Short-term electricity load time series prediction by machine learning model via feature selection and parameter optimization using hybrid cooperation search algorithm

Wen-jing Niu, Zhong-kai Feng*, Shu-shan Li, Hui-jun Wu and Jia-yang Wang

1 Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, People’s Republic of China
2 School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, People’s Republic of China
3 Power Dispatching Control Center, China Southern Power Grid, Guangzhou 510663, People’s Republic of China
4 Power Research Institute, China Southern Power Grid, Guangzhou 510080, People’s Republic of China

* Author to whom any correspondence should be addressed.
E-mail: myfellow@163.com

Keywords: load prediction, support vector machine, cooperation search algorithm, parameter optimization, feature selection, artificial intelligence, machine learning

Abstract
Reliable load time series forecasting plays an important role in guaranteeing the safe and stable operation of modern power system. Due to the volatility and randomness of electricity demand, the conventional forecasting method may fail to effectively capture the dynamic change of load curves. To satisfy this practical necessity, the goal of this paper is set to develop a practical machine learning model based on feature selection and parameter optimization for short-term load prediction. In the proposed model, the ensemble empirical mode decomposition is used to divide the original loads into a sequence of relatively simple subcomponents; for each subcomponent, the support vector machine is chosen as the basic predictor where the real-valued cooperation search algorithm (CSA) is used to seek the best model hyperparameters, while the binary-valued CSA is set as the feature selection tool to determine the candidate input variables; finally, the aggregation of all the submodules’ outputs forms the final forecasting result. The presented method is assessed by short-term load data from four provincial-grid dispatching centers in China. The experiments demonstrate that the proposed model can provide better results than several conventional models in short-term load prediction, while the emerging CSA method is an effective tool to determine the parameter combinations of machine learning method.

1. Introduction

Accurate short-term load forecasting information has a profound impact on not only the social and economic decision-making plan [1–3], but also the safe and stable operation of modern power system, like hydrothermal scheduling, unit commitment, security analysis, risk assessment, and equipment overhaul [4–6]. Generally, load forecasting is an important but challenging production task because the energy curve is usually influenced by a large number of factors at the same time, like economic level, industrial...
structure, weather condition, living style and population [7–9]. Due to the booming economic development and growing extreme climate events, the energy quantity and peak-valley difference of load curves becomes increasingly prominent in recent years [7–9]. Meanwhile, the installed capacity of distributed renewable energy (like wind and solar) with the features of strong volatility and high uncertainty grows rapidly, increasing the change frequency and amplitude of load demand curve [10–12]. Under this engineering background, short-term load forecasting is raising new challenges for the daily operation of power system [13–15].

Over the past few decades, many advanced methods have been developed for short-term load forecasting [16–19]. Based on the employed theory principle, the existing methods can be roughly divided into two different categories: one is the conventional approaches represented by multiple linear regression and autoregressive integrated moving average method [20–22], while another is an intelligent approaches represented by SVM, adaptive network-based fuzzy inference system, ANN and deep machine learning [23–26]. Conventional approaches based on linear analysis have the merits of easy implementation and broad adaptability, but may fail to capture the nonlinearity of load curves in many cases [27–30]. Intelligent methods can obtain satisfying forecasting results without explicitly deducing the underlying mechanism of load series, promoting their widespread applications in practical problems [31–34]. Given the above considerations, the intelligent approach is chosen as the focus of this paper. As one of the most famous intelligent methods, SVM is based on the structural risk minimization theory and Vapnik–Chervonenkis dimension [35–37]. In theory, SVM can guarantee the achievement of global optima for regression and classification problems [38–40]. However, the SVM performance often changes with the combinations of computation parameters [6]. To refill the research gaps, more effective methods should be developed to improve the SVM performance in the short-term load forecasting problem.

In order to overcome the SVM defects, two popular strategies (feature selection and parameter optimization) are employed in this paper. As an effective signal analysis tool, EEMD can divide the original load series into a set of subcomponents with different frequencies and resolutions [41–43]. EEMD has the merits of high flexibility and strong robustness compared with the traditional EMD method [44–46], promoting its wide applications in time series analysis and forecasting. Hence, EEMD is set as the data processing tool to identify the hidden information in the time series. Besides, due to the rapid development of information technology, a variety of swarm intelligence methods are successfully proposed to search for the ideal parameter combination of the SVM method [47–50]. Recently, an emerging method named as CSA is proposed to solve numerical optimization and engineering optimization problems [51–53]. In the CSA method, it is believed that each solution is equal to a staff in the teamwork while a group of staffs can converge to the promising areas around the best solution. After randomly placing the positions of the initial swarm in the state space, the team commutation operation is used to enhance the global exploitation; the reflective learning is used to improve local exploration, while the internal competition is used to guarantee the survival of the fittest. Hence, the RCSA method is employed to search for the optimal parameters of the SVM model for the first time. Moreover, feature selection proves to be an effective way in identifying candidate variables and reducing modeling difficulty [54–56]. In order to achieve this goal, the BCSA method is proposed for the first time. Based on the above analysis, an effective machine learning model using feature selection and parameter optimization is developed to predict electricity load time series. Specifically, the EEMD method is used to divide the load series into several different parts; and then the SVM method using the HCSA method is adopted to model the inputs–outputs relationships; and then the outputs of all the models are combined to produce the final forecasting result. Specially, the HCSA method is mainly composed of two different parts: real-valued zone for parameter optimization and binary-valued zone for feature selection. The feasibility and practicability of the proposed method is fully proved by the experimental results of several power grids in China.

To better understand this paper, the interesting insights of this paper lie in: (a) the RCSA method is used to optimize the hyperparameters of the SVM method; (b) the BCSA method is developed to select candidate input variables; (c) a hybrid method using EEMD, CSA and SVM are developed for short-term load prediction. For all we know, the above three points have not been reported by other research groups by far, providing useful technical reference for relevant engineering problems.

The reminder of this paper is organized as below: the technical details of the presented method is given in section 2. The performance of the presented method is verified in section 3, while the conclusions are given in section 4.

2. Methodologies

2.1. Ensemble empirical mode decomposition
EEMD is an effective signal analysis tool developed to divide the complex data series into a set of small and
simple subseries called IMF [57–60]. In the decomposing process of the EEMD method, the white noise with finite amplitude is added to the original signal to split various scales of sub-signals in the frequency space; and then each IMF subseries is equal to the ensemble mean of several trails [36, 61, 62]. Next, the execution steps of the EEMD method are summarized as below:

Step 1: define the number of ensembles $M$ and the amplitude of the white noise.

Step 2: add the white noise to the original signal $x(t)$ to reconstruct the modified data signal:

$$x_i(t) = x(t) + n_i(t)$$

where $n_i(t)$ and $x_i(t)$ are the $t$th value of the $i$th white noise and the $i$th modified signal.

