IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 February 2024, accepted 9 March 2024, date of publication 18 March 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3377097

RESEARCH ARTICLE

The Power Load Forecasting Model of Combined SaDE-ELM and FA-CAWOA-SVM Based on CSSA

ZUOXUN WANG[1,2], YANGYANG KU[2], AND JIAN LIU[2]
[1]School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China
[2]School of Information and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China
Corresponding author: Zuoxun Wang (wangzuoxun@126.com)

ABSTRACT The security and stability of the power grid are directly affected by the accuracy of power load forecasting. Additionally, it plays an important role in power system planning. In order to enhance forecasting accuracy, a combined forecasting model is proposed in this paper. Firstly, preprocessing of the original data is conducted through improved singular spectrum analysis. Subsequently, load data prediction is carried out by the adaptive evolutionary extreme learning machine (SaDE-ELM). Additionally, load data prediction is performed using the support vector machine model(SVM), which is optimized by the chaotic adaptive whale algorithm based on the firefly disturbance strategy (FA-CAWOA-LSSVM). In the final step, the weight coefficients of the two prediction models are calculated by the chaotic sparrow search algorithm (CSSA). The load prediction results are obtained through the weighted summation of the two predictions. Superior performance is demonstrated by the combined prediction model compared with other single prediction models. The data preprocessing method, based on improved singular spectrum analysis, effectively enhances prediction accuracy.

INDEX TERMS Singular spectrum analysis, combinatorial forecasting model, power load forecasting.

I. INTRODUCTION

The development of the power system is crucially impacted by power load forecasting. Accurate load forecasts for different future times are relied upon for the stable operation of the power system. A key aspect of power load forecasting is represented by the analysis of factors affecting power load. A mathematical model representing the intrinsic relationship between influencing factors and load changes is established. The power development trend and load data for the coming period are forecasted additionally.

At present, many achievements have been made in the direction of power load forecasting by experts and scholars at home and abroad. Common forecasting models, such as artificial neural networks(ANN), support vector machines(SVM), extreme learning machines(ELM), and so on, are employed.

A dynamic neural network for daily load prediction was proposed by Literature [1]. Improved accuracy and efficiency were achieved compared to traditional methods widely used in this field. A short-term load forecasting model based on a Bayesian Neural Network (BNN) learned by the Hybrid Monte Carlo Algorithm (HMC) was introduced by Literature [2]. The results indicated better performance than traditional networks and a strong generalization ability, addressing the overfitting problem. An architectural model for power load forecasting based on ANN was presented by Literature [3]. Short-term power loads are forecasted, offering a simple structure with a certain degree of reliability in forecasting accuracy. A regression model using large-scale linear programming support vector machines for short-term power load forecasting was proposed by Literature [4]. Experimental results show a relatively small prediction error. A recursive support vector machine based on genetic algorithm (RSVMG) for power load forecasting was introduced by Literature [5]. The penalty factor and kernel

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

parameters of the SVM are determined by genetic algorithm (GA). Empirical results show superiority over SVM, ANN, and regression models. A power load forecasting model based on ACF-GCLSSVM was proposed by Literature [6]. ACF is used to select useful feature vectors, and then Gray Wolf Optimization Algorithm (GWO) and cross-validation (CV) are employed to optimize parameters of least squares support vector machines (LSSVM). Validation demonstrates its effectiveness in significantly improving forecasting accuracy."

Swarm intelligence optimization algorithms have been developed. Some optimization algorithms are utilized to optimize the parameters of prediction models, such as ANN, SVM, and ELM by experts and scholars. Common swarm intelligence optimization algorithms, including particle swarm algorithm (PSO) [7], ant colony optimization algorithm (ACO) [8], fruit fly optimization algorithm (FOA) [9], whale optimization algorithm (WOA) [10], sparrow search algorithm (SSA) [11], Harris Hawk Optimization Algorithm (HHO) [12], and other new optimization algorithms with good performance, have been applied.

A prediction model was proposed in Literature [13], utilizing a PSO to optimize the Gaussian kernel function and penalty factor C in the LSSVM. The model was then employed to forecast future short-term loads, demonstrating good convergence, prediction accuracy, and training speed. A power load forecasting model based on SVM and ACO was introduced in Literature [14]. A feature selection mechanism based on ACO optimization was established to select features for the original data, and future load data were predicted with the SVM. Compared with single SVM and BP neural network (BPNN) in short-term load forecasting, this model exhibited much higher prediction accuracy. In Literature [15], a short-term power load forecasting model that combines Elman neural network (ENN) and PSO was proposed. It mainly searched for the optimal learning rate of ENN with the PSO algorithm, showing certain effectiveness and reliability in short-term load forecasting. A wavelet least squares support vector machine (W-LSSVM) model using the Improved Fruit Fly Optimization Algorithm (IFOA) was introduced in Literature [16]. The model utilized the Cauchy variation process to make fruit fly individual variation, replaced the Gaussian kernel function of the LSSVM with the wavelet kernel function, and optimized the W-LSSVM with IFOA to seek the optimal parameter and achieve prediction accuracy. The results verified the model's strong validity and feasibility in medium- and long-term power load forecasting. In Literature [17], the initial weights and thresholds of an ELM were proposed to be optimized with the Chaotic Sparrow Search Algorithm (CSSA) and improved by the Firefly Algorithm (FA). It overcame early local convergence with the chaotic strategy and enhanced the algorithm's global optimality-seeking ability with the Firefly perturbation strategy. The results showed that the model was superior to some other single models, exhibiting better accuracy and stability.

Various characteristic factors affect the load, and these factors are characterized by randomness and volatility. The importance of load data preprocessing for improving prediction accuracy is emphasized. Currently, commonly used preprocessing techniques include wavelet transform(WT) [18], empirical modal decomposition(EMD) [19], variational modal decomposition(VMD) [20], singular spectrum analysis [21], and more.

With the development of society, load data is increasingly influenced by complex factors. The intrinsic information of the data is difficult to fully uncover with the traditional single model. Its prediction performance is limited by some inherent shortcomings. However, the drawbacks of a single prediction method are compensated for by the combination of prediction models. Better stability and prediction accuracy are achieved by this approach.

In literature [22], a combined prediction model, based on three single prediction models, was proposed. The raw data underwent preprocessing using VMD-singular spectrum analysis. Subsequently, the ELM prediction model (CAWOA-ELM), the LSSVM prediction model (EOBL-CSSA-LSSVM), and the Elman prediction model were individually applied to predict future load data. Finally, the simulated annealing algorithm was used to calculate weighting coefficients for the three prediction models, and their weighted sum was obtained for the prediction results. The experiments demonstrated that the shortcomings of single models were overcome by the combined forecasting model, which exhibited superior forecasting performance. In literature [23], a power load forecasting model based on LSSVM, ELM, and a generalized regression neural network (GRNN) was proposed. The weight coefficients for the three forecasting models were calculated using the WOA, and a weighted summation of their results was performed. Experimental comparisons indicated that the model showed good forecasting performance in both short-term load forecasting and electricity price forecasting.

In conclusion, combined forecasting models outperform single forecasting models, and they are played by a crucial role with significant development potential in present and future electric load forecasting.

In this paper, a new combined prediction model is proposed, which is mainly composed of two single models, SaDE-ELM and FA-CAWOA-LSSVM, in which SaDE-ELM is suitable for large-scale samples as well as meeting the real-time prediction, while FA-CAWOA-LSSVM is suitable for dealing with high-dimensional data with strong generalization ability. They complement each other in load forecasting and can meet the forecasting requirements of various types of data and scenarios. Therefore, this paper uses these two single models to form a combined prediction model, which combines the advantages of the two single models, has better robustness and generalization performance, and has better applicability in the field of load forecasting.

The combined prediction model proposed in this paper is described as follows: firstly, we use the improved singular spectrum analysis method to reduce the noise of the original data, then we use SaDE-ELM and FA-CAWOA-LSSVM to predict the future load values, and finally, we use CSSA to calculate the weight coefficients of the two models, and then we weight and sum the prediction results of the two models. We use the real load data of a region as experimental data and compare the five single models with this model, and the results show that this model has better prediction accuracy and stability.

