & if\ (\mathrm{A}_{i+1}-\mathrm{A}_i)(\mathrm{F}_{i+1}-\mathrm{A}_i)>0\\ 0, & \text{otherwise}\end{cases}$
R^2	Coefficient of determination	$R^2=1-\dfrac{\sum_{i=1}^{N}(\mathrm{F}_i-\mathrm{A}_i)}{\sum_{i=1}^{N}(\mathrm{F}_i-\bar{F})}$

Additionally, grey relational analysis (GRA) is used to calculate the grey relational degree (GRD) for assessing the similarity between the curves of the forecasting values and true values. The calculation process is as follows:

For the comparison series $X_i=(X_i(1),X_i(2),\ldots,X_i(n))$, and the reference series $X_0=(X_0(1),X_0(2),\ldots,X_0(n))$, the standard form is

$$x_i(t)=\frac{X_i(t)-\frac{1}{n}\sum_{t=1}^{n}X_i(t)}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(X_i(t)-\frac{1}{n}\sum_{t=1}^{n}X_i(t)\right)}}. \tag{10}$$

The correlation coefficient between x_0 and x_i is

$$\xi_i(k)=\frac{\min_i\min_k|x_0(k)-x_i(k)|+\rho\max_i\max_k|x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\rho\max_i\max_k|x_0(k)-x_i(k)|},\quad \rho\in(0,\infty). \tag{11}$$

The GRD between x_0 and x_i is

$$r_i=\frac{1}{n}\sum_{k=1}^{n}\xi_i(k). \tag{12}$$

Further, the Diebold-Mariano (DM) test is implemented to estimate the difference between the two forecasting models [65]. The null hypothesis of the DM test is that the forecasting accuracy of the target model and its benchmark model is equal. The DM test using the mean square error as the loss function can be expressed as follows:

$$S=\frac{\bar{g}}{\left(\hat{V}_{\bar{g}}/N\right)^{1/2}}, \tag{13}$$

where $\bar{g}=1/N\sum_{t=1}^{N}\left[\left(x_t-\hat{x}_{A,t}\right)^2-\left(x_t-\hat{x}_{B,t}\right)^2\right]$ and $\hat{V}_{\bar{g}}=\gamma_0+2\sum_{l=1}^{\infty}\gamma_l,\left(\gamma_l=\mathrm{cov}\left(g_t,g_{t-l}\right)\right)$. $\hat{x}_{A,t}$ represents the forecasting values of the target model A and $\hat{x}_{B,t}$ represents the benchmark model B.

3.1.3 Benchmark models

To verify the effectiveness of our proposed dynamic decomposition-reconstruction-ensemble approach, several other models are presented for comparison. Accordingly, single models, classical decomposition-ensemble approaches, decomposition-ensemble approaches with reconstruction strategy, decomposition-ensemble approaches with secondary decomposition strategy, and our proposed model will be comprehensively evaluated for comparison. To compare the accuracy and robustness of each model objectively, each model employs the BPNN as the predictor and follows the same automatic optimization algorithm to determine its hyperparameters. Therefore, for the single model, the BPNN is adopted as the single benchmark. For classical decomposition-ensemble approaches, the ICEEMDAN is applied to decompose data, and BPNN is adopted in individual forecasting. For decomposition-ensemble approaches with reconstruction strategy, and decomposition-ensemble approaches with secondary decomposition strategy, five different reconstruction strategies, and two secondary decomposition strategies are introduced in this study and listed in **Table 5**. In sum, there are 10 forecasting models, namely, the single model (BP), classical decomposition-ensemble approach (ICEEMDAN-BP), five various decomposition-ensemble approaches with reconstruction strategy (i.e., ICEEMDAN-DCD-BP, ICEEMDAN-SE-BP, ICEEMDAN-FTC-BP, ICEEMDAN-KM-BP, and ICEEMDAN-RLJ-BP), decomposition-ensemble approaches with secondary decomposition strategy (i.e., CEEMD-VMD-BP, and WPD-CEEMD-BP) and our proposed method.

Table 5 Different reconstruction and secondary decomposition methods.

Strategy	Method	Focus	References
Reconstruction	Fine-to-coarse (FTC)	Mean	Zhu et al. [46]
	Run-length-judgment (RLJ)	Frequency	Wang et al. [39]
	Sample entropy (SE)	Complexity	Sun and Wang [37]
	K-means (KM)	Clustering	Sun et al. [51]
	Data-characteristic-driven (DCD)	Hybrid	Yu et al. [16]
Secondary decomposition	WPD-CEEMD	-	Gan et al. [43]
	CEEMD-VMD	-	Li et al. [34]

3.2 Empirical results

3.2.1. Data decomposition and reconstruction

To verify the effectiveness of our strategy, as shown in **Fig. 5,** we compared the forecasting accuracy of the dynamic decomposition reconstruction strategy at different stages (i.e. single BPNN, once decomposition-reconstruction with BPNN, twice decomposition-reconstruction with BPNN, and thrice decomposition-reconstruction with BPNN, etc.). Specifically, the numbers on the horizontal axis represent the corresponding stages of our proposed strategy, the VAR% represents the sum of the variances of the filtered components as a percentage of the variance of the original data, and the MAE represents the mean absolute error of true values and the forecasting values accumulated from all component. Obviously, as the process of decomposition and reconstruction

increases, the variance share of the filtered components keeps decreasing and the forecasting accuracy on the validation set keeps improving. In other words, it reveals that our proposed strategy effectively improves the forecasting accuracy by reasonably increasing the decomposition and reconstruction process.