Step 3: use the standard EMD method to divide the newly-obtained signal $x_i(t)$ into a series of IMFs

$$x_i(t) = \sum_{s=1}^{S} h_{i,s}(t) + r_{i,S}(t),$$

where $S$ is the number of IMFs, $h_{i,s}(t)$ is the $t$th value of the $s$th IMFs associated with $x_i(t)$, $r_{i,S}(t)$ is the residue representing the mean trend of $x_i(t)$.

Step 4: steps 2–3 are repeated $M$ times to produce an ensemble of IMFs. Then, the average of all the obtained IMFs possessing identical frequency is treated as the final result

$$\tilde{h}_s(t) = \frac{1}{M} \sum_{i=1}^{M} h_{i,s}(t)$$

where $\tilde{h}_s(t)$ is the $t$th value of the $s$th IMF subcomponent.

2.2. Hybrid CSA

In practice, many problems often have binary and real variables at the same time. To address this kind of problem, all the decision variables in HCSA method are encoded as real values using the same evolutionary optimizer. To effectively respond the actual problem, all the variables are divided into two groups: real zone and binary zone. As showed in figure 1, the variables in binary zone should be mapped into binary values via a transfer function while the variables in real zone require no processing. By this time, all the variables for the target problem will be known and then the fitness values of all the staffs can be obtained. The technical details of the HCSA algorithm are given in the following sections.

2.2.1. Real-valued CSA

Based on the team cooperation behaviors drawn in figure 2, the CSA method starts from an initial population randomly placed in the feasible decision space, and then iteratively makes use of three evolutionary operators to approach the promising area around the global optimal solution [51–53]: the team communication operator where each staff gain knowledge from leaders to improve the job performance; the reflective learning operator where each staff has the chance to draw lessons from the past and avoid making similar mistakes as far as possible in the future;
and the internal competition operator where elite staffs are conserved and under-performing staffs are dismissed. Generally, the team communication operator helps the population explore in the unknown space with a large search step; the reflective learning operator is used to enhance the local search ability of the swarm with a small search step; the internal competition operator gradually improves the solution’s quality from generation to generation.

Without loss of generality, for the \( J \)-variable minimization optimization problem, the detailed procedures of the CSA method with \( I \) staffs are given as below:

(a) **Team building phase.** The job tasks of all the staffs at hand are randomly assigned in the feasible decision space, which can be expressed as below:

\[
x^k_j = \phi \left( x, \bar{x}_j \right), \quad i \in [1, I], \ j \in [1, J], \ k = 1
\]

where \( x^k_j \) is the \( j \)th value of the \( i \)th staff at the \( k \)th iteration. \( \phi(L, U) \) is the function producing a random number uniformly distributed in the range of \( [L, U] \).

(b) **Team communication operator.** The staffs with superior behaviors are seen as the candidate members in the supervisor and directors board; and then each staff improves the professional skills by gaining favorable information from excellent staffs in both supervisory and directors boards, which can be described as follows:

\[
u^{k+1}_{ij} = \begin{cases} u^{k+1}_{ij} & \text{if } u^{k+1}_{ij} \leq \bar{x}_j \\ \bar{x}_j & \text{if } u^{k+1}_{ij} > \bar{x}_j \end{cases}
\]

where \( \bar{x}_j \) and \( x_j \) are the upper or lower limits of the \( j \)th value of the target problem.

(c) **Reflective learning operator.** In daily work, each staff has the chance to draw lessons from the past to reduce the possible mistakes in the future, which can be described as below:

\[
u^{k+1}_{ij} = \begin{cases} \bar{v}^{k+1}_{ij} & \text{if } v^{k+1}_{ij} \geq \bar{c}_j \\ v^{k+1}_{ij} & \text{if } v^{k+1}_{ij} < \bar{c}_j \end{cases}
\]

\[
u^{k+1}_{ij} = \begin{cases} v^{k+1}_{ij} & \text{if } v^{k+1}_{ij} \leq \bar{c}_j \\ \bar{v}^{k+1}_{ij} & \text{if } v^{k+1}_{ij} > \bar{c}_j \end{cases}
\]
where $v_{i,j}^{k+1}$ is the $j$th value of the $i$th reflective staff at the $k+1$th cycle.

$$c_j = \left( \bar{x}_j + x_{-j} \right) \cdot 0.5$$  \hspace{1cm} (13)

where $F(x)$ is the fitness value of the staff $x$.

The pseudo-code of the CSA method for solving global optimization problem is given as:

1. Set the objective function and physical constraints of the target problem.
2. **Initialization**
   3. Create the initial swarm in the feasible search space by equation (4).
   4. Calculate the fitness values of all initial solutions.
   5. Both group and reflective solutions are set as its initial one.
3. **End Initialization**
4. **Repeat search process**
   8. Update $I$ personal best-known solutions for the current swarm.
   9. Update $M$ global best-known solutions found by far.
   10. Obtain $I$ group solutions by team communication operator equations (5)–(9).
   11. Obtain $I$ feasible reflective solutions by equations (10)–(13).
   12. Compute the fitness values of both group and reflective solutions.
   13. Choose $I$ better solutions by internal competition operator in equation (14).
5. **While** (maximum iterations is not met)
   15. Global best-known individual is seen as the final solution of the target problem.

2.2.2. **Binary-valued CSA**

To effectively solve the discrete optimization problem, this section introduces the BCSA whose search process is similar to the RCSA method. In BCSA, the elements of each staff in the current swarm will be turned into the binary variable (i.e. 0 or 1) via the aid of the classical sigmoid function defined as below:

$$b_{x_{i,j}}^{k+1} = \begin{cases} 0 & \text{if } (S < 0.5) \\ 1 & \text{if } (S \geq 0.5) \end{cases}, \quad i \in [1,I], j \in [1,J], k \in [1,K]$$  \hspace{1cm} (15)

where $w$ and $S$ are the original and modified values, respectively. $b_{x_{i,j}}^{k+1}$ is the binary variable associated with the $j$th value of the $i$th staff at the $k$th cycle.