II. MATERIALS AND METHODS

A. SINGULAR SPECTRUM ANALYSIS

Singular spectrum analysis is considered a popular and effective tool for studying nonlinear time series and extracting useful information. A one-dimensional time series is considered, and its trajectory matrix is shown in equation (1).

$$H = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \cdots & \vdots \\ x_M & x_1 & \cdots & x_N \end{bmatrix} \tag{1}$$

where N is the sequence length and M is the window length, $K = N - M + 1$, calculate HH^T and perform singular value decomposition to get M singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0$, then extract the useful information, i.e., select the first r larger singular values. Finally, the signal is reconstructed.

B. A NOISE REDUCTION METHOD BASED ON IMPROVED SINGULAR SPECTRUM ANALYSIS

Singular spectrum analysis is considered a popular and effective tool for studying nonlinear time series and extracting useful information. Its mathematical basis is matrix decomposition theory, which mainly builds the original data into Toeplitz-type matrices, and then performs singular value decomposition (SVD) on the matrices to obtain the eigenvectors and singular values, and then reconstructs the sequences by selecting the singular values, which ultimately achieves the denoising purpose. The optimization focuses on calculating the fourth root of the variance of the singular values [24].

$$P(n) = \left(\frac{1}{n} \sum_{i=1}^{n} (\sigma_i - \bar{\sigma})^4 \right)^{\frac{1}{4}} \tag{2}$$

where n is the number of singular values, σ_i is the ith singular value, and $\bar{\sigma}$ is the average of all singular values.

In order for the matrix to better resolve the information in the time series and suppress the noise, $P(n)$ should be maximized. Therefore, the maximum window length is shown in (3).

$$n_{best} = \arg\max [P(n)] \tag{3}$$

A differential spectral threshold method is employed to choose the appropriate singular value for signal reconstruction. The difference spectrum expression is given by (4).

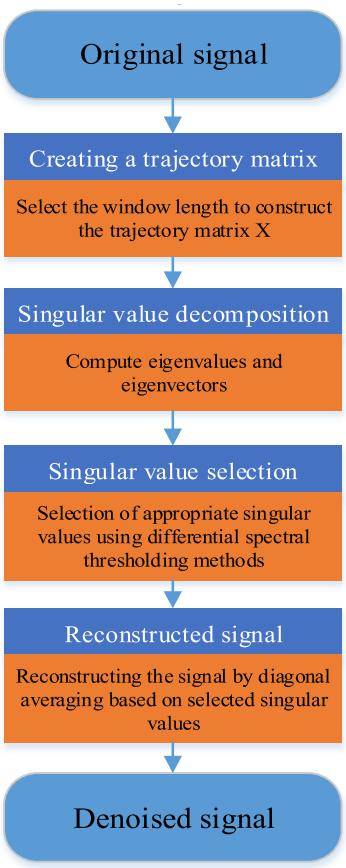


FIGURE 1. Flowchart of the algorithm to improve the singular spectrum analysis method.

$$b_i = \sigma_i - \sigma_{i+1} (i = 1, 2, \cdots, r-1) \tag{4}$$

where σ_i is the singular value and r is the number of singular values.

The singular value threshold is:

$$T = \rho(M - m) + m \tag{5}$$

where ρ is the threshold weighting coefficient in the interval [0, 1], M is the maximum value of the difference spectrum, and m is the minimum value of the difference spectrum.

We use the threshold T to extract useful information from the time series by selecting the singular values.

$$\delta = i | b_i < T \text{ and } b_{i+1} < T \text{ and } b_{i+2} < T \quad \text{i} = 1, 2, \cdots, \text{r-3} \tag{6}$$

where T is the singular value threshold, the singular values are ranked from largest to smallest, and the first δ singular values are selected to reconstruct the time series.

The detailed steps of the singular spectrum noise reduction method are summarized as follows:

1) The only hyperparameter in the improved singularity spectrum, the weight coefficient, is defined.

2) The time series is input $f(t)$, the appropriate window is selected using (2) and (3), and then the trajectory matrix (1) is built.

3) The eigenvalues and eigenvectors of HH^T are calculated, and then the left singular vector U and the right singular vector V are calculated.

4) The threshold of the singular value is calculated using (4) and (5), the appropriate singular value is selected using (6), and then the useful information is extracted.

5) Matrix reconstruction is performed with the extracted singular values, the reconstructed singular matrix is R, and the diagonally reconstructed signals are averaged to finally obtain the time series after noise removal.

The algorithmic flow of the improved singular spectrum analysis method is shown in Fig. 1.

C. ELM

The initial weights and thresholds of ELM, a single hidden-layer neural network, are randomly determined, and the weights of the output layer are computed using inverse matrix theory. The derivation theory of the ELM network is as follows.

The output of the hidden layer of the ELM network $h_i(x)$ is:

$$h_i(x) = g(w_i x + b_i) \tag{7}$$

where x is the network input, w_i is the input layer weight, b_i is the input layer threshold, and $g(\cdot)$ is the activation function.

The output f_L of the output layer of the ELM network is given by:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = H(x)\beta \tag{8}$$

where $\beta = [\beta_1, \cdots \beta_L]^T$ is the output layer weight matrix, $H(x) = [h_1(x), \cdots, h_L(x)]$ is the hidden layer output matrix, and L is the number of nodes in the hidden layer.

During network training, w_i and b_i are determined randomly, and the hidden layer output is obtained from (7), where $H = [h_1, h_2, \cdots, h_L]$.

$$T = H\beta \tag{9}$$

where, T is the network output and is the output layer weight matrix. Then:

$$\hat{\beta} = H^{+}T \tag{10}$$

where $\hat{\beta}$ is the optimal output weight of the network and H^+ is the generalized inverse matrix of H.

D. SADE-ELM

The optimization of input weights and thresholds of the extreme learning machine by the SaDE-ELM is achieved through the use of an adaptive differential evolutionary algorithm for output layer weights [25]. The introduction of the SaDE-ELM is characterized by the following steps.

1) The population is initialized by randomly generating parameters for the hidden layer nodes of the extreme learning machine, forming the first generation of N individual vectors.

$$\theta_{k,G} = [a^T_{1,(k,G)}, \cdots, a^T_{L,(k,G)}, b_{1,(k,G)}, \cdots, b_{L,(k,G)}] \tag{11}$$

where a_j and b_j $(j = 1, \cdots, L)$ are randomly generated input weights and thresholds, L represents the number of hidden layer nodes of the extreme learning machine, G represents the number of times the population has evolved, and $k = 1, 2, \cdots, N$.

2) The fitness function is determined, and the network output weights β can be obtained from (10). The calculation of the fitness function is performed using the following equation.

$$fitness_{k,G} = \sqrt{\frac{\sum_{i=1}^{N} \left\| \sum_{j=1}^{L} \beta_j g\left(a_{j,(k,G)}, b_{j,(k,G)}, x_i\right) - y_i \right\|}{N}} \tag{12}$$

where x_i is the input to the training set and y_i is the actual target value of the training set.

3) Mutation and crossover are performed, and for each target vector in the current generation, the test vector generation strategy is selected from the candidate pool constructed from the four strategies.

Use $p_{l,G}$ to represent the probability of strategy 1 being selected in the G generation $(l = 1, 2, 3, 4)$, set the learning period to LP, and let $p_{l,G}$ be updated according to the following rules.

If $G > LP$ and each strategy has an equal probability of being chosen, then $p_{l,G} = 1/4$

If $G > LP$, then $p_{l,G} = S_{t,G} \Big/ \sum_{l=1}^{4} S_{t,G}$, and $S_{t,G} = \sum_{g=G-LP}^{G-l} nS_{l,g} \Big/ \left(\sum_{g=G-LP}^{G-l} nS_{l,g} + \sum_{g=G-LP}^{G-l} nf_{l,g}\right) + \varepsilon$

where $nS_{l,g}$ and $nf_{l,g}$, respectively, represent the number of experimental vectors generated by strategy l of the generation g population entering the next generation evolutionary population and the number discarded, $\varepsilon = 0.001$, to prevent a zero success rate.