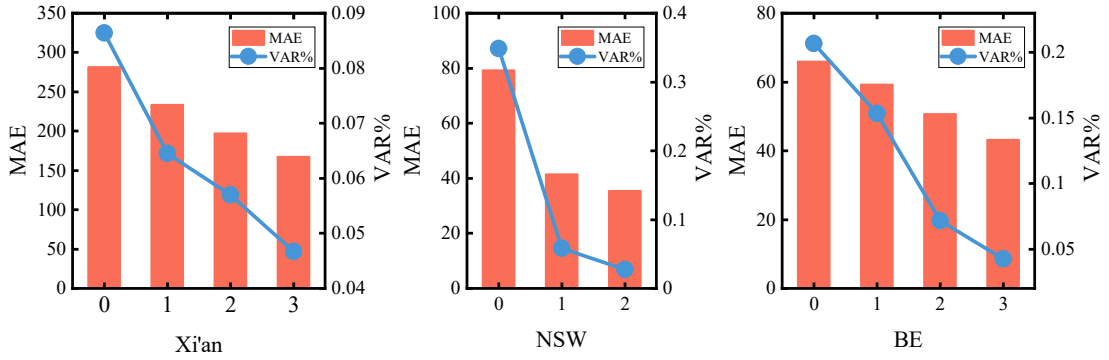

Fig. 5. Visualization of the variance share of the filtered components and the evaluated MAE at different stages for three datasets.

3.2.2. Forecasting performance comparisons

Three short-term load datasets are applied to validate the superiority of our proposed approach. Accordingly, the multi-step ahead forecasting performance (i.e., one-step, three-step, and five-step ahead) are shown in **Table 6**. Moreover, the performance evaluation results in **Table 6** are visualized in **Fig.6** in the form of radar charts, where the performances for three forecasting horizons are presented in a subplot. After a thorough analysis of **Table 6** and **Fig. 6**, the conclusions can be drawn as follows:

(1) First, compared with the single model, classical decomposition-ensemble approaches, decomposition-ensemble approaches with reconstruction strategy, and decomposition-ensemble approaches with secondary decomposition strategy, our proposed model in this study performs the best performance, with the three metrics (i.e., MAE, RMSE, and MAPE), in all forecasting horizons and datasets. Moreover, as can be seen from the case of Xi'an in **Fig. 6**, the variation of each metric of our proposed approach is smaller as the forecasting horizons increase. This may be attributed to the fact that our proposed approach combines the advantages of the other two strategies and dynamically applies the decomposition and reconstruction process several times to fully simplify the complex data into components that can be easily simulated by the model, which further reduces the nonlinearity and complexity of the components and thus enhances the forecasting accuracy effectively. This result also shows that our proposed dynamic decomposition-reconstruction-ensemble approach improves the classical decomposition-ensemble approach significantly and can be a promising alternative for short-term load forecasting.

(2) Second, when comparing the classical decomposition-ensemble approach with the

corresponding single model, the ICEEMDAN-BP performs significantly better than the single model in all forecasting horizons and datasets. The main reason is that the decomposition technique can simplify the complex load data and explore the hidden factors, thus further improving the forecasting accuracy. Furthermore, this result also confirms the effectiveness of the ICEEMDAN technique for load data decomposition and decomposition-ensemble approach.

(3) Third, compared with the classical decomposition-ensemble approach, the decomposition-ensemble approaches with reconstruction strategy and secondary decomposition strategy do not have better performance in all cases. One possible reason is that both strategies determine the modeling details of reconstruction or secondary decomposition on the basis of a specific problem and cannot be dynamically adjusted, which will lead to poor migration of these strategies, making both strategies provide poor results in some cases.

Table 6. Performance of load forecasting models for all datasets.

Cases	Horizon	One-step			Three-step			Five-step		
	Models	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Xi'an	BP	384.074	512.003	1.679	464.178	559.831	2.115	525.758	660.456	2.351
	ICEEMDAN-BP	267.539	316.656	1.211	275.569	337.204	1.245	315.011	376.605	1.421
	ICEEMDAN-DCD-BP	200.685	255.930	0.892	231.956	315.922	0.999	278.857	332.103	1.263
	ICEEMDAN-SE-BP	201.181	260.044	0.882	288.846	372.308	1.270	330.347	428.708	1.455
	ICEEMDAN-FTC-BP	329.383	401.886	1.495	385.121	504.234	1.679	396.928	527.326	1.745
	ICEEMDAN-KM-BP	290.580	348.213	1.299	376.984	465.109	1.649	392.883	482.109	1.772
	ICEEMDAN-RLJ-BP	271.471	345.627	1.179	275.176	365.685	1.198	325.808	426.467	1.431
	CEEMD-VMD-BP	205.816	260.445	0.890	232.964	290.773	1.041	284.030	357.435	1.254
	WPD-CEEMD-BP	283.581	341.270	1.264	317.480	398.227	1.412	440.432	544.395	1.953
	Our proposed	**175.595**	**223.525**	**0.780**	**180.928**	**218.366**	**0.827**	**216.512**	**284.170**	**0.970**
NSW	BP	75.055	97.918	0.877	163.840	225.962	1.888	259.797	343.881	3.065
	ICEEMDAN-BP	63.506	79.841	0.728	124.157	157.181	1.449	159.421	198.148	1.920
	ICEEMDAN-DCD-BP	50.486	63.922	0.582	124.756	155.304	1.469	144.741	175.930	1.707
	ICEEMDAN-SE-BP	51.395	64.320	0.608	117.709	149.919	1.392	151.445	190.467	1.777
	ICEEMDAN-FTC-BP	70.251	93.116	0.815	114.910	156.262	1.334	162.388	211.533	1.864
	ICEEMDAN-KM-BP	60.587	79.966	0.703	134.244	165.279	1.586	215.545	263.340	2.570
	ICEEMDAN-RLJ-BP	57.115	72.333	0.657	170.878	218.606	2.018	228.108	293.161	2.644
	CEEMD-VMD-BP	72.107	132.062	0.820	121.709	172.360	1.414	141.507	194.319	1.643
	WPD-CEEMD-BP	68.673	86.993	0.828	125.965	172.009	1.489	195.696	272.795	2.302
	Our proposed	**44.437**	**59.343**	**0.530**	**66.736**	**83.024**	**0.774**	**81.531**	**96.795**	**0.972**
BE	BP	130.115	160.014	1.411	191.658	243.405	2.073	253.102	318.114	2.813
	ICEEMDAN-BP	97.539	124.616	1.025	122.552	137.033	1.381	212.443	238.877	2.366
	ICEEMDAN-DCD-BP	62.350	76.916	0.664	105.587	126.636	1.166	160.916	203.744	1.729
	ICEEMDAN-SE-BP	82.322	104.051	0.878	111.408	138.654	1.241	154.630	205.340	1.690
	ICEEMDAN-FTC-BP	94.423	119.856	1.029	151.207	195.855	1.656	201.917	266.978	2.192
	ICEEMDAN-KM-BP	96.214	120.765	1.075	187.999	235.484	2.074	255.801	322.528	2.792