2.3. **Support vector machine**

SVM is a classical statistical machine learning method for regression and classification problem. In the optimization process, SVM takes advantages of the structural risk minimization principle to produce the global optimal solution [63–65]. Mathematically,
SVM for a regression problem with $N$ training samples can be described as:

$$y_i = w \cdot \phi(x_i) + b, \quad i = 1, 2, \ldots, N$$ \hspace{1cm} (17)$$

where $x_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$ are the input and output variables of the $i$th sample. $\phi(x_i)$ is the non-linear transfer function. $w$ and $b$ are the weight vector and bias, which can be obtained by solving the following constrained optimization problems:

$$\begin{align*}
\min & \quad \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^{N} (\xi_i + \xi_i^*) \\
\text{subject to:} & \quad w \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\
& \quad y_i - w \cdot \phi(x_i) - b \leq \varepsilon + \xi_i
\end{align*}$$ \hspace{1cm} (18)

where $\xi_i$ and $\xi_i^*$ are two slack variables for penalizing the training errors of the loss function over the error tolerance $\varepsilon$. $C$ is the parameter used to determine the empirical risk.

To effectively solve the above problem, the Lagrangian function is reconstructed as below:

$$L = \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^{N} (\xi_i + \xi_i^*) - \sum_{i=1}^{N} a_i \left[ \varepsilon + \xi_i^* + w \cdot \phi(x_i) + b - y_i \right]$$
$$- \sum_{i=1}^{N} \left( \eta_i \xi_i + \eta_i^* \xi_i^* \right) - \sum_{i=1}^{N} a_i^* \left[ \varepsilon + \xi_i^* + y_i - w \cdot \phi(x_i) - b \right]$$ \hspace{1cm} (19)

where $\eta_i, \eta_i^*, a_i, \text{and } a_i^*$ denote the Lagrange multiples.

Based on the famous Karush–Kuhn–Tucker condition, the dual form of the problem in equation (19) can be expressed as below:

$$\begin{align*}
\max & \quad f(a, a^*) = \sum_{i=1}^{N} y_i (a_i - a_i^*) - \varepsilon \sum_{i=1}^{N} (a_i - a_i^*) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) \\
\text{subject to:} & \quad \sum_{i=1}^{N} (a_i - a_i^*) = 0 \\
& \quad 0 \leq a_i, a_i^* \leq C
\end{align*}$$ \hspace{1cm} (20)

After solving the above constrained optimization problem, the desired weight vector can be derived, and then the regression function for the SVM model becomes

$$f(x) = \sum_{i=1}^{N} (a_i^* - a_i) K(x_i, x) + b$$ \hspace{1cm} (21)

where $K(x_i, x)$ is the kernel function meeting the Mercer’s condition. Here, the radial basis function $K(x_i, x) = \exp \left(-\|x_i - x\|^2 / 2\sigma^2 \right)$ is chosen as the kernel function, where $\sigma$ is the value of the standard deviation.

### 2.4. The proposed hybrid method for short-term load time series prediction

As mentioned above, the load curve in practice usually presents the characteristics of strong nonlinearity, randomness and instability due to the growing economy, changeable electric demand and other uncertain factors. In order to accurately trace the
Figure 3. Sketch map of the hybrid method for load time series prediction.

dynamic process of load time series, this paper introduces a hybrid load forecasting method linking the advanced signal decomposition tool and population-based metaheuristic algorithm into the classical machine learning model. Figure 3 shows the sketch map of the proposed method. It can be clearly found that the presented method is composed of three different parts: (a) the first is the decomposition phase where the EEMD method is driven to extract different resolution subcomponents hidden in the original load series; (b) the second is the modeling phase where SVM is used to simulate the possible inputs–outputs relationship hidden in each subseries, and the HCSA is used to find the best combinations of input variables and model parameters; (c) the last is the ensemble phase where the outputs of all the optimized SVM models are combined to produce the final forecasting value. By this way, the proposed method is able to effectively identify the inherent property of nonlinear load series and provides strong technical support for smart grid operation.

Then, the execution procedures of the proposed method in load prediction are given as below:

Step 1: the original load series is decomposed into a group of subseries. For each subseries, the dataset is normalized into the range of [0,1] and then divided into training and testing samples.

Step 2: use the SVM method to construct the forecasting model for all the subcomponents, where the HCSA method is driven to search for the satisfying combination of model parameters and candidate inputs. In the search process, the RCSA method is in charge of optimizing model parameters while the BCSA method is used to eliminate the redundant information from multiple possible input variables.

Step 3: the forecasting results of all the optimized forecasting model are combined to produce an aggregated output. In practice, for each forecasting model, the newly-obtained variable should be normalized into the preset range of [0,1] while the simulated outputs should be renormalized to the original range of the studied signal.

3. Case studies

3.1. Study area and dataset

In practice, it is preferred that the developed load forecasting model can work in different cases. Thus, to fully test the robust performance of the proposed method, the load series of four China’s provincial-grid dispatching centers (A–D) in both summer and winter are chosen for comparison. The reasons lie in that the load series of four power grids have different features, like numbers of peak loads, occurrence times of both peak and valley loads, total loads, as well as peak-valley differences. If not robust enough, the developed method may be suitable for one power grid and fail to work for another grid; but not vice versa. As illustrated in figure 4, the loads in summer are considered to reflect the huge energy demand caused by the high temperature weathers; besides, the load in winter are considered in the second case. Table 1 shows the statistical information of four power grids in summer and winter, demonstrating the huge differences of various load curves. For the studied power grids, about 1600 quarter-hour load series in both winter and summer are collected, where the first 75% observation are chosen for model training while the left is used for testing the performance of the forecasting model. The forecasting window as a quarter-hour. The experiments are executed on a personal computer with 3.30 GHz processor and 8 GB RAM.

3.2. Evaluation indicators

To fully check the quality of various forecasting model, the RMSE, MAE, MARE, and NSE are chosen as the performance evaluation indexes for comparison. Generally, the better the forecasting mode, the larger the values of NSE and R, the smaller values of RMSE, MAE and MARE. Then, the definitions of five statistical measures are given as below:
Figure 4. Load curves of four power grids in summer and winter. (a) Summer. (b) Winter.

Table 1. Statistical information of four power grids in summer and winter.

| Index | Summer | Winter |
|-------|--------|--------|
|       | Grid A | Grid B | Grid C | Grid D | Grid A | Grid B | Grid C | Grid D |
| Max   | 25788.0| 63845.0| 50319.0| 25393.0| 21549.0| 62228.0| 44864.0| 22618.0|
| Mean  | 19273.8| 53519.7| 38568.2| 21063.7| 16526.6| 53255.8| 35001.0| 18111.6|
| Min   | 12273.0| 44115.0| 26953.0| 15255.0| 11095.0| 43840.0| 25234.0| 13619.0|
| Std.  | 3438.5 | 4701.4 | 5341.7 | 2274.1 | 2771.6 | 4587.5 | 5078.6 | 2269.6 |

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (22) \]

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \times 100\% \quad (23) \]

\[ \text{MARE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (24) \]

\[ R = \frac{\sum_{i=1}^{n} [y_i - y_{avg}](\hat{y}_i - \hat{y}_{avg})}{\sqrt{\sum_{i=1}^{n} (y_i - y_{avg})^2(\hat{y}_i - \hat{y}_{avg})^2}} \quad (25) \]

\[ \text{CE} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - y_{avg})^2} \quad (26) \]

where \( n \) is the number of samples. \( y_i \) and \( \hat{y}_i \) are the \( i \)th observed and forecasted load values, \( y_{avg} \) and \( \hat{y}_{avg} \) are the mean value of the observed and forecasted load series.