Select the variant strategy with the highest probability in the strategy pool to generate the variant vector $v_{k,G}$, generate the experimental vector after crossover as $u_{k,G}$, then calculate the fitness values of the experimental vector and the original vector, and select the vector with the best fitness value as the individual of generation $G + 1$.

The generation rule for the experimental vect $u_{k,G}$ is:

$$u^j_{k,G} = \begin{cases} v^j_{k,G}, & if\ (rand_j \le CR) or\ (j = j_{rand}) \\ \theta^j_{k,G}, & otherwise \end{cases} \tag{13}$$

where CR is the cross-factor of the control perturbation parameter obeying a normal distribution with values in [0, 1], $rand_j$ is a random number in [0, 1], and j_{rand} is a random positive integer in [1, G].

4) The result obtained in step 3 is used as the input weights and hidden layer bias of the ELM.

5) A portion from the load data dataset is selected as a training set, and its network optimal input weights and biases are computed using steps 1-4.

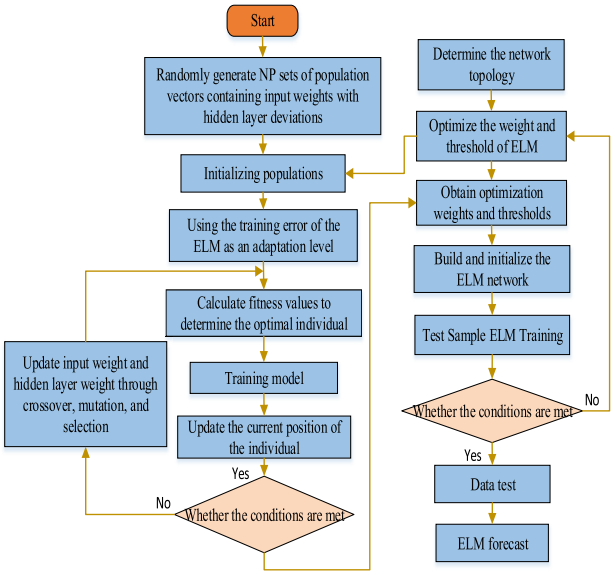


FIGURE 2. SaDE-ELM prediction model.

6) The remaining data sequences are used as a test set, and their corresponding results are output based on the weight update rule.

The intuitive process of the SaDE-ELM prediction model is depicted in Fig 2.

E. LSSVM MODEL

The model known as the LSSVM is used for classification and regression, with its basic idea being to infer the corresponding output value y from any input sample x. For a given set of training data samples, set (x_i, y_i), where $i = 1, 2, 3, \cdots, l$, Its prediction model expression is as follows:

$$y = f(x) = \omega\phi(x) + b \tag{14}$$

where ω is the weights; b is the bias term, taken as a constant; and $\phi(x)$ is the kernel function, which represents a nonlinear mapping from a low-dimensional space to a high dimensional space.

The expression and constraints of its optimization objective are formulated as follows.

$$\begin{aligned} \min\ & \frac{1}{2}\|\omega\| + C\sum\nolimits_{i=1}^{N}(\delta_i + \delta_i') \\ s.t.\quad & y_i - \left(\omega^T x_i + b\right) \le \varepsilon + \delta_i \\ & \left(\omega^T x_i + b\right) - y_i \le \varepsilon + \delta_i' \\ & \delta_i, \delta_i' \ge 0 \end{aligned} \tag{15}$$

Its dyadic form is expressed as follows:

$$\begin{aligned} \max\ & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i - \alpha_i'\right)\left(\alpha_j - \alpha_j'\right)x_i^T x_j \\ & -\sum_{i=1}^{N}\left[\left(\varepsilon - y_i\right)\alpha_i + \left(\varepsilon + y_i\right)\alpha_i'\right] \end{aligned}$$

TABLE 1. Kernel functions of support vector machines.

Name	Function
Linear kernel	$\kappa(x_i, x_j) = x_i^T x_j$
Polynomial kernel	$\kappa(x_i, x_j) = \left(x_i^T x_j\right)^d$
Gaussian kernel	$\kappa(x_i, x_j) = \exp(-\left\|x_i - x_j\right\|^2 / 2\sigma^2)$
Laplace kernel	$\kappa(x_i, x_j) = \exp(-\left\|x_i - x_j\right\| / \sigma)$
Sigmoid kernel	$\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$

$$\begin{aligned} s.t.\quad & \sum_{i=1}^{N}\left(\alpha_i - \alpha_i'\right) = 0 \\ & 0 \le \alpha_i, \alpha_i' \le C, i = 1, 2, \cdots, N \end{aligned} \tag{16}$$

where C is the penalty factor, δ_i and δ_i' are the relaxation factors and ε is the loss function.

In the prediction of nonlinear samples, the data is typically converted from low-dimensional to high-dimensional using a kernel function. The selection of the kernel function is crucial, and the commonly used kernel functions are shown in Table 1.

After determining the kernel function, the penalty factor C and the kernel parameter g are determined. In this paper, these two parameters are mainly optimized using the improved whale optimization algorithm.

F. WOA ALGORITHM

The feeding behavior of whale groups is simulated by the whale algorithm. It is a concise and easy-to-implement heuristic optimization algorithm. Loose requirements on the objective function conditions and less parameter control characterize the algorithm. However, it has the disadvantages of low solution accuracy and being prone to local optimization [26].

The position of each whale in the D-dimensional space is:

$$X = (x_1, x_2, \cdots, x_D) \tag{17}$$

1) SURROUNDING THE PREY

Swimming towards the optimal position, the position update equation is as follows:

$$X_i^{t+1} = X_{best}^t - A\left|C * X_{best}^t - X_i^t\right| \tag{18}$$

$$A = 2ar_1 - a \tag{19}$$

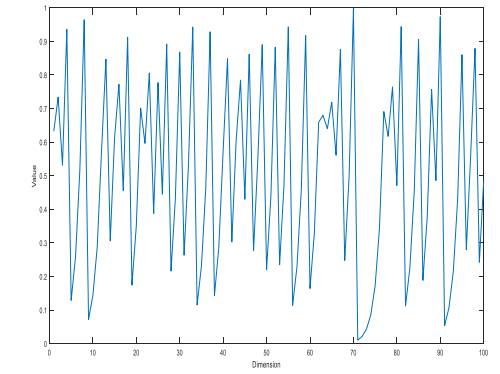
$$a = 2 - \frac{2t}{T} \tag{20}$$

$$C = 2r_2 \tag{21}$$

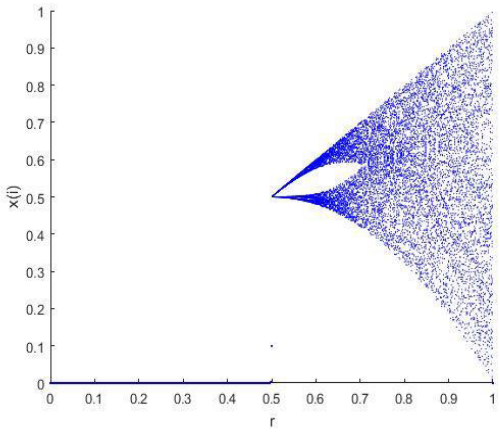
where X_{best}^t is the current optimal whale position, r_1, r_2 is a random number of [0, 1], t denotes the current number of iterations, and T denotes the maximum number of iterations.

Swimming towards the location of a random whale, the location of this whale is updated as follows:

$$X_i^{t+1} = X_{rand}^t - A\left|C * X_{rand}^t - X_i^t\right| \tag{22}$$



(a)



(b)

FIGURE 3. Tent chaotic mapping ((a) The Distribution Map of Tent Chaotic Maps; (b) Illustrations of Tent Chaotic Maps).

where X_{rand}^{t} is the position of a randomly selected whale in the current population. When $|A < 1|$, the whale chooses to swim toward the optimal individual. When $|A \geq 1|$, the whale chooses to swim toward the random individual.

