

Short-Term Electricity Load Forecasting Based on GWO-VMD-ISSA-LSTM Model

Qian Li, Yuan Cao

Abstract—With the increasing proportion of renewable energy in the power system, short-term electricity load forecasting is crucial for optimizing grid operation and improving energy efficiency. In this paper, we propose a comprehensive model that integrates the Grey Wolf Optimizer (GWO), Variational Mode Decomposition (VMD), Improved Sparrow Search Algorithm (ISSA), and Long Short-Term Memory networks (LSTM) to enhance the precision of short-term electricity load forecasting. Firstly, the raw power load data are preprocessed using the VMD optimized by the GWO algorithm to adaptively decompose them into a collection of different frequency components in order to eliminate the limitations of the traditional methods in dealing with non-smooth signals. Secondly, the ISSA is used to optimize the LSTM parameters to further enhance the generalization ability and prediction accuracy of LSTM. Finally, the optimized LSTM is utilized to forecast each frequency component, resulting in the final forecasted value. The final experimental results indicate that the model outperforms the traditional LSTM and other benchmark models on the actual electricity dataset, indicating that the proposed GWO-VMD-ISSA-LSTM model has high practical value and potential for generalization.

Index Terms—electricity load forecasting, VMD, ISSA, LSTM.

I. INTRODUCTION

THE safe and stable operation of the power system hinges on accurate electricity load forecasting[1]. With the socio-economic development and the growth of people's demand for energy, the stability and reliability of power systems are facing unprecedented challenges. Especially in recent years, with the large-scale grid integration of renewable energy sources such as wind and solar, the uncertainty of the electricity load system has increased significantly[2]. Accurate short-term electricity load forecasting not only help power companies to rationalize their power generation schedules and reduce unnecessary energy wastage, but also facilitate the effective integration of renewable energy, therefore boosting the overall efficiency and profitability of the power system's operations[3][4].

Early short-term electricity load forecasting primarily relied on statistical methods, such as time series analysis and regression analysis. A widely employed method is the Autoregressive Integrated Moving Average (ARIMA) model. Reference[5]proposes an ARIMA-based short-term electricity forecasting model, which performs well in handling linear

time series data but its ability to capture nonlinear characteristics is limited. To cope with the seasonal variations, [6]used the Seasonal Autoregressive Integral Sliding Average Model (SARIMA), which further improves the prediction accuracy. However, these traditional methods have limitations in dealing with complex non-linear relationships and cannot produce excellent results for electricity load forecasting[7]. With the advancement of machine learning technology, an increasing number of studies have begun to explore its application in electric power load forecasting. Reference [8] proposes a short-term electricity load forecasting model based on Support Vector Machine (SVM), which significantly improves the forecasting accuracy by choosing a suitable kernel function and parameter optimization. Reference[9]explores the possibility of using Random Forest (RF) for electricity load forecasting, which performs well in dealing with high-dimensional data and non-linear relationships. Even though machine learning has solved many problems and produced better predictions, overfitting still occurs when the data dimensions are large. In recent years, deep learning techniques have received widespread attention due to their powerful non-linear fitting capabilities, and various deep learning models such as Artificial Neural Network (ANN)[10], Markov Chain (MC)[11], Extreme Learning Machine (ELM)[12], Random Forest (RF)[13], and Long Short-Term Memory networks (LSTM)[14]have been used for electricity load forecasting, and these deep learning models produced better predictions than traditional machine learning methods[15]. Reference [16] compares LSTM with other forecasting models and shows that LSTM models can achieve better results than other models in both long-term and short-term forecasting. However, a single method may encounter obstacles in the form of elevated error rates, substantial computational complexity, and diminished computational efficiency. Therefore, attaining a high level of prediction accuracy can pose a significant challenge[17]. In order to enhance the prediction accuracy further, many studies have begun to explore the combination of signal processing techniques and optimization algorithms. Several widely-applied signal processing methods encompass wavelet decomposition[18], Empirical Mode Decomposition (EMD)[19], Ensemble Empirical Mode Decomposition (EEMD)[20], as well as Variational Mode Decomposition (VMD)[21], and more. Experimental verification in [22]confirms that VMD is capable of effectively addressing the issue of modal overlap and exhibits greater stability compared to traditional signal decomposition techniques, and it is widely used in research related to electricity forecasting. In [23], the VMD technique is employed to decompose the multivariate load sequences within an integrated energy system. Prior to inputting these sequences into a deep learning fusion model for prediction, distinct feature sequences are individually constructed. This

Manuscript received December 10, 2024; revised March 13, 2025. This work was supported in part by the Doctoral Research Initiation Fund of Shandong University of Technology under Grant No. 417037.