3.3. Simulation results

3.3.1. Case 1: validation of the proposed method in summer

3.3.1.1. Decomposing results

The original load data series is decomposed into several subseries by the EEMD method. Figure 5 illustrates the obtained decomposition results for two
power grids. It can be found that obvious differences exist in the obtained subcomponents: the first subseries has the largest amplitude and frequency while the other subseries gradually changes the frequency and amplitude. Specially, the first two IMFs have no obvious changing rules and fluctuates frequently, reflecting the random high-frequency features of the original loads; the 3rd IMFs basically varies by hours, reflecting the hourly-level feature of original loads; the 4th and 5th IMFs are the high-frequency periodic components varying in days; the following 6–8th IMFs roughly represent low-frequency fluctuations varying in weeks; while the last IMF denote the overall trend of the load signal. Hence, the feasibility of the EEMD method in identifying internal features of the target loads is successfully verified.

### 3.3.1.2. Model development

Given the inputs has a great influence on the model performance, the set of the candidate input variables should be determined before optimizing the forecasting model. For the proposed method, the PACF is used to identify the initial input sets. Generally, the variable at lag $t$ can be chosen as the possible influence factor as the relevant PACF value falls into the confidence interval. Based on the PACF value and other relevant information, the initial set of input variables for various subseries can be roughly determined.

Several artificial intelligence methods are introduced for comparison, including ANN, ELM, SVM, SVM couple with EMD (SVM–EMD), SVM couple with EEMD (SVM–EEMD). For the standard ANN, ELM and SVM methods, the original loads are chosen as the target object. The trial-and-error method is used to select the number of hidden neurons in both ANN and ELM, while the grid search method is used to identify the computational parameters of SVM. For both SVM–EMD and SVM–EEMD method, the decomposed results by EMD or EEMD are selected as the target object, while the relevant parameters are also determined by the grid search method. For the proposed method, the decomposed result produced by the EEMD method are selected as the target object, while the decomposed result produced by the EEMD method are selected as the target object, while the relevant parameters are also determined by the grid search method. For the proposed method, the decomposed result produced by the EEMD method are selected as the target object, while the relevant parameters are also determined by the grid search method. For the proposed method, the decomposed result produced by the EEMD method are selected as the target object, while the relevant parameters are also determined by the grid search method.

### 3.3.1.3. Simulation results

Table 2 shows the statistical results of various methods for four power grids in summer. It can be seen that in the load prediction works, there are obvious differences in the performances of three artificial intelligence methods (like SVM, ELM and ANN), demonstrating the importance of the training technique and model theory; secondary, the modified SVM methods are able to yield better results than the standard SVM method, proving the role of decomposition-ensemble tool in improving the model performances; moreover, the proposed method has the ability of producing the best performances with respect to various indexes in different power grids, demonstrating the feasibility of feature selection and parameter optimization. For instance, as compared with the RMSE values of the ELM and SVM approaches, the proposed forecasting model can make about 52.25% and 35.03% reductions for grid A during the training phase; compared with the MAE values of the grid A.
Table 2. Statistical results of various methods for four power grids in summer.

|         | Training |         |         |         |         |         |
|---------|----------|---------|---------|---------|---------|---------|
|         | RMSE     | MAE     | MAPE    | R       | NSE     |         |
| Grid A  |          |         |         |         |         |         |
| ANN     | 716.7929 | 518.5718| 2.9104  | 0.9874  | 0.9582  | 969.0185|
| ELM     | 677.0698 | 483.4867| 2.7156  | 0.9888  | 0.9627  | 878.2691|
| SVM     | 497.6590 | 370.6275| 2.0408  | 0.9936  | 0.9798  | 679.4857|
| SVM–EMD | 494.4460 | 367.5201| 2.0228  | 0.9936  | 0.9801  | 682.8092|
| SVM–EEMD| 452.7589 | 348.4554| 1.8876  | 0.9945  | 0.9833  | 649.8781|
| Proposed| 323.3099 | 210.9559| 1.1608  | 0.9958  | 0.9915  | 261.5060|
| Grid B  |          |         |         |         |         |         |
| ANN     | 844.0808 | 677.3304| 1.2554  | 0.9878  | 0.9687  | 924.2439|
| ELM     | 829.9612 | 538.9670| 1.0033  | 0.9870  | 0.9697  | 787.9220|
| SVM     | 807.2059 | 506.8761| 0.9382  | 0.9871  | 0.9713  | 740.9188|
| SVM–EMD | 797.1339 | 491.2034| 0.9102  | 0.9873  | 0.9721  | 717.4005|
| SVM–EEMD| 711.0988 | 415.9180| 0.7742  | 0.9898  | 0.9778  | 584.2385|
| Proposed| 504.3095 | 306.3620| 0.5607  | 0.9944  | 0.9888  | 472.4968|
| Grid C  |          |         |         |         |         |         |
| ANN     | 965.0460 | 753.2017| 1.9577  | 0.9871  | 0.9678  | 928.1296|
| ELM     | 825.3206 | 604.0199| 1.5655  | 0.9901  | 0.9765  | 778.6545|
| SVM     | 848.4581 | 549.8705| 1.4258  | 0.9888  | 0.9751  | 686.1739|
| SVM–EMD | 725.8955 | 495.4683| 1.2869  | 0.9919  | 0.9818  | 611.3340|
| SVM–EEMD| 724.0771 | 479.1950| 1.2389  | 0.9918  | 0.9819  | 591.5048|
| Proposed| 648.1260 | 350.9830| 0.9175  | 0.9927  | 0.9855  | 443.2897|
| Grid D  |          |         |         |         |         |         |
| ANN     | 663.1577 | 531.9854| 2.6504  | 0.9799  | 0.9200  | 576.8217|
| ELM     | 542.0487 | 421.5031| 2.0930  | 0.9849  | 0.9466  | 456.9889|
| SVM     | 480.3600 | 333.8053| 1.6551  | 0.9873  | 0.9615  | 348.8931|
| SVM–EMD | 442.7864 | 310.2736| 1.5439  | 0.9876  | 0.9644  | 320.6406|
| SVM–EEMD| 440.6193 | 278.7966| 1.3947  | 0.9860  | 0.9647  | 276.8319|
| Proposed| 342.1828 | 156.6660| 0.7870  | 0.9984  | 0.9797  | 163.5963|

Figure 6. Improvements of the proposed method compared with the proposed methods during the testing phase in summer.
decomposition-coordination technique. The proposed method outperforms the control methods in approximating the dynamic change of the load series because its trend lines in different cases are closer to the ideal value. Thus, the feasibility of the proposed method in forecasting load series is proved again.