ICEEMDAN-RLJ-BP	114.828	140.758	1.215	168.629	202.962	1.830	211.633	258.052	2.357
CEEMD-VMD-BP	67.890	83.208	0.732	90.910	111.194	1.001	125.268	153.939	1.345
WPD-CEEMD-BP	64.021	78.936	0.687	143.884	180.315	1.534	214.103	264.177	2.369
Our proposed	**35.237**	**43.047**	**0.389**	**66.576**	**84.651**	**0.747**	**100.031**	**128.889**	**1.110**

Note: The numbers in bold indicate the best results.

To provide a comprehensive experimental analysis for our proposed model, a widely used statistical model (i.e. ARIMA) was also selected as the benchmark model, while six additional evaluation metrics (i.e. AE, FB, U_1, U_2, DS, and R^2) are employed to evaluate the forecasting performance. Accordingly, the one-step-ahead forecasting performance is illustrated in **Table 7**. We can still draw similar conclusions from **Table 7**. First, our proposed model almost has the best forecasting performance in all metrics, which means that our proposed method can forecast load time series effectively in forecasting accuracy (measured with AE and R^2), precise direction (measured with DS and FB), and equality (measured with U_1 and U_2). And these results prove the validity of our proposed model and the potential to be a promising forecasting tool for STLF. Second, the ICEEMDAN-BP model is generally superior to the BP model in terms of U_1, U_2, DS, and R2, due to the effectiveness of the decomposition technique. Finally, although the ICEEMDAN-DCD-BP model is generally better than the ICEEMDAN-BP model, not all decomposition-ensemble approaches based on reconstruction or secondary decomposition strategies are superior to ICEEMDAN-BP, for example, ICEEMDAN-FTC-BP does not have any advantage over ICEEMDAN-BP in most metrics.

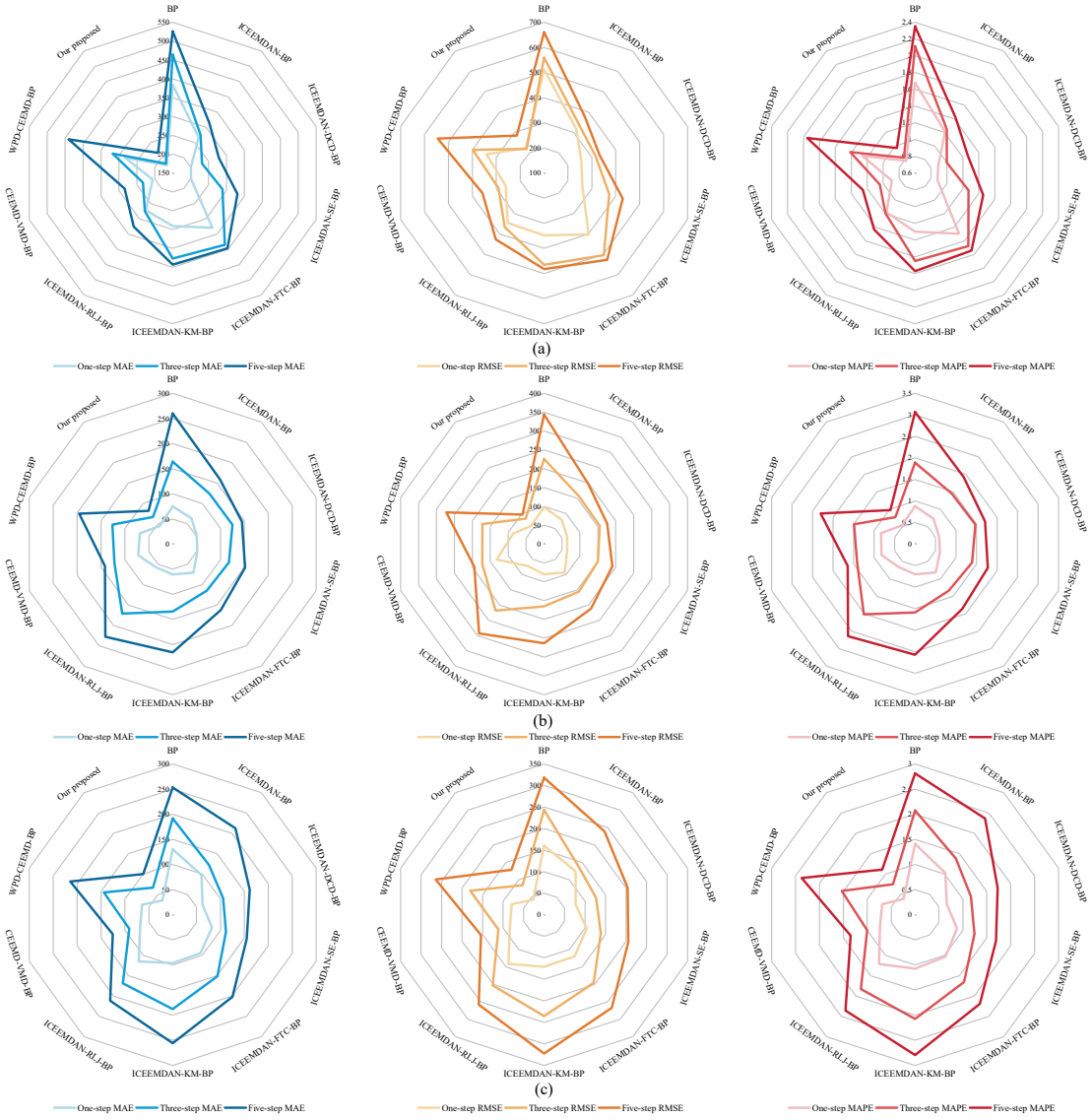

Fig. 6 Radar chart for visualized observation of the metrics and the variation tendency in various forecasting horizons obtained by all models: (a) Xi'an, (b) NSW, and (c) BE.