Qian Li is a postgraduate student of School of Mathematics and Statistics, Shandong University of Technology, Zibo, China. (e-mail: 1978181085@qq.com).

Yuan Cao is an associate professor of School of Mathematics and Statistics, Shandong University of Technology, Zibo, China (corresponding author to provide phone:15653315613; e-mail: yuancs@sdut.edu.cn).

approach mitigates the risk of overfitting during the training phase and enhances the accuracy of the predictions.

Considering that the LSTM model has too many parameters, so if we use the method of setting parameters manually, it is likely to affect the final prediction results due to subjectivity. The incorporation of intelligent optimization algorithms can effectively address and overcome this limitation. Some of the commonly used algorithms include Particle Swarm Optimization (PSO)[24], Genetic Algorithm (GA)[25], Whale Optimization Algorithm (WOA)[26], Sparrow Search Algorithm (SSA)[27], the Grey Wolf Optimizer (GWO) algorithm[28]and so on. A VMD-SSA-LSTM algorithm has been proposed in [29]. By utilizing the VMD, the electricity load data is decomposed into modal functions exhibiting diverse characteristics and frequencies. Subsequently, these processed data are employed to train LSTM models, facilitated by the sparrow search algorithm. This integrated approach can be effectively utilized for short-term power load forecasting. Reference[30] proposed an Improved Sparrow Search algorithm (ISSA) based on the improved algorithm with higher optimization efficiency and accuracy when doing a short-term power prediction study. At the same time, considering that there are also key parameters when decomposing power sequence signals using VMD, such as the modal decomposition number K and the quadratic penalty coefficients , it has been shown in[31] that using the Grey Wolf Optimizer (GWO) algorithm to improve VMD can more adequately extract the sequence features, thus improving the stability and accuracy of the subsequent prediction model.

In summary, this paper introduces a novel methodology that integrates Variational Mode Decomposition (VMD) optimized through the Grey Wolf Optimizer (GWO) with Long Short-Term Memory networks (LSTM) enhanced by the Improved Sparrow Search Algorithm (ISSA), collectively termed as the GWO-VMD-ISSA-LSTM model. The objective of this hybrid approach is to elevate the precision of short-term electricity load forecasting and furnish robust data insights for the optimal dispatch of power systems. The subsequent sections delve into the technical specifics of the proposed method, the experimental setup, and a comprehensive analysis of its outcomes.

II. METHOD AND THEOREM

A. Grey Wolf Optimization Algorithm

The grey wolf optimization algorithm is a kind of intelligent optimization algorithm proposed by the famous scholars Mirjalili et al[32]by drawing on the process of wolf predation, which is widely used in all kinds of optimization models for parameter optimization because of the advantages of few initial parameters, high computational efficiency and good performance of optimization search. There are 4 different types of wolves in the grey wolf optimization algorithm model: the wolf responsible for the decision-making part of the hunting process is labelled as wolf; for the rest of the population, the wolves are labelled as , and wolves according to the population hierarchy. The calculation procedure is divided into three parts: encirclement, pursuit and attack. The corresponding algorithms are shown below. (1) Encirclement. In the hunting process, grey wolves firstly need to encircle

the target prey, and encircle it within the range circle of the wolf pack. The distance between the prey and the grey wolf is calculated by the formula:

$$D = |CX_p(t) - X(t)| \quad (1)$$

$$X(t + 1) = X_p(t) - AD \quad (2)$$

Where: t is the number of iterations of the algorithm; $X_p(t)$ is the location of the optimal solution of the algorithm; A and C are the coefficient factors of the algorithms, and the corresponding computational relationship between them is:

$$A = 2ar_1 - a \quad (3)$$

$$C = 2r_2 \quad (4)$$

where: a is the number that decreases monotonically from 2 to 0 as t increases; r_1 and r_2 are the random number in the interval 0 to 1.

(2) Pursuit. Having surrounded the prey, the β , δ and ω wolves will go to capture the target, led by the α wolves. Because all the wolves in the above process are changing randomly, we need to update the position of the corresponding prey according to the position of each wolf group, and the calculation expression is:

$$\begin{cases} D_\alpha = |C_1X_\alpha(t) - X(t)| \\ D_\beta = |C_2X_\beta(t) - X(t)| \\ D_\delta = |C_3X_\delta(t) - X(t)| \end{cases} \quad (5)$$

$$\begin{cases} X_1 = X_\alpha(t) - A_1D_\alpha \\ X_2 = X_\beta(t) - A_1D_\beta \\ X_3 = X_\delta(t) - A_1D_\delta \end{cases} \quad (6)$$

$$X_p(t + 1) = \frac{X_1 + X_2 + X_3}{3} \quad (7)$$

where D_α , D_β and D_δ denote the distances between α wolves, β wolves, δ wolves and other wolves, respectively. (3) Attack. The attack process of grey wolf optimization algorithm is the optimization process of the algorithm, which is accomplished by the incremental change of α in (3). Employing the GWO algorithm to optimize the parameters of VMD during the construction of the short-term electricity load forecasting model can mitigate the errors arising from empirical parameter settings.