3.3.2. Case 2: validation of the proposed method in winter

Table 3 shows the statistical results of various methods for four power grids in winter. It can be found that for four power grids, the forecasting results of the proposed method are better than five control methods during both training and testing phases in terms of various indexes; besides, three standard methods are inferior to the decomposition and ensemble-based forecasting method. For instance, as used to forecast the load of grid D, the proposed method can make about 53.82%, 48.29%, 37.16%, 31.49% and 25.76% reductions as compared with the RMSE values of the ANN, ELM, SVM, SVM–EMD and SVM–EEMD methods during the testing phase. Thus, the superior performance of the proposed method for load forecasting is proved.
Table 3. Statistical results of various methods for four power grids in winter.

| Grid | Method     | RMSE | MAE   | MAPE | R   | NSE | RMSE | MAE   | MAPE | R   | NSE |
|------|------------|------|-------|------|-----|-----|------|-------|------|-----|-----|
| A    | ANN        | 466.3756 | 335.9871 | 2.1694 | 0.9908 | 0.9728 | 342.9006 | 273.8941 | 1.6416 | 0.9948 | 0.9826 |
|      | ELM        | 450.0194 | 322.9625 | 2.0784 | 0.9912 | 0.9746 | 329.8254 | 264.8980 | 1.5825 | 0.9951 | 0.9839 |
|      | SVM        | 387.0731 | 289.0081 | 1.8525 | 0.9935 | 0.9812 | 299.8512 | 240.8496 | 1.4429 | 0.9957 | 0.9867 |
|      | SVM–EMD    | 317.5099 | 219.8450 | 1.4082 | 0.9948 | 0.9874 | 227.3179 | 184.2451 | 1.1118 | 0.9972 | 0.9924 |
|      | SVM–EEMD   | 297.4857 | 199.4481 | 1.2728 | 0.9952 | 0.9889 | 207.7816 | 167.9622 | 1.0147 | 0.9974 | 0.9936 |
|      | Proposed   | 277.8122 | 151.1230 | 0.9617 | 0.9952 | 0.9903 | 170.1500 | 132.5490 | 0.8010 | 0.9979 | 0.9957 |
| B    | ANN        | 1001.5242 | 815.4982 | 1.5733 | 0.9879 | 0.9513 | 997.2770 | 834.2780 | 1.5825 | 0.9873 | 0.9545 |
|      | ELM        | 1035.7502 | 832.0363 | 1.5926 | 0.9873 | 0.9479 | 941.6978 | 817.6565 | 1.5427 | 0.9900 | 0.9594 |
|      | SVM        | 1052.7062 | 862.7715 | 1.6502 | 0.9878 | 0.9462 | 962.5749 | 842.6516 | 1.5910 | 0.9901 | 0.9576 |
|      | SVM–EMD    | 897.6125 | 695.7813 | 1.3259 | 0.9892 | 0.9609 | 799.7289 | 679.0669 | 1.2799 | 0.9924 | 0.9707 |
|      | SVM–EEMD   | 808.8271 | 613.3215 | 1.1674 | 0.9903 | 0.9682 | 714.2517 | 598.7002 | 1.0147 | 0.9934 | 0.9767 |
|      | Proposed   | 485.5940 | 296.8861 | 0.5557 | 0.9943 | 0.9885 | 426.2931 | 285.9316 | 0.5267 | 0.9959 | 0.9917 |
| C    | ANN        | 768.8131 | 535.5623 | 1.5589 | 0.9896 | 0.9777 | 639.1345 | 504.5127 | 1.4404 | 0.9920 | 0.9830 |
|      | ELM        | 752.4032 | 477.2716 | 1.3821 | 0.9899 | 0.9787 | 568.8456 | 440.1853 | 1.2469 | 0.9936 | 0.9865 |
|      | SVM        | 702.3083 | 439.5180 | 1.2670 | 0.9915 | 0.9814 | 529.8426 | 396.5874 | 1.1216 | 0.9947 | 0.9883 |
|      | SVM–EMD    | 696.1356 | 400.6597 | 1.1542 | 0.9912 | 0.9817 | 476.1403 | 356.9416 | 1.0096 | 0.9956 | 0.9906 |
|      | SVM–EEMD   | 692.0658 | 384.4934 | 1.1026 | 0.9912 | 0.9819 | 466.3754 | 347.0769 | 0.9778 | 0.9957 | 0.9910 |
|      | Proposed   | 584.0246 | 277.4214 | 0.7891 | 0.9936 | 0.9871 | 391.8466 | 259.4235 | 0.7170 | 0.9969 | 0.9936 |
| D    | ANN        | 410.1866 | 294.5324 | 1.6477 | 0.9873 | 0.9675 | 376.2928 | 315.6420 | 1.8218 | 0.9927 | 0.9715 |
|      | ELM        | 381.9772 | 263.1803 | 1.4763 | 0.9897 | 0.9718 | 336.0588 | 281.8090 | 1.3631 | 0.9942 | 0.9773 |
|      | SVM        | 338.1524 | 218.4644 | 1.2217 | 0.9907 | 0.9779 | 276.3230 | 230.1308 | 1.3345 | 0.9957 | 0.9846 |
|      | SVM–EMD    | 308.6905 | 201.4284 | 1.1257 | 0.9922 | 0.9816 | 253.6314 | 211.5786 | 1.2196 | 0.9962 | 0.9871 |
|      | SVM–EEMD   | 365.0591 | 193.2180 | 0.8033 | 0.9958 | 0.9743 | 234.0752 | 193.3402 | 1.1106 | 0.9967 | 0.9890 |
|      | Proposed   | 207.9075 | 134.3764 | 0.7376 | 0.9958 | 0.9917 | 173.7683 | 132.7429 | 0.7403 | 0.9970 | 0.9939 |

Figure 8. Improvements of the proposed method compared with the proposed methods during the testing phase in winter.

Figure 8 shows the comparison between the proposed method and other methods in terms of four statistical indicators. It can be clearly observed that a single forecasting method (like SVM and ELM) can make obvious improvements with respect to various performance indexes in the testing phase. Besides, the results of all the methods can be improved to different degree compared with the proposed method. Thus, we can draw the conclusion that the proposed method is a powerful method for short-term load time series forecasting.