Table 7. Further analysis of all forecasting models with one-step ahead in all datasets.

Cases	Models	AE	FB	U_1	U_2	DS	R^2
Xi'an	Arima	281.2520	0.0125	0.0100	0.4362	0.9094	0.9713
	BP	173.2240	0.0077	0.0113	0.4803	0.8780	0.9633
	ICEEMDAN-BP	-203.9550	-0.0090	0.0069	0.3035	0.8920	0.9860
	ICEEMDAN-DCD-BP	-49.2236	-0.0022	0.0056	0.2372	**0.9373**	0.9908
	ICEEMDAN-SE-BP	-23.4191	-0.0010	0.0057	0.2383	0.9233	0.9905
	ICEEMDAN-FTC-BP	-78.9521	-0.0035	0.0088	0.3890	0.8606	0.9774
	ICEEMDAN-KM-BP	-41.5521	-0.0018	0.0077	0.3307	0.8850	0.9830
	ICEEMDAN-RLJ-BP	-181.5470	-0.0080	0.0076	0.3136	0.8955	0.9833
	CEEMD-VMD-BP	147.2890	0.0065	0.0057	0.2373	0.9233	0.9905
	WPD-CEEMD-BP	129.4300	0.0058	0.0075	0.3209	0.8676	0.9837
	Our proposed	**20.1283**	**0.0009**	**0.0049**	**0.2083**	0.9233	**0.9930**

	Arima	3.1212	0.0004	0.0049	0.3047	0.9059	0.9942
	BP	6.5693	0.0008	0.0057	0.3477	0.9338	0.9924
	ICEEMDAN-BP	-48.7930	-0.0057	0.0046	0.2695	0.9199	0.9950
	ICEEMDAN-DCD-BP	2.9357	0.0003	0.0037	0.2209	0.9338	0.9968
	ICEEMDAN-SE-BP	-19.8599	-0.0023	0.0037	0.2298	0.9408	0.9967
NSW	ICEEMDAN-FTC-BP	44.9459	0.0052	0.0054	0.3270	0.9338	0.9932
	ICEEMDAN-KM-BP	**2.0196**	**0.0002**	0.0046	0.2801	0.9547	0.9950
	ICEEMDAN-RLJ-BP	30.0563	0.0035	0.0042	0.2504	0.9373	0.9959
	CEEMD-VMD-BP	-24.0027	-0.0028	0.0076	0.4212	0.9268	0.9862
	WPD-CEEMD-BP	-44.6603	-0.0052	0.0050	0.3223	0.9268	0.9940
	Our proposed	2.7712	0.0003	**0.0034**	**0.2193**	**0.9652**	**0.9972**
	Arima	-7.7665	-0.0008	0.0088	1.4430	0.4555	0.9700
	BP	58.0855	0.0063	0.0087	1.4126	0.5753	0.9707
	ICEEMDAN-BP	-84.8403	-0.0092	0.0067	1.0258	0.7021	0.9822
	ICEEMDAN-DCD-BP	34.2277	0.0037	0.0042	0.6590	0.8082	0.9932
	ICEEMDAN-SE-BP	-63.0196	-0.0068	0.0056	0.8748	0.7329	0.9876
BE	ICEEMDAN-FTC-BP	38.8811	0.0042	0.0065	1.0654	0.6164	0.9836
	ICEEMDAN-KM-BP	-60.6486	-0.0066	0.0065	1.1175	0.6370	0.9833
	ICEEMDAN-RLJ-BP	102.9360	0.0113	0.0077	1.1911	0.7295	0.9773
	CEEMD-VMD-BP	32.9355	0.0036	0.0045	0.7291	0.8048	0.9921
	WPD-CEEMD-BP	29.7286	0.0032	0.0043	0.6823	0.7911	0.9929
	Our proposed	**-7.2195**	**-0.0008**	**0.0023**	**0.3893**	**0.8630**	**0.9979**

Note: The numbers in bold indicate the best results.

3.2.3. Further assessment with stability and grey relational analysis

In addition to forecasting error, forecasting stability is also an essential indicator for measuring the performance of models in this study. It is commonly assumed that a smaller variance means higher stability. From **Table 8** we can see that our proposed model has the smallest variance among all models. The variance of our proposed model on the Xi'an, NSW, and BE datasets are 0.0098, 0.0051, and 0.0022, respectively. Our proposed model improves both forecasting accuracy and robustness. The ICEEMDAN-BP model has a smaller variance than the BP model in all datasets, which also implies the effectiveness of the decomposition technique in improving the robustness. And compared to the classical decomposition-ensemble approach, not all decomposition-ensemble approaches with reconstruction or secondary decomposition strategies have smaller variance.

To further evaluate the validity of our proposed model, the grey relational degree was also employed in this study. The GRD is calculated from the GRA and employed to measure the correlation between the forecasts and observations. As is known to all, a higher GRD value means better forecasting. **Table 9** shows the results of GRD, from which we can find that the conclusions deduced above can be verified again here.

Table 8. The variance of the forecasting errors (%).

Models	Xi'an	NSW	BE
Arima	0.0276	0.0101	0.0314

BP	0.0441	0.0128	0.0269
ICEEMDAN-BP	0.0116	0.0052	0.0084
ICEEMDAN-DCD-BP	0.0120	0.0052	0.0054
ICEEMDAN-SE-BP	0.0124	0.0051	0.0073
ICEEMDAN-FTC-BP	0.0318	0.0085	0.0154
ICEEMDAN-KM-BP	0.0235	0.0084	0.0142
ICEEMDAN-RLJ-BP	0.0156	0.0055	0.0098
CEEMD-VMD-BP	0.0083	0.0194	0.0069
WPD-CEEMD-BP	0.0202	0.0084	0.0061
Our proposed	**0.0098**	**0.0051**	**0.0022**

Note: The numbers in bold indicate the best results.

Table 9. The results of the grey relational degree.