2) BUBBLE NET

During this phase, the whale's position is updated with the following formula:

$$X_i^{t+1} = \left|X_{best}^{t} - X_i^{t}\right| * e^{bl} * \cos(2\pi l) + X_{best}^{t} \tag{23}$$

where b is a constant and l is a random number uniformly distributed in $[-1, 1]$.

Before each action, a strategy is chosen at random by the whale.

G. FA-CAWOA ALGORITHM

To address the issue of the algorithm easily falling into a local optimum, improvements to WOA are proposed in this paper.

1) TENT CHAOS MAPPING STRATEGY

Tent chaotic mapping has the characteristics of randomness, periodicity, and regularity, so we use the Tent chaotic mapping strategy for population initialization. The distribution map and bifurcation map of Tent chaotic mapping are shown in Fig. 3. The mathematical model of tent chaos is defined as follows:

$$z_{k+1} = \begin{cases} z_k / \beta & z_k \in (0, \beta] \\ (1 - z_k)/(1 - \beta) & z_k \in (\beta, 1] \end{cases} \tag{24}$$

2) NONLINEAR CONVERGENCE FACTOR

The original algorithm convergence factor an is linearly reduced from 2 to 0, so that the global search and local search of the algorithm are difficult to balance, and it is easy to make the algorithm fall into the local optimum when it converges. In this paper, we propose a nonlinear convergence factor with the following expression:

$$a = 1 + \cos(\mu \frac{t}{T}\pi + \varphi) \tag{25}$$

where t is the current iteration number, T is the maximum iteration number, and the two parameters μ and φ are selected as $\mu = 1$ and $\varphi = 0$ respectively.

3) ADAPTIVE WEIGHTS AND ADAPTIVE THRESHOLDS

Local optimization is often encountered by the whale algorithm in later iterations. To improve optimization seeking efficiency, adaptive weights and adaptive thresholds are introduced in this paper.

The following adaptive weight expression is proposed in this paper.

$$w = e^{-(t/T)^4} \tag{26}$$

The following optimization strategy is employed.

$$X_i^{t+1} = \omega X_{best}^{t} - A\left|C * X_{best}^{t} - X_i^{t}\right| \tag{27}$$

$$X_i^{t+1} = \omega X_{rand}^{t} - A\left|C * X_{rand}^{t} - X_i^{t}\right| \tag{28}$$

$$X_i^{t+1} = \left|X_{best}^{t} - X_i^{t}\right| * e^{bl} * \cos(2\pi l) + \omega X_{best}^{t} \tag{29}$$

In this paper, the constant threshold in the original whale algorithm is replaced by the adaptive threshold strategy. The expression is given as follows.

$$p' = 1 - \log_2\left(1 + \frac{2t}{T}\right) \tag{30}$$

4) FIREFLY OPTIMIZATION ALGORITHM

In the Firefly Algorithm (FA), individuals are attracted by the light emitted by other fireflies. Three hypothetical states exist: 1) mutual attraction among all fireflies; 2) attraction directly proportional to their individual brightness, causing fireflies to be actively drawn towards those brighter than themselves; brightness and attraction both decrease with distance. 3) If there are no fireflies brighter than a specific one, it will move randomly [27].

The relative fluorescence brightness of fireflies is given by:

$$I = I_0 * e^{-\gamma r_{i,j}} \tag{31}$$

where I_0 is the maximum fluorescence brightness of the firefly, and the better the value of the objective function, the

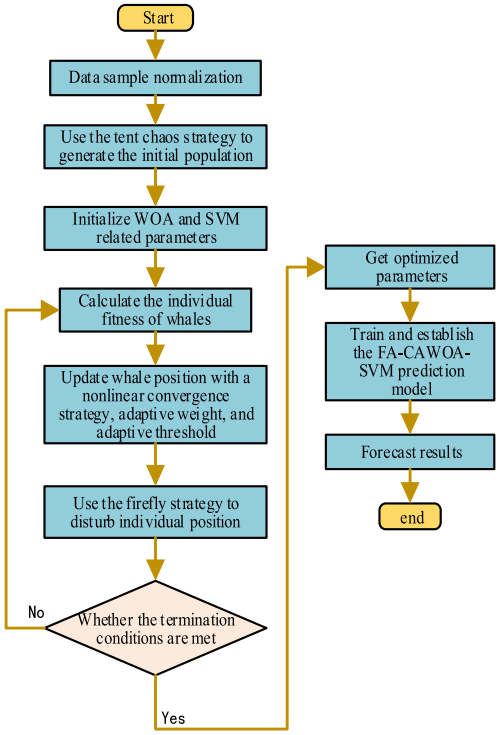


FIGURE 4. A prediction model based on FA-CAWOA-LSSVM.

higher its own brightness; γ is the light intensity absorption coefficient; and $r_{i,j}$ is the Euclidean distance between fireflies i and j.

The formula for the attraction of a firefly is given by:

$$\beta = \beta_0 * e^{-\gamma r_{i,j}^2} \tag{32}$$

where β_0 is the maximum degree of attraction.

The formula for updating the firefly location is given by:

$$x_i = x_i + \beta \times (x_j - x_i) + \alpha \times (rand - 0.5) \tag{33}$$

where x_i and x_j are the positions of fireflies i and j, respectively; $\alpha \in [0, 1]$ is the step factor; and rand is a random number on [0, 1] obeying a normal distribution.

H. FA-CAWOA BASED LSSVM PREDICTIVE MODELING

A prediction model is proposed in this paper, utilizing FA-CAWOA to optimize the penalty factor C and kernel parameter g of LSSVM. Improvements are made by FA-CAWOA, enhancing the prediction performance of the model based on WOA. The algorithmic flow of the model is outlined as follows.

1) The raw data is normalized, and a dataset is generated, divided into a test set and a training set.

2) The population is initialized, and tent chaotic mapping is employed to generate an initial population spread globally. The iteration number of the whale algorithm, the number of the initial population, the parameter of the nonlinear convergence factor, the kernel parameter C of the support vector machine, and the range of the penalty factor g are set.

3) The optimal weight value is found. The weight factors ω_1 and ω_2 of the two prediction models are determined using CSSA.

4) The convergence factor, weight, and threshold of the whale algorithm are updated using the nonlinear convergence factor, adaptive weight, and adaptive threshold formulas.

5) The individual whale position is updated according to the whale algorithm position update strategy.

6) The position of individual whales is perturbed using the firefly perturbation strategy to prevent the algorithm from falling into a local optimum.

7) It is judged whether the stopping condition is satisfied according to the iteration value and accuracy requirement; if the condition is satisfied, proceed to step 8; otherwise, go to step 3.

8) The optimal value is obtained, and the optimized parameters C and g are used to build the FA-CAWOA-SVM prediction model.

Fig. 4 displays the SVM prediction model based on FA-CAWOA.

I. CSSA ALGORITHM

The sparrow search algorithm (SSA) is a newly introduced swarm intelligence optimization algorithm. Strong optimization ability and fast convergence speed are possessed by it. However, a drawback exists in its susceptibility to falling into local optima [11]. In this paper, the population is initialized using tent chaotic mapping to enhance the algorithm's global optimality seeking ability.

Among sparrows, discoverers and followers are found. The responsibility of searching for the food source lies with the discoverer. The followers are provided with the location of the food source by the discoverer. The location update expression is as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\dfrac{-i}{\alpha \cdot M}\right) & R_2 < ST \\ X_{i,j}^{t} + Q \cdot L & R_2 \geq ST \end{cases} \tag{34}$$

where t is the current number of iterations, X_{ij}^{t} is the position of the i-th sparrow in the j-th dimension, M is the maximum number of iterations, α is a [0, 1] random number, Q is a random number obeying the normal distribution, $R_2 \in [0, 1]$ and $ST \in [0.5, 1]$ are respectively are the early warning value and safety value, and L is a $1 \times j$ matrix with values all 1.