Figure 9 shows the results of various forecasting methods for four power grids in the testing phase. It can be seen that the developed methods can obtain satisfying forecasting results for four power grids, while the proposed method has the best performance among in different scenarios since its trend lines are closer to the observed load line than the other methods. Thus, the proposed method can provide superior performances as used to forecast the load data series in different cases.

3.3.3. Case 3: comparison with different swarm intelligence methods

To further test the performance of the proposed method, the classical GA and PSO are chosen to optimize the SVM parameters. Table 4 lists the statistical results of various SVM methods using different strategies for four power grids. In SVM–EEMD–PSO and SVM–EEMD–GA, the original load is divided into several subcomponents while the SVM method based on PSO or GA are driven for modeling. For table 4, it can be found that three SVM variants outperforms the standard SVM method,
demonstrating the effectiveness of decomposing technique and swarm intelligence; besides, the proposed method yields better or same statistical indexes as compared with two other SVM variants during both training and testing, proving the superiority of the CSA approach in parameter optimization. Therefore, it can be concluded that the proposed method is an effective tool to accurately forecast the load series.

3.4. Result discussions
From the above analysis, it can be clearly found that in short-term load time series prediction, the developed model based on the dynamic combinations of three effective methods is capable of producing better performances than several traditional methods with distinct improvements in the statistical indexes. Based on the EEMD method, the complex load time series with the stochastic and time-dependent features can be transformed into a series of relatively stable subseries, making an obvious reduction in the modeling complexity; and then the HCSCA method is driven to optimize the computational parameters of the SVM model for each subcomponent, which can help improve the network compactness and generalization ability of the forecasting model; finally, the result is obtained by summing the outputs of all the SVM models, producing a synergetic effect in the final predicted value. Thus, the hybrid model based on the

Figure 9. Load forecasting results of various methods for four power grids in winter. (a) Grid A, (b) Grid B, (c) Grid C, (d) Grid D.
decomposition-and-ensemble idea can enhance the forecasting performances compared with several conventional methods.

Although the feasibility of the proposed method is proved in the real-world load demand of several power grids in China, its practical application may face some possible defects, one is the credibility of the EEMD method since its decomposing results change with the features (like length or magnitude) of the studied dataset; the second is the local minimum of the SVM model since even with the aid of the HCSA method, it is still difficult to guarantee the best choice of the input variables and network parameters. Thus, future research can be directed to the developments of the data decomposition tool with more stable performance, evolutionary algorithm with strong global search ability [66], input variable determination tool and forecasting model with stronger generalization ability. Besides, there is no uncertainty quantification in the resultant forecasting result because the deterministic load value is preferred in the decision-making process made by operators and managers of power grids [67, 68]. For better and reliable predictions, it is quite necessary to consider the relevant input variables (like weather conditions and date attribute) in the uncertain forecasting model in the future.

4. Conclusion

In order to improve the short-term electricity load prediction accuracy, this research develops a practical machine learning model based on feature selection and parameter optimization. In the proposed method, the EEMD is employed to divide the complex load data into a set of subcomponents that will be modeled by the famous SVM method; and then the HCSA is used to find the best combinations of model hyperparameters and input variables; finally, the outputs of all the submodules are used to form the final forecasting results. The load series of four provincial-grid dispatching centers are used to verify the performances of the proposed method. The simulations demonstrate that the presented method can effectively enhance the forecasting accuracy of several traditional methods in terms of all the employed assessment indexes. Hence, a powerful machine learning model with robust performances is presented to forecast the short-term load demand curve in practice. For the policy and decision makers, the proposed methodology can provide strong technical support information to determine the satisfying operation plan of various power resources (like hydropower, thermal and wind), rationally arrange the maintenance plans of the generating units, maintain the safety and stability of power grid. By this way, the fuel cost and pollutant emission of electric power system can be sharply improved in the long run, which can effectively increase the economy benefits, environmental and social benefits.

Data availability statement

Due to the strict security requirements from the departments, some or all data, models, or code generated or used in the study are proprietary or confidential in nature and may only be provided with restrictions (e.g. anonymized data).

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (52009012), Natural Science Foundation of Hubei Province (2020CFB340), National Natural Science Foundation of China (51709119). The writers would like to thank editors and reviewers for their valuable comments and suggestions.

Table 4. Statistical results of various SVM methods using different strategies for four power grids.

| Grid Method | RMSE | MAE | MAPE | R | NSE | RMSE | MAE | MAPE | R | NSE |
|-------------|------|-----|------|---|-----|------|-----|------|---|-----|
| A SVM       | 392.8276 | 247.0508 | 1.4585 | 0.9901 | 0.9798 | 286.6775 | 221.6965 | 1.4256 | 0.9940 | 0.9872 |
| SVM–EEMD–GA | 275.8251 | 158.6130 | 0.9439 | 0.9950 | 0.9900 | 184.3116 | 144.7555 | 0.8792 | 0.9974 | 0.9947 |
| SVM–EEMD–PSO | 268.9747 | 156.7671 | 0.9309 | 0.9953 | 0.9905 | 183.8701 | 144.5092 | 0.8774 | 0.9974 | 0.9947 |
| SVM–EEMD–PSO | 277.8547 | 160.9429 | 0.8907 | 0.9957 | 0.9913 | 176.8341 | 138.2800 | 0.8438 | 0.9976 | 0.9951 |
| SVM         | 703.2616 | 501.4699 | 0.9336 | 0.9891 | 0.9758 | 1131.1170 | 789.8251 | 1.6931 | 0.9825 | 0.9395 |
| SVM–EEMD–GA | 490.7092 | 292.8152 | 0.5433 | 0.9941 | 0.9882 | 464.4228 | 317.9422 | 0.6388 | 0.9951 | 0.9898 |
| SVM–EEMD–PSO | 523.9341 | 299.3658 | 0.5569 | 0.9933 | 0.9865 | 468.4656 | 323.9565 | 0.6522 | 0.9951 | 0.9896 |
| SVM         | 778.5265 | 532.4314 | 0.5251 | 0.9952 | 0.9904 | 456.1988 | 311.5914 | 0.6236 | 0.9952 | 0.9902 |
| SVM–EEMD–GA | 571.3678 | 269.2787 | 0.7604 | 0.9936 | 0.9871 | 354.6176 | 250.7333 | 0.8017 | 0.9967 | 0.9930 |
| SVM–EEMD–PSO | 594.9879 | 272.6848 | 0.7712 | 0.9931 | 0.9860 | 358.6236 | 255.5455 | 0.8153 | 0.9966 | 0.9929 |
| SVM         | 442.0539 | 284.0067 | 0.5251 | 0.9952 | 0.9904 | 456.1988 | 311.5914 | 0.6236 | 0.9952 | 0.9902 |
| SVM–EEMD–GA | 525.9341 | 299.3658 | 0.5569 | 0.9933 | 0.9865 | 468.4656 | 323.9565 | 0.6522 | 0.9951 | 0.9896 |
| SVM–EEMD–PSO | 491.7011 | 266.9856 | 0.7556 | 0.9953 | 0.9905 | 337.8191 | 243.3704 | 0.7927 | 0.9969 | 0.9937 |
| SVM         | 348.6657 | 235.8153 | 1.3078 | 0.8984 | 0.9781 | 368.6000 | 269.1371 | 1.6919 | 0.9872 | 0.9684 |
| SVM–EEMD–GA | 346.5720 | 137.2891 | 1.5270 | 0.9897 | 0.9761 | 1038.6650 | 729.6898 | 2.6952 | 0.9814 | 0.9403 |
| SVM–EEMD–PSO | 323.8329 | 134.3829 | 0.7333 | 0.9959 | 0.9918 | 172.5329 | 130.0298 | 0.7549 | 0.9967 | 0.9931 |
| SVM         | 181.9104 | 131.6025 | 0.7213 | 0.9790 | 0.9940 | 162.4667 | 126.0261 | 0.7352 | 0.9970 | 0.9939 |
References