Horizon		One-step			Three-step			Five-step	
Models	Xi'an	NSW	BE	Xi'an	NSW	BE	Xi'an	NSW	BE
BP	0.7474	0.9199	0.8268	0.7251	0.8108	0.7548	0.7249	0.8017	0.6649
ICEEMDAN-BP	0.8519	0.9451	0.9040	0.8212	0.8329	0.9032	0.7954	0.8129	0.8252
ICEEMDAN-DCD-BP	0.8220	0.9402	0.9170	0.8173	0.8389	0.8437	**0.8244**	0.7884	0.7756
ICEEMDAN-SE-BP	0.8364	0.9455	0.8964	0.7463	0.8563	0.7999	0.6945	0.7979	0.7755
ICEEMDAN-FTC-BP	0.7656	0.9281	0.8611	0.7290	0.8516	0.7771	0.7125	0.7835	0.7238
ICEEMDAN-KM-BP	0.7940	0.9306	0.8668	0.7362	0.8382	0.7327	0.7127	0.7166	0.7216
ICEEMDAN-RLJ-BP	0.8182	0.9399	0.8849	0.7539	0.8147	0.7547	0.7389	0.7826	0.6980
CEEMD-VMD-BP	0.8640	0.9200	0.9065	0.8052	0.8753	0.8548	0.8022	0.8538	0.8205
WPD-CEEMD-BP	0.7906	0.9349	0.9075	0.7421	0.8552	0.8194	0.6913	0.7950	0.7350
Our proposed	**0.8656**	**0.9467**	**0.9406**	**0.8804**	**0.9402**	**0.8685**	0.8118	**0.9222**	**0.8324**

Note: The numbers in bold indicate the best results.

3.2.4. Further assessment with DM test and forecasting validity degrees

In this section, DM-test and forecasting validity degrees are adopted to further evaluate the performance of all models and their results are illustrated in **Table 10**. As can be seen from **Table 10**, our proposed model is significantly superior to all benchmark models. For the case BE, all p-values are well below the 1% significance level under different horizons, which indicates that our proposed dynamic decomposition-reconstruction-ensemble approach significantly outperforms all other benchmark models at the 99% confidence level from the perspective of statistical validation. For the case Xi'an, except for ICEEMDAN-SE-BP (95% significant) and CEEMD-VMD-BP (95% significant) at the horizons of one-step, benchmark models have 99% significant differences from our proposed model. Similarly, for the case NSW, except for ICEEMDAN-DCD-BP (95% significant) and ICEEMDAN-SE-BP (95% significant) at the horizons of one-step, benchmark models have 99% significant differences from our proposed model too. Furthermore, the FVD also verifies the excellent forecasting performance of our model. The FVD is a metric to evaluate the comprehensive accuracy of models in this study. In general, the larger the value of the FVD, the higher the comprehensive forecasting accuracy of the model. In other words, the model can achieve high forecasting validity only when it achieves excellent forecasting accuracy in all periods. As

shown in **Table 10**, our proposed model achieves the largest FVD values in all cases. The above results show that our proposed model has the smallest forecasting error and significantly improves the comprehensive forecasting accuracy and stability over all benchmark models.

Table 10. Comparison results of the DM test and forecasting validity degrees.

Cases	Horizon	One-step			Three-step			Five-step		
	Models	DM	p-value	FVD	DM	p-value	FVD	DM	p-value	FVD
Xi'an	BP	-10.196*	0.000	0.969	-14.803*	0.000	0.964	-14.318*	0.000	0.959
	ICEEMDAN-BP	-7.959*	0.000	0.980	-7.970*	0.000	0.978	-7.084*	0.000	0.976
	ICEEMDAN-DCD-BP	-2.679*	0.008	0.984	-3.478*	0.001	0.982	-5.124*	0.000	0.979
	ICEEMDAN-SE-BP	-2.546**	0.011	0.984	-7.067*	0.000	0.977	-7.514*	0.000	0.974
	ICEEMDAN-FTC-BP	-10.859*	0.000	0.974	-10.188*	0.000	0.970	-10.200*	0.000	0.968
	ICEEMDAN-KM-BP	-9.529*	0.000	0.978	-11.826*	0.000	0.972	-11.631*	0.000	0.970
	ICEEMDAN-RLJ-BP	-6.784*	0.000	0.979	-6.096*	0.000	0.978	-7.658*	0.000	0.974
	CEEMD-VMD-BP	-2.432**	0.016	0.985	-5.238*	0.000	0.982	-4.028*	0.000	0.978
	WPD-CEEMD-BP	-8.153*	0.000	0.979	-8.999*	0.000	0.975	-12.418*	0.000	0.967
	Our proposed	-	-	**0.986**	-	-	**0.986**	-	-	**0.982**
NSW	BP	-7.261*	0.000	0.984	-10.855*	0.000	0.964	-13.262*	0.000	0.944
	ICEEMDAN-BP	-5.750*	0.000	0.988	-9.241*	0.000	0.975	-11.298*	0.000	0.966
	ICEEMDAN-DCD-BP	-2.153**	0.032	0.990	-10.351*	0.000	0.975	-9.812*	0.000	0.971
	ICEEMDAN-SE-BP	-2.386**	0.018	0.989	-8.126*	0.000	0.975	-9.589*	0.000	0.969
	ICEEMDAN-FTC-BP	-6.745*	0.000	0.985	-7.298*	0.000	0.975	-9.412*	0.000	0.967
	ICEEMDAN-KM-BP	-4.693*	0.000	0.987	-10.949*	0.000	0.973	-14.022*	0.000	0.956
	ICEEMDAN-RLJ-BP	-3.923*	0.000	0.989	-13.220*	0.000	0.964	-13.049*	0.000	0.954
	CEEMD-VMD-BP	-4.138*	0.000	0.980	-6.891*	0.000	0.973	-7.002*	0.000	0.969
	WPD-CEEMD-BP	-9.008*	0.000	0.985	-8.477*	0.000	0.971	-10.397*	0.000	0.955
	Our proposed	-	-	**0.990**	-	-	**0.987**	-	-	**0.984**
BE	BP	-16.866*	0.000	0.976	-13.937*	0.000	0.964	-13.788*	0.000	0.950
	ICEEMDAN-BP	-12.515*	0.000	0.982	-12.454*	0.000	0.979	-15.679*	0.000	0.963
	ICEEMDAN-DCD-BP	-9.384*	0.000	0.989	-8.673*	0.000	0.981	-7.522*	0.000	0.970
	ICEEMDAN-SE-BP	-11.618*	0.000	0.985	-11.965*	0.000	0.978	-8.218*	0.000	0.969
	ICEEMDAN-FTC-BP	-13.303*	0.000	0.982	-11.659*	0.000	0.970	-10.150*	0.000	0.960
	ICEEMDAN-KM-BP	-14.452*	0.000	0.981	-14.537*	0.000	0.963	-14.211*	0.000	0.951
	ICEEMDAN-RLJ-BP	-15.845*	0.000	0.980	-15.232*	0.000	0.970	-12.536*	0.000	0.960
	CEEMD-VMD-BP	-10.746*	0.000	0.988	-6.286*	0.000	0.983	-4.153*	0.000	0.977
	WPD-CEEMD-BP	-9.021*	0.000	0.988	-11.174*	0.000	0.974	-12.326*	0.000	0.959
	Our proposed	-	-	**0.993**	-	-	**0.986**	-	-	**0.980**