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{x_{wj}^{t} - x_{ij}^{t}}{i^2}\right) & i > \dfrac{n}{2} \\ X_{p}^{t+1} + \left|X_{ij}^{t} - X_{pj}^{t+1}\right| \cdot A^{+} \cdot L & other \end{cases} \tag{35}$$

where X_{pj} is the $t+1$ iteration optimal position, X_{wj} is the t iteration global worst position, A is a $1 \times j$ matrix of -1 or 1, and $A^{+} = A^{T}(AA^{T})^{-1}$.

In each generation, vigilantes are assigned to check the foraging area for danger. An alarm is signaled when danger is encountered. Their position is moved using the position

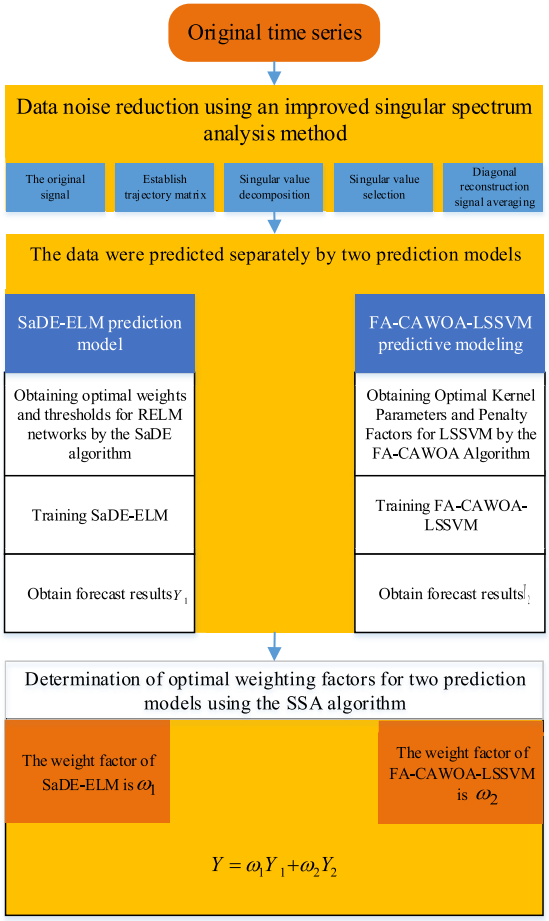


FIGURE 5. The specific process of the proposed combined prediction model in this paper.

update expression.

$$X_{ij}^{t} = \begin{cases} X_{bj}^{t} + \beta\left(X_{ij}^{t} - X_{bj}^{t}\right) & f_i > f_g \\ X_{ij}^{t} + k \cdot \left(\dfrac{X_{ij}^{t} - X_{bj}^{t}}{(f_i - f_w) + \xi}\right) & f_i = f_g \end{cases} \tag{36}$$

where X_{bj} is the global optimal position. β is the step control parameter. $k \in [0, 1]$. ξ is the very small real number. f_g and f_w are the current sparrow's optimal and worst fitness values, respectively.

J. CSSA-BASED PREDICTION MODEL OF SADE-ELM WITH FA-CAWOA-SVM

A combined prediction model is introduced in this paper. It integrates data noise reduction methods, swarm intelligence optimization algorithms, and machine learning algorithms. The specific steps of the proposed model are as follows:

1) Data preprocessing is performed, employing improved singular spectrum analysis for noise reduction of the raw data.

2) A forecasting model is established using SaDE-ELM and FA-CAWOA-LSSVM for the combined prediction model.

3) Optimal weights are determined using CSSA for the two prediction models.

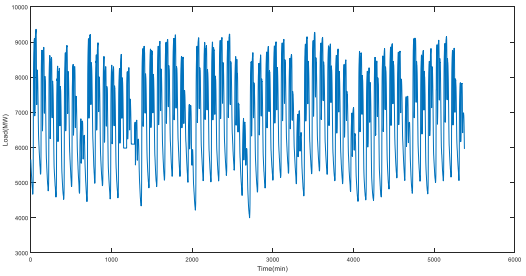


FIGURE 6. Raw data on electrical loads.

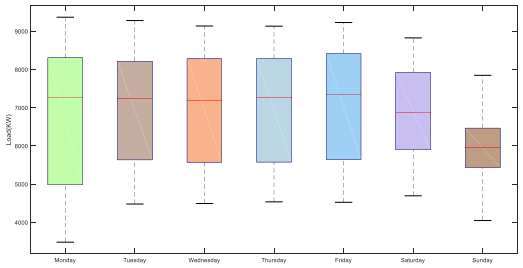


FIGURE 7. Box plot of seven sets of data subsets.

TABLE 2. Model performance evaluation criteria.

Metrics	The Formula
RMSE	$RMSE = \sqrt{\sum_{i=1}^{N}\left(x_i - y_i\right)^2 / N}$
MAPE	$MAPE = \sum_{i=1}^{N}\left\|\left(x_i - y_i\right)/x_i\right\| / N \times 100\%$
MAE	$MAE = \sum_{i=1}^{N}\left\|x_i - y_i\right\| / N$

4) Model prediction is conducted using SaDE-ELM and FA-CAWOA-LSSVM to obtain the prediction results Y_1 and Y_2, respectively.

5) The prediction results of the two models are waited and summed to obtain the final prediction result, denoted as $Y = \omega_1 Y_1 + \omega_2 Y_2$.

The prediction process of the prediction model in this paper is depicted in Fig. 5.

III. SIMULATION EXPERIMENT

A. DIVISION OF THE DATA SET

Eight weeks of actual electric power load data were selected for simulation experiments, spanning from July 6, 2018, 0:00, to August 30, 2018, 24:00, in a region. Data, totaling 96 sets daily and resulting in 5376 sets of experimental data, were measured every fifteen minutes during the day. The training set utilized the first seven weeks, comprising 4704 data sets, while the test set used the eighth week, consisting of 672 data sets. All data were categorized into 7 subsets based on Monday to Sunday types to build and verify the proposed prediction model separately for each subset. The raw data trend is depicted in Fig. 6, revealing a discernible pattern in data distribution. The weekly power consumption pattern is illustrated by box plots in Fig. 7, with the highest on Monday,

TABLE 3. Predictive modeling based on raw data (MAPE).

Weekly	GRNN	BP	FA-CAWOA-LSSVM	WOA-LSSVM	SaDE-ELM	Combinatorial model
Monday	2.97	5.11	2.63	2.95	2.61	2.08
Tuesday	2.26	4.87	1.76	1.76	2.67	1.67
Wednesda	4.60	5.17	3.56	3.90	2.32	1.95
Thursday	1.84	2.32	1.74	1.76	1.83	1.75
Friday	1.86	3.30	1.01	1.02	2.77	1.02
Saturday	2.45	4.17	1.89	1.97	2.96	1.87
Sunday	5.50	6.46	3.82	3.99	3.86	3.79

TABLE 4. Predictive modeling based on improved singular spectrum analysis (MAPE).

Weekly	GRNN	BP	FA-CAWOA-LSSVM	WOA-LSSVM	SaDE-ELM	Combinatorial model
Monday	2.60	5.09	2.39	2.63	2.17	1.71
Tuesday	2.07	4.19	1.19	1.19	1.55	1.18
Wednesda	4.15	4.99	2.51	2.95	1. 83	1.44
Thursday	1.91	2.59	1.72	1.73	1.75	1.72
Friday	1.74	3.11	0.87	1.01	2.15	0.86
Saturday	2.42	3.33	1.80	1.82	2.43	1.77
Sunday	4.69	5.48	3.24	3.56	3.18	3.15

TABLE 5. FA-CAWOA-LSSVM prediction accuracy and optimal parameters.