[1] Piao M J, Li Y P and Huang G H 2014 Development of a stochastic simulation-optimization model for planning electric power systems—a case study of Shanghai, China Energy Convers. Manage. 86 111–24

[2] Catalão J P S, Pousinho H M I and Mendes V M F 2011 Short-term wind power forecasting in Portugal by neural networks and wavelet transform Renew. Energy 36 1245–51

[3] Catalão J P S, Pousinho H M I and Mendes V M F 2011 Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal IEEE Trans. Sustain. Energy 2 50–9

[4] Ahmad T, Chen H, Shair J and Xu C 2019 Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management Int. J. Refrig. 98 399–409

[5] Ansari M R, Amjady N and Vatani B 2014 Stochastic security-constrained hydrothermal unit commitment considering uncertainty of load forecast, inflows to reservoirs and unavailability of units by a new hybrid decomposition strategy IET Gener. Transm. Distrib. 8 1900–15

[6] Barman M and Dev Choudhury N B 2020 A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India Sustain. Cities Soc. 61 102311

[7] Chalal M L, Medjdoub B, Benai N and Shahrabi R 2020 Big Data to support sustainable urban energy planning: the EnvEnergy project Front. Eng. Manage. 7 259–74

[8] Ziyoun G and Lixing Y 2019 Energy-saving operation approaches for urban rail transit systems Front. Eng. Manage. 6 139–51

[9] Lu Z, Dan Z, Liwen M, Ran Y and Shuang W 2019 Operation management of green ports and shipping networks: overview and research opportunities Front. Eng. Manage. 6 152–62

[10] Li Z, Li Y, Liu Y, Wang P, Lu R and Gooi H B 2020 Deep learning based densely connected network for load forecasting IEEE Trans. Power Syst. (https://doi.org/10.1109/TPWRS.2020.3048359)

[11] Niu W J, Feng Z K, Liu S, Chen Y B, Xu Y S and Zhang J 2021 Multiple hydropower reservoirs operation by hyperbolic grey wolf optimizer based on elitism selection and adaptive mutation Water Resour. Manage. 35 573–91

[12] Liu Y, Li Y, Gooi H B, Jian Y, Xin H, Jiang X and Pan J 2019 Distributed robust energy management of a multicrogrid system in the real-time energy market IEEE Trans. Sustain. Energy 10 396–406

[13] Dai S, Niu D and Li Y 2018 Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey Wolf optimization algorithm Energies 11 163

[14] Dong B, Li Z, Rahman S M M and Vega R 2016 A hybrid model approach for forecasting future residential electricity consumption Energy Build. 117 341–51

[15] Duan M, Darvishan A, Mohammadit R, Wakil K and Abedinia O 2018 A novel hybrid prediction model for aggregated loads of buildings by considering the electric vehicles Sustain. Cities Soc. 41 205–19

[16] Hu J, Zhao G, Mu X, Hörmann G, Tian P, Gao P and Sun W 2020 Effect of soil and water conservation measures on regime-based suspended sediment load during floods Sustain. Cities Soc. 55 102044

[17] Ilbégzi M, Gholeihi M and Dehghanbanadaki A 2020 Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm Sustain. Cities Soc. 61 102343

[18] Khalid B and Javid N 2020 A survey on hyperparameters optimization algorithms of forecasting models in smart grid Sustain. Cities Soc. 61 102275

[19] Rafiei M, Niknam T, Aghaei J, Shafie-Khah M and Catalão J P S 2018 Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine IEEE Trans. Smart Grid 9 6961–71

[20] Loy-Benitez J, Li Q, Nam K and Yoo C 2020 Sustainable subway indoor air quality monitoring and fault-tolerant ventilation control using a sparse autoencoder-driven sensor self-validation Sustain. Cities Soc. 52 101847

[21] Luo X J, Owedaile L O, Ajayi O A and Akinade O O 2020 Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads Sustain. Cities Soc. 61 102283

[22] Mahgoub A O, Gowid S and Ghani S 2020 Global evaluation of WBGT and SET indices for outdoor environments using thermal imaging and artificial neural networks Sustain. Cities Soc. 60 102182

[23] Ni K, Wang J, Tang G and Wei D 2019 Research and application of a novel hybrid model based on a deep neural network for electricity load forecasting: a case study in Australia Energies 12 2467

[24] Wen S, Hu R, Yang Y, Huang T, Zeng Z and Song Y D 2019 Memristor-based echo state network with online least mean square IEEE Trans. Syst. Man Cybern. Syst. 49 1787–96

[25] Zhao T, Wang J and Zhang Y 2019 Day-ahead hierarchical probabilistic load forecasting with linear quantile regression and empirical copulas IEEE Access 7 80969–79

[26] Niu W J, Feng Z K, Feng E F, Xu Y S and Min Y W 2021 Parallel optimization and swarm intelligence based artificial intelligence model for multi-step-ahead hydrological time series prediction Sustain. Cities Soc. 66 102686

[27] Cao Y, Cao Y, Guo Z, Huang T and Wen S 2020 Global exponential synchronization of delayed memristive neural networks with reaction–diffusion terms Neural Netw. 123 70–81

[28] Wen S, Dong M, Yang Y, Zhou P, Huang T and Chen Y 2020 End-to-end detection-segmentation network for face labeling IEEE Trans. Emerg. Top. Comput. Intell. 10–1109

[29] Wang Y, Cao Y, Guo Z, Huang T and Wen S 2020 Event-based sliding-mode synchronization of delayed memristive neural networks via continuous/periodic sampling algorithm Appl. Math. Comput. 383 125379