Notes: * represents the 1% significance level; ** represents the 5% significance level. The numbers in bold indicate the best results.

4.Conclusions

By dynamically combining two proven and effective strategies, the reconstruction strategy and the secondary decomposition strategy, we propose a dynamic decomposition-reconstruction-

ensemble approach for STLF. In fact, by introducing the decomposition-reconstruction process based on the dynamic classification and filtering and giving the criteria for determining the components that need to be decomposed again, our proposed model improves the existing decomposition-ensemble forecasting framework. Specifically, our study answers the question of which components need to be reconstructed, which components need to be decomposed, and how many times they need to be decomposed in a decomposition-ensemble framework. Our proposed model takes advantage of decomposition techniques, complexity analysis, reconstruction strategies, secondary decomposition strategies, and a neural network optimized by an automatic hyperparameter optimization algorithm. In our proposed dynamic decomposition-reconstruction strategy, all decomposed components are reconstructed and decomposed accordingly on the basis of their complete analysis and evaluated forecasting performance. In particular, our proposed dynamic decomposition-reconstruction strategy is carefully developed by an in-depth analysis of the characteristics of the data and the extent to which they are simulated by the model.

In the experiment designs, various evaluation perspectives such as forecasting accuracy, precise direction, equality, stability, correlation, comprehensive accuracy, and statistical tests are introduced to verify the superiority of our proposed model. In addition, one-step, three-step, and five-step ahead forecasts are performed on three datasets to verify the applicability of our proposed model. Moreover, five reconstruction strategy models and two secondary decomposition strategy models with related literature are compared with our proposed model in this study. The experimental results showed that our proposed model is superior to all benchmark models for almost all datasets and evaluation metrics. For example, in the dataset of Xi'an, our proposed model is ahead of BP, ICEEMDAN-BP, ICEEMDAN-DCD-BP, ICEEMDAN-SE-BP, ICEEMDAN-FTC-BP, ICEEMDAN-KM-BP, ICEEMDAN-RLJ-BP, CEEMD-VMD-BP, and WPD-CEEMD-BP models by 53.522%, 35.557%, 12.502%, 11.554%, 47.793%, 39.948%, 33.839%, 12.355%, and 38.284%, respectively, for one-step-ahead forecasting in the metrics of MAPE. In summary, our proposed model makes full use of the advantages of each part and fills the shortcomings of the reconstruction strategy and secondary decomposition strategy models, which can be an alternative method for STLF.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This research work was supported by the National Natural Science Foundation of China under Grant No. 71774130, No. 72101197, and No. 71988101, the Fundamental Research Funds for the Central Universities under Grant No. SK2021007.

References

[1] Lebotsa ME, Sigauke C, Bere A, Fildes R, Boylan JE. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. Appl Energy. 2018;222:104-18.http://dx.doi.org/10.1016/j.apenergy.2018.03.155

[2] Barman M, Choudhury NBD. Season specific approach for short-term load forecasting based on hybrid FA-SVM and similarity concept. Energy. 2019;174:886-96.http://dx.doi.org/10.1016/j.energy.2019.03.010