	Monday	Tuesday	Wednesday	Thursday
Gam	0.0012	12.4332	0.6307	0.0136
Sig	100	70.08	98.05	100
MAPE(%)	2.39	1.19	2.51	1.72
	Friday	Saturday	Sunday	
Gam	1.2384	0.0100	0.1035	
Sig	86.75	100	45.65	
MAPE(%)	0.87	1.80	3.24	

the lowest on Saturday and Sunday, and weekdays showing a more balanced consumption. It is evident that Monday's data is more scattered, while Sunday's data is more concentrated.

B. MODEL PERFORMANCE EVALUATION

To verify the reliability of the proposed model in this article, evaluation criteria for model accuracy, including root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), were selected. The expression of these criteria is shown in Table 2.

C. PRE-PROCESSING EXPERIMENTS FOR POWER LOAD DATA

Electricity load data is influenced by various factors, and there is significant noise in the data. The accuracy of load forecasting is greatly reduced when raw data is used directly. An improved method of singular spectrum analysis is proposed in this paper to preprocess raw load data, thereby enhancing prediction accuracy.

TABLE 6. Selection of hidden layer units in SaDE-ELM.

	Monday	Tuesday	Wednesday	Thursday
Number of hidden layer neurons	500	200	100	100
MAPE(%)	2.17	1.55	1.83	1.75
	Friday	Saturday	Sunday	
Number of hidden layer neurons	800	100	200	
MAPE(%)	2.15	2.43	3.18	

TABLE 7. Hidden layer cell selection for SaDE-ELM.

	Monday	Tuesday	Wednesday	Thursday
W1	0.4072	0.9884	0.4501	0.8767
W2	0.5929	0.0116	0.5499	0.1232
MAPE(%)	1.71	1.18	1.44	1.72
	Friday	Saturday	Sunday	
W1	0.9913	0.9988	0.2648	
W2	0.0087	0.0087	0.7362	
MAPE(%)	0.86	1.77	3.15	

To verify the effectiveness of the denoising method, a comparison is made with five single models: CRNN, BP, FA-CAWOA-LSSVM, WOA-LSSVM, and SaDE-ELM. MAPE is employed as the evaluation criterion for the models proposed in this paper.

Model prediction results based on the original data and improved singular spectrum analysis are displayed in Tables 3 and 4, respectively. Varied reductions in MAPEs for combined models, ranging from the highest at 29.34% to the lowest at 1.71%, with an average reduction of 16.13%, are revealed through comparative analysis. The average reduction of MAPEs for the four single models is 7.03%, 6.75%, 15.13%, 12.67%, and 18.17%. Significant noise reduction is observed in the Tuesday, Wednesday, and Sunday subsets among the seven data subsets. Evident improvement in prediction accuracy is observed for SaDE-ELM in the six models. This verifies that the model's prediction accuracy is effectively enhanced by the improved singular spectrum denoising method.

D. ANALYSIS OF FA-CAWOA-LSSVM PREDICTION RESULTS

The radial basis function is employed as the kernel function in some LSSVM parameter choices. The optimization of penalty factors and kernel parameters is carried out using FA-CAWOA. To test the effectiveness of the model in predicting detail components, it is applied to predict each of the seven data subsets. The optimization process for both parameters, using MAPE as the measure, is presented in Table 5.

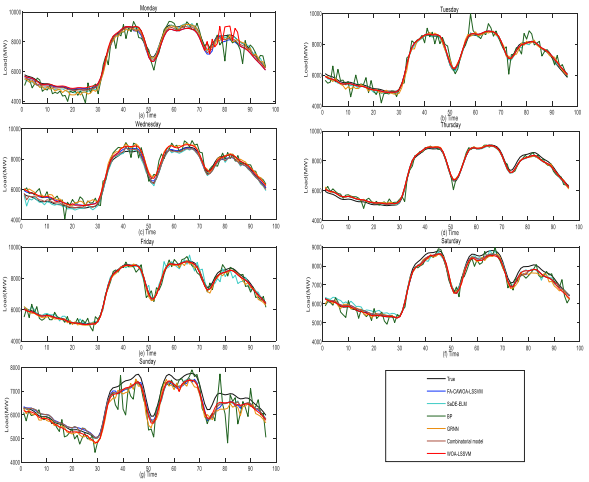


FIGURE 8. Actual values versus six model predictions in each 15-minute power load data set ((a) predictions in the Monday subset; (b) predictions in the Tuesday subset; (c) predictions in the Wednesday subset; (d) predictions in the Thursday subset; (e) predictions in the Friday subset; (f) predictions in the Saturday subset; and (g) predictions in the Week 7 subset).

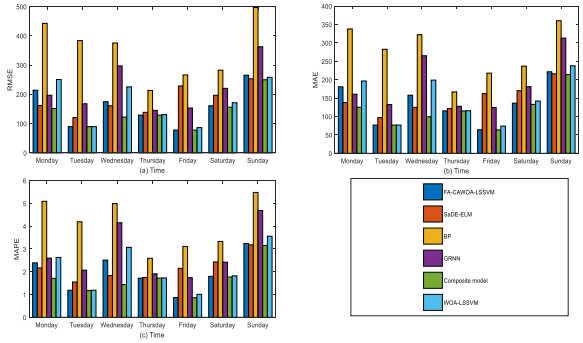


FIGURE 9. Evaluation metrics of different forecasting models ((a) comparison of RMSE of different forecasting models; (b) comparison of MAE of different forecasting models; (c) comparison of MAPE of different forecasting models).

E. ANALYSIS OF SADE-ELM PREDICTION RESULTS

The input layer weights and hidden layer bias of the ELM are determined randomly. Optimal prediction performance for the ELM can be hindered by this randomness. The optimal input layer weights and hidden layer bias of the ELM are found using SaDE. To test the effectiveness of the model for load prediction, seven subsets of data are predicted separately. For the selection of hidden layer units, as shown in Table 6, MAPE is used as a measure of the model.

F. CSSA OPTIMIZES THE WEIGHTING FACTORS OF TWO PREDICTION MODELS

The weight factor of each model in the proposed combination model is found using the CSSA algorithm. Different characteristics are possessed by each model; a poorer prediction effect may be observed in a single model. Various advantages of a single model are combined in the prediction effect of the combination model, resulting in a stronger prediction effect. During the training process, appropriate contribution weights for the two models are determined. Finally, the final result is obtained by combining the prediction results of the two models through the weights. For weighted optimization using CSSA, we set the warning value ST = 0.6; the proportion of discoverers PD = 0.7; and the proportion of joiners SD = 0.2. The weight of FA-CAWOA-LSSVM is defined as W1, and the weight of SaDE-ELM is defined as W2. The optimal weights of the two models are shown in Table 7.

IV. SIMULATION EXPERIMENT ANALYSIS

To effectively illustrate the good performance of the proposed model, predictions for each of the seven data subsets are made using the combined model and five single models. The five single models, namely GRNN, BP, FA-CAWOA-LSSVM, WOA-LSSVM, and SaDE-ELM, are employed. The comparison of model trends is displayed in Fig. 8, and the zoomed-in graph is included in Appendix.

The proposed combination model was compared with other single models in Fig. 7. It can be observed that the data is fitted closer to the true value by the combination model. Following closely are FA-CAWOA-LSSVM and WOA-LSSVM. Poorer prediction results from Friday to Sunday are exhibited by SaDE-ELM, but better results are observed from Monday to Thursday. In terms of accuracy, the MAPE of our combined model is superior to the other five models for all subsets except Thursday. For Thursday, our combined model is the same as FA-WOA-LSSVM.

To highlight the superiority of the combined prediction models proposed in this paper, the RMSE, MAE, and MAPE for the six prediction models in the seven data subsets are separately calculated. The results are presented in Table 8. The predictive performance of the models is visually compared in bar charts, as depicted in Figure 9. When combined, it is found that the proposed combined model outperforms other single models in the prediction results of all seven data sets. Its prediction stability is also found to be higher than that of other single models. Even in the face of poor data, such as the subset of data on Sundays, high stability is demonstrated by the combined prediction model, which combines the advantages of a single prediction model and ultimately achieves better prediction results than other single models.