[30] Wen S, Chen J, Wu Y, Yan Z, Cao Y, Yang Y and Huang T 2020 CKFO: convolution kernel first operated algorithm with applications in memristor-based convolutional neural network IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 1

[31] Wang S, Cao Y, Guo Z, Yan Z, Wen S and Huang T 2020 Periodic event-triggered synchronization of multiple memristive neural networks with switching topologies and parameter mismatch IEEE Trans. Cybern. 1–11

[32] Xiao J, Zhong S and Wen S 2021 Improved approach to the problem of the global Mittag-Leffler synchronization for fractional-order multidimension-valued BAM neural networks based on new inequalities Neural Netw. 133 87–100

[33] Wang H, Wei G, Wen S and Huang T 2020 Impulsive disturbance on stability analysis of delayed quaternion-valued neural networks Appl. Math. Comput. 390 126580

[34] Wen S, Zeng Z and Huang T 2013 Associative learning of integrate-and-fire neurons with memristor-based synapses Neural Process. Lett. 38 69–80

[35] Wang W C, Xu D M, Chau K W and Chen S 2013 Improved algorithm of forecasting models in smart grid IEEE Trans. Sustain. Energy 3 50–57

[36] Tan Q F, Lei X H, Wang X, Wang H, Wen X, Ji Y and Kang A Q 2018 An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach J. Hydroinf. 20 767–80

[37] Khalid B, Zeinal Hamadani A and Bijari M 2012 A fuzzy intelligent approach to the classification problem in gene expression data analysis Knowl.-Based Syst. 27 465–74

[38] Lu J, Yang X and Zhang G 2008 Support vector machine-based multi-source multi-attribute information
integration for situation assessment Expert Syst. Appl. 34 1333–40
[39] Lin J Y, Cheng C T and Chau K W 2006 Using support vector machines for long-term discharge prediction Hydrolog. Sci. J. 51 599–612
[40] Li L, Zhao X, Tseng M and Tan R R 2020 Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm J. Clean. Prod. 242 118447
[41] Wang W C, Chau K W, Qiu L and Chen Y B 2015 Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition Environ. Res. 139 46–54
[42] Wang W C, Chau K W, Xu D M and Chen X Y 2015 Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition Water Resour. Manage. 29 2639–75
[43] Wang D, Wei S, Luo H, Yue C and Grunder O 2017 A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine Sc. Total Environ. 580 719–33
[44] Rezaie-Balf M, Kim S, Fallah H and Alaghmand S 2019 Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: application on the perennial rivers in Iran and South Korea J. Hydrod. 572 470–85
[45] Chen P, Wang Y, You G J and Wei C 2017 Comparison of methods for non-stationary hydrologic frequency analysis: case study using annual maximum daily precipitation in Taiwan J. Hydrod. 545 197–211
[46] Fijani E, Barzegar R, Deo R, Tziritis E and Konstantinos S 2019 Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters Sc. Total Environ. 648 839–53
[47] Wang W C, Chau K W, Cheng C T and Qiu L 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series J. Hydrod. 374 294–306
[48] Niu W J and Feng Z K 2021 Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management Sustain. Cities Soc. 64 102562
[49] Fu W, Tan J, Li C, Zou Z, Li Q and Chen T 2018 A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm Entropy 20 626
[50] Zhao E, Zhang Z and Bohlool N 2020 Cost and load forecasting by an integrated algorithm in intelligent electricity supply network Sustain. Cities Soc. 60 102243
[51] Feng Z K, Niu W J and Liu S 2021 Cooperation search algorithm: a novel metaheuristic evolutionary intelligence algorithm for numerical optimization and engineering optimization problems Appl. Soft Comput. 98 106734
[52] Feng Z and Niu W 2021 Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions Knowl.-Based Syst. 211 106580
[53] Feng Z, Niu W, Tang Z, Xu Y and Zhang H 2021 Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction J. Hydrod. 595 126062
[54] Houssein E H, Hosney M E, Oliva D, Mohamed W M and Hassaballah M 2020 A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery Comput. Chem. Eng. 133 106668
[55] Rodrigues D, Pereira L A M, Nakamura R Y M, Costa K A P, Yang X S, Souza A N and Papa J P 2014 A wrapper approach for feature selection based on bat algorithm and optimum-path forest Expert Syst. Appl. 41 2250–8
[56] Aljarah I, Mafarja M, Heidari A A, Faris H, Zhang Y and Mirjalili S 2018 Asynchronous accelerating multi-leader salp chains for feature selection Appl. Soft Comput. J. 71 964–79
[57] Jiang Y, Huang G, Yang Q, Yan Z and Zhang C 2019 A novel probabilistic wind speed prediction approach using real time refined variational model decomposition and conditional kernel density estimation Energy Convers. Manage. 185 758–73
[58] Qian Z, Lei X, Zareipour H and Chen N 2019 A review and discussion of decomposition-based hybrid models for wind energy forecasting applications Appl. Energy 235 939–53
[59] Liu G, Zhou J, Jia B, He F, Yang Y and Sun N 2019 Advance short-term wind energy quality assessment based on instantaneous standard deviation and variogram of wind speed by a hybrid method Appl. Energy 238 643–67
[60] Cheng Y, Wang Z, Chen B, Zhang W and Huang G 2019 An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis ISA Trans. 91 218–34
[61] Huang Y, Schmitt F G, Lu Z and Liu Y 2009 Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis J. Hydrod. 373 103–11
[62] Zhang X, Peng Y, Zhang C and Wang B 2015 Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences J. Hydrod. 530 137–52
[63] Chau K W and Wu C L 2010 A hybrid model coupled with singular spectrum analysis for daily rainfall prediction J. Hydroinf. 12 458–73
[64] Wu C L, Chau K W and Li Y S 2009 Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques Water Resour. Res. 45
[65] Wu C L, Chau K W and Li Y S 2008 River stage prediction based on a distributed support vector regression J. Hydrod. 358 96–111
[66] Feng Z K, Liu S, Niu W J, Li S S, Wu H J and Wang Y J 2020 Ecological operation of cascade hydropower reservoirs by elite-guide gravitational search algorithm with Levy flight local search and mutation J. Hydrod. 581
[67] Du Y, Li Y, Duan C, Gou H B and Jiang L 2021 Adjustable uncertainty set constrained unit commitment with operation risk reduced through demand response IEEE Trans. Ind. Inform. 17 1154–65
[68] Li Y, Ni Z, Zhao T, Zhong T, Liu Y, Wu L and Zhao Y 2020 Supply function game based energy management between electric vehicle charging stations and electricity distribution system considering quality of service IEEE Trans. Ind. Appl. 56 5932–43