[3] Chen KJ, Chen KL, Wang Q, He ZY, Hu J, He JL. Short-term load forecasting with deep residual networks. IEEE Trans Smart Grid. 2019;10:3943-52.http://dx.doi.org/10.1109/Tsg.2018.2844307
[4] Hu YS, Li JG, Hong MN, Ren JZ, Lin RJ, Liu Y, et al. Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm-A case study of papermaking process. Energy. 2019;170:1215-27.http://dx.doi.org/10.1016/j.energy.2018.12.208
[5] Żymełka P, Szega M. Issues of an improving the accuracy of energy carriers production forecasting in a computer-aided system for monitoring the operation of a gas-fired cogeneration plant. Energy. 2020;209:118431.http://dx.doi.org/10.1016/j.energy.2020.118431
[6] Hobbs BF, Jitprapaikulsarn S, Konda S, Chankong V, Loparo KA, Maratukulam DJ. Analysis of the value for unit commitment of improved load forecasts. IEEE Trans Power Syst. 1999;14:1342-8.http://dx.doi.org/10.1109/59.801894
[7] Lee CM, Ko CN. Short-term load forecasting using lifting scheme and ARIMA models. Expert Syst Appl. 2011;38:5902-11.http://dx.doi.org/10.1016/j.eswa.2010.11.033
[8] Dudek G. Pattern-based local linear regression models for short-term load forecasting. Electr Power Syst Res. 2016;130:139-47.http://dx.doi.org/10.1016/j.epsr.2015.09.001
[9] Li DC, Chang CJ, Chen CC, Chen WC. Forecasting short-term electricity consumption using the adaptive grey-based approach-An Asian case. Omega-Int J Manage Sci. 2012;40:767-73.http://dx.doi.org/10.1016/j.omega.2011.07.007
[10] Yang D, Guo J-e, Li J, Wang S, Sun S. Knowledge Mapping in Electricity Demand Forecasting: A Scientometric Insight. Front Energy Res. 2021;9.http://dx.doi.org/10.3389/fenrg.2021.771433
[11] Wang SX, Wang X, Wang SM, Wang D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. Int J Electr Power Energy Syst. 2019;109:470-9.http://dx.doi.org/10.1016/j.ijepes.2019.02.022
[12] Zhang XB, Wang JZ, Zhang KQ. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm. Electr Power Syst Res. 2017;146:270-85.http://dx.doi.org/10.1016/j.epsr.2017.01.035
[13] Çevik HH, Çunkaş M. Short-term load forecasting using fuzzy logic and ANFIS. Neural Computing and Applications. 2015;26:1355-67.http://dx.doi.org/10.1007/s00521-014-1809-4
[14] Song JJ, Wang JZ, Lu HY. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. Appl Energy. 2018;215:643-58.http://dx.doi.org/10.1016/j.apenergy.2018.02.070
[15] Mohan N, Soman KP, Kumar SS. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Appl Energy. 2018;232:229-44.http://dx.doi.org/10.1016/j.apenergy.2018.09.190
[16] Yu LA, Wang ZS, Tang L. A decomposition-ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Appl Energy. 2015;156:251-67.http://dx.doi.org/10.1016/j.apenergy.2015.07.025
[17] Xiao L, Wang J, Dong Y, Wu J. Combined forecasting models for wind energy forecasting: A case study in China. Renewable Sustainable Energy Rev. 2015;44:271-88.http://dx.doi.org/https://doi.org/10.1016/j.rser.2014.12.012
[18] Fan CD, Ding CK, Zheng JH, Xiao LY, Ai ZY. Empirical mode decomposition based multi-objective deep belief network for short-term power load forecasting. Neurocomputing. 2020;388:110-23.http://dx.doi.org/10.1016/j.neucom.2020.01.031

[19] Li WQ, Chang L. A combination model with variable weight optimization for short-term electrical load forecasting. Energy. 2018;164:575-93.http://dx.doi.org/10.1016/j.energy.2018.09.027
[20] Tang L, Wang S, He KJ, Wang SY. A novel mode-characteristic-based decomposition ensemble model for nuclear energy consumption forecasting. Ann Oper Res. 2015;234:111-32.http://dx.doi.org/10.1007/s10479-014-1595-5
[21] Tian CS, Hao Y, Hu JM. A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization. Appl Energy. 2018;231:301-19.http://dx.doi.org/10.1016/j.apenergy.2018.09.012
[22] Sun XL, Hao J, Li JP. Multi-objective optimization of crude oil-supply portfolio based on interval prediction data. Ann Oper Res. 2020.http://dx.doi.org/10.1007/s10479-020-03701-w
[23] Zhang WY, Qu ZX, Zhang KQ, Mao WQ, Ma YN, Fan X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Convers Manage. 2017;136:439-51.http://dx.doi.org/10.1016/j.enconman.2017.01.022
[24] Zhang ZC, Hong WC. Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. Nonlinear Dyn. 2019;98:1107-36.http://dx.doi.org/10.1007/s11071-019-05252-7
[25] Sharma E, Deo RC, Prasad R, Parisi AV. A hybrid air quality early-warning framework: An hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms. Sci Total Environ. 2020;709:135934.http://dx.doi.org/10.1016/j.scitotenv.2019.135934
[26] Bento PMR, Pombo JAN, Calado MRA, Mariano SJPS. Optimization of neural network with wavelet transform and improved data selection using bat algorithm for short-term load forecasting. Neurocomputing. 2019;358:53-71.http://dx.doi.org/10.1016/j.neucom.2019.05.030
[27] He FF, Zhou JZ, Feng ZK, Liu GB, Yang YQ. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Appl Energy. 2019;237:103-16.http://dx.doi.org/10.1016/j.apenergy.2019.01.055
[28] Jiang M, Jia L, Chen Z, Chen W. The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. Ann Oper Res. 2020.http://dx.doi.org/10.1007/s10479-020-03690-w
[29] Sulandari W, Subanar, Lee MH, Rodrigues PC. Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. Energy. 2020;190:116408.http://dx.doi.org/10.1016/j.energy.2019.116408
[30] Kong XY, Li C, Wang CS, Zhang YS, Zhang J. Short-term electrical load forecasting based on error correction using dynamic mode decomposition. Appl Energy. 2020;261:114368.http://dx.doi.org/10.1016/j.apenergy.2019.114368
[31] Sun W, Huang CC. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. J Cleaner Prod. 2020;243:118671.http://dx.doi.org/10.1016/j.jclepro.2019.118671
[32] Li HT, Bai JC, Li YW. A novel secondary decomposition learning paradigm with kernel extreme learning machine for multi-step forecasting of container throughput. Physica A. 2019;534:122025.http://dx.doi.org/10.1016/j.physa.2019.122025
[33] Yang D, Guo J-e, Sun S, Han J, Wang S. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. Appl Energy. 2022;306:117992.http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2021.117992