It is found that the FA-CAWOA-LSSVM proposed in this paper yields better prediction results than the WOA-LSSVM. This improvement is achieved by employing the firefly perturbation strategy with chaotic initialization population to enhance the global optimality searching ability of the WOA. Furthermore, the local and global searching ability of the WOA is enhanced by using the nonlinear convergence factor with adaptive weights and adaptive thresholds. This enables the WOA to search for better LSSVM parameters, ultimately leading to an improvement in the prediction performance of the support vector machine. It is also observed that, except for Tuesday, the RMSE, MAE, and MAPE of FA-CAWOA-LSSVM are smaller than those of WOA-LSSVM, highlighting the superior prediction ability of FA-CAWOA-LSSVM.

TABLE 8. Comparison of the prediction accuracy of different prediction models in seven data sets.

Day	RMSE	MAE	MAPE
Monday			
FA-CAWOA-LSSVM	214.02	180.57	2.39
SaDE-ELM	161.48	137.83	2.17
BP	442.53	337.81	5.09
GRNN	196.73	160.44	2.60
combinatorial model	151.60	125.11	1.71
WOA-LSSVM	250.74	196.45	2.63
Tuesday			
FA-CAWOA-LSSVM	89.52	76.82	1.19
SaDE-ELM	120.22	97.07	1.55
BP	383.65	282.60	4.19
GRNN	167.66	132.54	2.07
combinatorial model	89.36	76.62	1.18
WOA-LSSVM	89.45	76.78	1.19
Wednesday			
FA-CAWOA-LSSVM	174.57	157.83	2.51
SaDE-ELM	160.79	124.87	1.83
BP	375.36	322.10	4.99
GRNN	297.03	264.99	4.15
combinatorial model	122.18	99.29	1.44
WOA-LSSVM	225.36	198.86	3.07
Thursday			
FA-CAWOA-LSSVM	128.84	115.34	1.72
SaDE-ELM	138.44	121.49	1.75
BP	213.18	166.81	2.59
GRNN	145.10	127.70	1.91
combinatorial model	128.61	114.96	1.72
WOA-LSSVM	130.68	115.69	1.73
Friday			
FA-CAWOA-LSSVM	77.74	63.99	0.87
SaDE-ELM	228.48	162.10	2.15
BP	266.26	217.76	3.11
GRNN	153.10	124.48	1.74
combinatorial model	77.63	63.56	0.86
WOA-LSSVM	86.07	73.90	1.01
Saturday			
FA-CAWOA-LSSVM	160.82	136.22	1.80
SaDE-ELM	196.69	170.06	2.43
BP	282.48	236.81	3.33
GRNN	220.36	180.96	2.42
combinatorial model	155.86	132.97	1.77
WOA-LSSVM	171.16	141.92	1.82
Sunday			
FA-CAWOA-LSSVM	235.38	221.42	3.24
SaDE-ELM	253.09	215.82	3.18
BP	496.72	360.11	5.48
GRNN	362.56	312.99	4.69
combinatorial model	249.74	213.67	3.15
WOA-LSSVM	257.91	237.77	3.56

V. CONCLUSION

A combined power load forecasting model is investigated in this paper, which integrates a power load preprocessing method, swarm intelligence optimization algorithm, and machine learning. The improved singular spectrum method is used to preprocess the raw data. Two distinct forecasting models, FA-CAWOA-LSSVM and SaDE-ELM, are proposed for the prediction of future loads. Weighting coefficients for the two models are calculated using the CSSA algorithm. The results of the two models are then weighted and summed for prediction. The advantages of individual forecasting models are leveraged by the combined forecasting model presented in this paper. Superior performance, compared to a single forecasting model, is demonstrated. The following conclusions are yielded by specific prediction experiments on seven data subsets

1) Upon comparing Tables 3 and 4, it is observed that load forecasting accuracy is effectively enhanced by the proposed improved singular spectrum data preprocessing method. An average MAPE reduction of 16.13% is achieved by the combined forecasting model, with notable reductions of 29.34% on Tuesday and 26.15% on Wednesday.

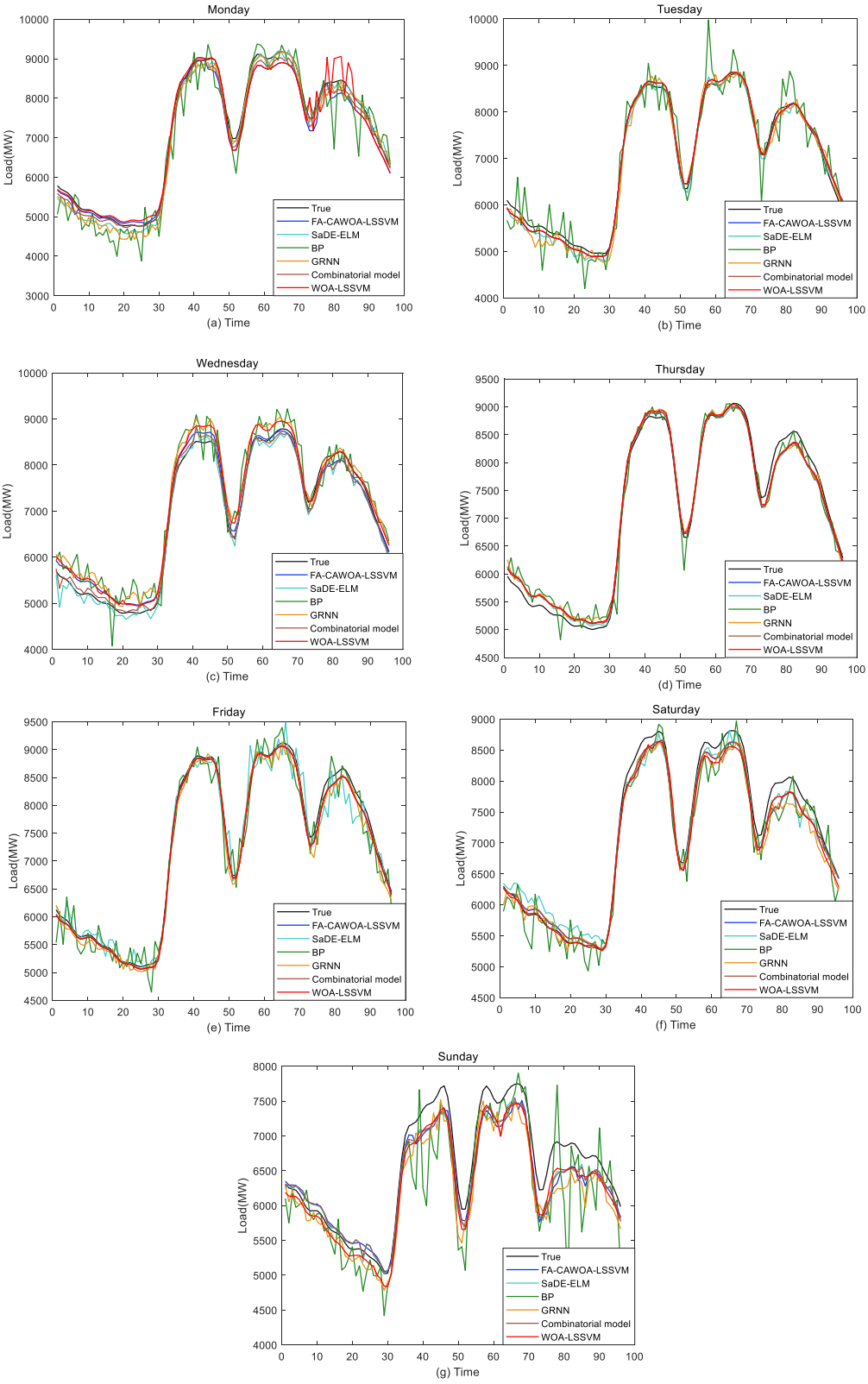
2) The comparison in Table 8 reveals that WOA-LSSVM is outperformed by FA-CAWOA-LSSVM in terms of RMSE, MAE, and MAPE. The average reductions are 7.33%, 7.61%, and 7.41%, respectively, demonstrating the superiority of the proposed FA-CAWOA-LSSVM.

3) The comparison in Table 8 reveals that higher forecasting accuracy is achieved by the combined forecasting model proposed in this paper compared to the single model. The RMSE decreases by at least 1.32% and at most 70.84% from week 1 to Sunday. The MAE decreases by at least 1% and at most 70.81%, and the MAPE decreases by at least 0.94% and at most 71.14%.

The combined prediction model proposed in this paper outperforms a single prediction model. The prediction accuracy is significantly enhanced by the data noise reduction method based on improved singular spectrum analysis. Notably, the prediction effect of FA-CAWOA-LSSVM is greatly improved, validating the reliability of the proposed FA-CAWOA optimization algorithm.

Although the combined prediction model proposed in this paper has high accuracy, it also has some limitations, such as 1) higher complexity of the operation, which requires larger computational resources and time. 2) model selection, if the single model selected is not accurate enough or cannot complement each other's advantages, it may affect the final prediction results. In future research, we can reduce the training time by setting the appropriate number of training times so that the model can be trained without affecting the prediction accuracy, and we can also choose a more appropriate single model or improve the hyperparameters of the model, which will in turn improve the prediction accuracy of the combined prediction model. In addition, this paper is limited by using fewer input features and ignoring factors such as the correlation between historical load data and holidays. In future power load forecasting, we will try to collect more feature data to obtain more accurate results.

APPENDIX


Monday
Load(MW)
(a) Time
Tuesday
(b) Time
Wednesday
(c) Time
Thursday
(d) Time
Friday
(e) Time
Saturday
(f) Time
Sunday
(g) Time
True
FA-CAWOA-LSSVM
SaDE-ELM
BP
GRNN
Combinatorial model
WOA-LSSVM

REFERENCES

[1] M. Mordjaoui, S. Haddad, A. Medoued, and A. Laouafi, "Electric load forecasting by using dynamic neural network," *Int. J. Hydrogen Energy*, vol. 42, no. 28, pp. 17655–17663, Jul. 2017.
[2] D.-X. Niu, H.-F. Shi, and D. D. Wu, "Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm," *Appl. Soft Comput.*, vol. 12, no. 6, pp. 1822–1827, Jun. 2012.
[3] L. Hernandez, C. Baladrón, J. Aguiar, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Short-term load forecasting for microgrids based on artificial neural networks," *Energies*, vol. 6, no. 3, pp. 1385–1408, Mar. 2013.
[4] P. Rivas-Perea, J. Cota-Ruiz, D. G. Chaparro, A. Q. Carreón, F. J. E. Aguilera, and J.-G. Rosiles, "Forecasting the demand of short-term electric power load with large-scale LP-SVR," *Smart Grid Renew. Energy*, vol. 4, no. 6, pp. 449–457, 2013.
[5] P.-F. Pai and W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electr. Power Syst. Res.*, vol. 74, no. 3, pp. 417–425, Jun. 2005.
[6] A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and least squares support vector machines," *Knowl.-Based Syst.*, vol. 163, pp. 159–173, Jan. 2019.
[7] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.
[8] C. Blum, "Ant colony optimization: Introduction and recent trends," *Phys. Life Rev.*, vol. 2, no. 4, pp. 353–373, Dec. 2005.
[9] X. Yuan, X. Dai, J. Zhao, and Q. He, "On a novel multi-swarm fruit fly optimization algorithm and its application," *Appl. Math. Comput.*, vol. 233, pp. 260–271, May 2014.
[10] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.
[11] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020.
[12] T. Dokeroglu and E. Sevinc, "An island parallel Harris hawks optimization algorithm," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18341–18368, Nov. 2022.
[13] S. Qiang and Y. Pu, "Short-term power load forecasting based on support vector machine and particle swarm optimization," *J. Algorithms Comput. Technol.*, vol. 13, Jan. 2019, Art. no. 174830181879706.
[14] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2531–2539, Mar. 2010.
[15] K. Xie, H. Yi, G. Hu, L. Li, and Z. Fan, "Short-term power load forecasting based on Elman neural network with particle swarm optimization," *Neurocomputing*, vol. 416, pp. 136–142, Nov. 2020.
[16] N. Dongxiao, M. Tiannan, and L. Bingyi, "Power load forecasting by wavelet least squares support vector machine with improved fruit fly optimization algorithm," *J. Combinat. Optim.*, vol. 33, no. 3, pp. 1122–1143, Apr. 2017.
[17] Z. Wang, X. Wang, C. Ma, and Z. Song, "A power load forecasting model based on FA-CSSA-ELM," *Math. Problems Eng.*, vol. 2021, pp. 1–14, Apr. 2021.
[18] Z. Wang and L. Xu, "Fault detection of the power system based on the chaotic neural network and wavelet transform," *Complexity*, vol. 2020, pp. 1–15, Dec. 2020.
[19] Z. Zhang and W.-C. Hong, "Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm," *Nonlinear Dyn.*, vol. 98, no. 2, pp. 1107–1136, Oct. 2019.
[20] S. Zhao, J. Zhou, Y. Liu, J. Zhang, and J. Cui, "Application of adaptive filtering based on variational mode decomposition for high-temperature electromagnetic acoustic transducer denoising," *Sensors*, vol. 22, no. 18, p. 7042, Sep. 2022.
[21] M. R. Müller, G. Gaio, E. M. Carreno, A. D. P. Lotufo, and L. A. Teixeira, "Electrical load forecasting in disaggregated levels using fuzzy ARTMAP artificial neural network and noise removal by singular spectrum analysis," *Social Netw. Appl. Sci.*, vol. 2, no. 7, p. 1218, Jul. 2020.
[22] X. Wang, X. Gao, Z. Wang, C. Ma, and Z. Song, "A combined model based on EOBL-CSSA-LSSVM for power load forecasting," *Symmetry*, vol. 13, no. 9, p. 1579, Aug. 2021.
[23] Z. Shang, Z. He, Y. Song, Y. Yang, L. Li, and Y. Chen, "A novel combined model for short-term electric load forecasting based on whale optimization algorithm," *Neural Process. Lett.*, vol. 52, no. 2, pp. 1207–1232, Oct. 2020.
[24] W. Xue, Y. Luo, Y. Yang, and Y. Huang, "Noise suppression for GPR data based on SVD of window-length-optimized Hankel matrix," *Sensors*, vol. 19, no. 17, p. 3807, Sep. 2019.
[25] J. Cao, Z. Lin, and G.-B. Huang, "Self-adaptive evolutionary extreme learning machine," *Neural Process. Lett.*, vol. 36, no. 3, pp. 285–305, Dec. 2012.
[26] I. Aljarah, H. Faris, and S. Mirjalili, "Optimizing connection weights in neural networks using the whale optimization algorithm," *Soft Comput.*, vol. 22, no. 1, pp. 1–15, Jan. 2018.
[27] V. Kumar and D. Kumar, "A systematic review on firefly algorithm: Past, present, and future," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 3269–3291, Jun. 2021.

ZUOXUN WANG received the B.E. degree in motor and control from Shandong University of Technology, in 1997, the M.E. degree in computer application technology from Northeast Electric Power University, in 2005, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, in 2015.

He is currently a Professor with the School of Information and Electronic Engineering, Shandong Technology and Business University. He is the author/coauthor of over 40 journal articles and conference papers. His research interests include intelligent robots, control theory, and control engineering.

YANGYANG KU received the B.Eng. degree in electrical engineering and automation from the Haojing College, Shaanxi University of Science and Technology, in 2020. He is currently pursuing the M.E. degree in electronic information with the Qilu University of Technology.

His research interest includes power load forecasting.

JIAN LIU received the B.Eng. degree in mechatronic engineering from the Jingchu University of Technology, in 2020. He is currently pursuing the M.E. degree in control science and engineering with the Qilu University of Technology.

His research interests include intelligent control and electro-hydraulic control of construction machinery.