[34] Li HT, Jin F, Sun SL, Li YW. A new secondary decomposition ensemble learning approach for carbon price forecasting. Knowledge-Based Syst. 2021;214:106686.http://dx.doi.org/10.1016/j.knosys.2020.106686
[35] Yin H, Dong Z, Chen YL, Ge JF, Lai LL, Vaccaro A, et al. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. Energy Convers Manage. 2017;150:108-21.http://dx.doi.org/10.1016/j.enconman.2017.08.014
[36] Liu ZG, Sun WL, Zeng JJ. A new short-term load forecasting method of power system based on EEMD and SS-PSO. Neural Comput Appl. 2014;24:973-83.http://dx.doi.org/10.1007/s00521-012-1323-5
[37] Sun W, Wang YW. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. Energy Convers Manage. 2018;157:1-12.http://dx.doi.org/10.1016/j.enconman.2017.11.067
[38] Paramesh Kumar N, Vijayabaskar S, Murali L. Forecasting biofuel production using adaptive integrated optimization network model. Fuel. 2021;283:118764.http://dx.doi.org/10.1016/j.fuel.2020.118764
[39] Wang S-p, Hu A-m, Wu Z-x, Liu Y-q, Bai X-w. Multiscale combined codel based on run-length-judgment method and Its application in oil price forecasting. Math Prob Eng. 2014;2014:1-9.http://dx.doi.org/10.1155/2014/513201
[40] Yang WD, Wang JZ, Niu T, Du P. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. Appl Energy. 2019;235:1205-25.http://dx.doi.org/10.1016/j.apenergy.2018.11.034
[41] Sun N, Zhou JZ, Chen L, Jia BJ, Tayyab M, Peng T. An adaptive dynamic short-term wind speed forecasting model using secondary decomposition and an improved regularized extreme learning machine. Energy. 2018;165:939-57.http://dx.doi.org/10.1016/j.energy.2018.09.180
[42] Liu H, Tian HQ, Liang XF, Li YF. Wind speed forecasting approach using secondary decomposition algorithm and elman neural networks. Appl Energy. 2015;157:183-94.http://dx.doi.org/10.1016/j.apenergy.2015.08.014
[43] Gan K, Sun SL, Wang SY, Wei YJ. A secondary-decomposition-ensemble learning paradigm for forecasting PM2.5 concentration. Atmos Pollut Res. 2018;9:989-99.http://dx.doi.org/10.1016/j.apr.2018.03.008
[44] Sun ZX, Zhao MY, Dong Y, Cao X, Sun HX. Hybrid model with secondary decomposition, randomforest algorithm, clustering analysis and long short memory network principal computing for short-term wind power forecasting on multiple scales. Energy. 2021;221:119848.http://dx.doi.org/10.1016/j.energy.2021.119848
[45] Zhang X, Lai KK, Wang SY. A new approach for crude oil price analysis based on empirical mode decomposition. Energy Economics. 2008;30:905-18.http://dx.doi.org/10.1016/j.eneco.2007.02.012
[46] Zhu BZ, Shi XT, Chevallier J, Wang P, Wei YM. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. J Forecast. 2016;35:633-51.http://dx.doi.org/10.1002/for.2395
[47] Xie G, Qian Y, Wang S. A decomposition-ensemble approach for tourism forecasting. Ann Touris Res. 2020;81:102891.http://dx.doi.org/10.1016/j.annals.2020.102891
[48] Fu WL, Wang K, Tan JW, Zhang K. A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting. Energy Convers Manage.

2020;205:112461.http://dx.doi.org/10.1016/j.enconman.2019.112461
[49] Ruiz-Aguilar JJ, Turias I, Gonzalez-Enrique J, Urda D, Elizondo D. A permutation entropy-based EMD-ANN forecasting ensemble approach for wind speed prediction. Neural Comput Appl. 2021;33:2369-91.http://dx.doi.org/10.1007/s00521-020-05141-w
[50] Duan JD, Wang P, Ma WT, Tian X, Fang S, Cheng YL, et al. Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network. Energy. 2021;214:118980.http://dx.doi.org/10.1016/j.energy.2020.118980
[51] Sun SL, Wang SY, Zhang GW, Zheng JL. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. Sol Energy. 2018;163:189-99.http://dx.doi.org/10.1016/j.solener.2018.02.006
[52] Zhu JM, Liu JP, Wu P, Chen HY, Zhou LG. A novel decomposition-ensemble approach to crude oil price forecasting with evolution clustering and combined model. Int J Mach Learn Cybern. 2019;10:3349-62.http://dx.doi.org/10.1007/s13042-019-00922-9
[53] Wang ZC, Chen LR, Ding ZN, Chen HY. An enhanced interval PM2.5 concentration forecasting model based on BEMD and MLPI with influencing factors. Atmos Environ. 2020;223:117200.http://dx.doi.org/10.1016/j.atmosenv.2019.117200
[54] Xiang L, Li JX, Hu AJ, Zhang Y. Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Convers Manage. 2020;220:113098.http://dx.doi.org/10.1016/j.enconman.2020.113098
[55] Sun W, Huang CC. A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network. Energy. 2020;207:118294.http://dx.doi.org/10.1016/j.energy.2020.118294
[56] Jiménez-Cordero A, Morales JM, Pineda S. A novel embedded min-max approach for feature selection in nonlinear support vector machine classification. Eur J Oper Res. 2021;293:24-35.http://dx.doi.org/10.1016/j.ejor.2020.12.009
[57] Seera M, Lim CP, Kumar A, Dhamotharan L, Tan KH. An intelligent payment card fraud detection system. Ann Oper Res. 2021.http://dx.doi.org/10.1007/s10479-021-04149-2
[58] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomed Signal Process Control. 2014;14:19-29.http://dx.doi.org/10.1016/j.bspc.2014.06.009
[59] Bandt C, Pompe B. Permutation entropy: A natural complexity measure for time series. Phys Rev Lett. 2002;88:174102.http://dx.doi.org/10.1103/PhysRevLett.88.174102
[60] Ding SF, Su CY, Yu JZ. An optimizing BP neural network algorithm based on genetic algorithm. Artif Intell Rev. 2011;36:153-62.http://dx.doi.org/10.1007/s10462-011-9208-z
[61] Li C. Designing a short-term load forecasting model in the urban smart grid system. Appl Energy. 2020;266:114850.http://dx.doi.org/10.1016/j.apenergy.2020.114850
[62] Wu Z, Zhou C, Xu F, Lou W. A CS-AdaBoost-BP model for product quality inspection. Ann Oper Res. 2020.http://dx.doi.org/10.1007/s10479-020-03798-z
[63] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc.; 2011. p. 2546-54.
[64] Jiang P, Ma XJ. A hybrid forecasting approach applied in the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms. Appl Math Modell. 2016;40:10631-49.http://dx.doi.org/10.1016/j.apm.2016.08.001

[65] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat. 1995;13:253-63.http://dx.doi.org/10.2307/1392185