B. Variational Modal Decomposition

VMD is a novel adaptive decomposition estimation method for non-smooth signals proposed by Dragomiretskiy [33]in 2014. The method uses a non-recursive approach to process the original signal, so that the modal functions obtained are characteristically different from each other, thus effectively circumventing the occurrence of modal aliasing. Compared with the traditional empirical modal decomposition (EMD) and ensemble empirical modal decomposition (EEMD), VMD not only addresses the modal aliasing issue present in EMD and EEMD, but also provides higher frequency resolution and more stable decomposition results. In addition, the adaptive bandwidth adjustment capability of the VMD helps preserve the physical characteristics of the signal while reducing noise interference. These features make VMD an effective tool more suitable for the analysis

of complex signals, so it is very suitable for non-stationary sequences. In this paper, after optimization with the GWO algorithm, the power load sequence is decomposed by VMD and then multiple modal functions with different frequency characteristics are obtained. Assuming that the original signal is decomposed into components, its constrained variational model is:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^K u_k = f(t) \end{cases} \quad (8)$$

where $f(t)$ is the original undecomposed signal; K is the number of decomposed modal components; j is the imaginary unit; $\delta(t)$ is the unit shock function; ∂_t is the derivative of $\delta(t)$; $\{u_k\} = \{u_1, u_2, \dots, u_K\}$ and $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_K\}$ are the k th modal component and the centre frequency, respectively. Adding the quadratic penalty factor and the Lagrange multiplication operator transforms Eq. (8) into an unconstrained variational model:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left[\lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right] \end{aligned} \quad (9)$$

where: L is the Lagrange operator; α is the quadratic penalty factor; $\lambda(t)$ is the variation function of the Lagrange operator. By employing the multiplier alternating direction method to update each modal component $\{u_k\}$ and its corresponding center frequency cyclically, the final modal component and center frequency can be expressed as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i=1, i \neq k}^K \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (10)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (11)$$

where: $\hat{f}(\omega)$, $\hat{u}_i^n(\omega)$, $\hat{\lambda}^n(\omega)$, $\hat{u}_k^n(\omega)$ are the Fourier transforms of $f(t)$, $u_i^n(t)$, $\lambda_i^n(t)$, $u_k^n(t)$ respectively; n is the number of iterations, and is the frequency. And $\hat{u}_k^{n+1}(\omega)$, ω_k^{n+1} are the updated first modal component and centre frequency, respectively.

The cycle stops when the set condition is satisfied, and its stopping condition is:

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1}(\omega) - \hat{u}_k^n(\omega)\|_2^2}{\|\hat{u}_k^n(\omega)\|_2^2} < \varepsilon \quad (12)$$

Where, ε is the convergence threshold.

By applying the inverse Fourier transform, the original electricity sequence is decomposed into k IMFs (Intrinsic Mode Function) components, allowing for adaptive segmentation of the original electricity signal in the frequency domain.

C. Improved Sparrow Search Algorithm

Sparrow Search Algorithm (SSA) is an optimization algorithm proposed by scholars drawing on the behavior of a flock of sparrows in search of food[34]. The sparrow population contains explorers and followers, firstly the explorers

in the population are used to find the location of the food, and then the followers follow the best-performing explorer to move the location until the population searches for the globally optimal solution. In this case, sparrows close to natural enemies need to move to a safe area to obtain food. Explorer locations are updated below:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \times \exp\left(-\frac{1}{\alpha \times it}\right) & \text{if } R_2 < ST \\ X_{i,j}^t + Q \times L & \text{if } R_2 \geq ST \end{cases} \quad (13)$$

where: t is the current number of iterations; it is the maximum number of iterations; ST is a predefined safety threshold; R_2 is the value of the warning signal emitted by individuals that have detected a predator; is the individual position information of the first sparrow in dimension; $\alpha \in (0, 1]$ is a random number; Q is a random number obeying a normal distribution; and L is a $1 \times d$ -dimensional matrix.

The follower positions are updated below:

$$\begin{aligned} X_{i,j}^{t+1} = & \begin{cases} Q \times \exp\left(\frac{X_{worst} - X_{i,j}^t}{i^2}\right) & \text{if } \iota > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1} + 1| A^+ L & \text{if } \iota \leq \frac{n}{2} \end{cases} \\ & A^+ = A^T (A A^T)^{-1} \end{aligned} \quad (14)$$

where: X_p is the position of the follower; X_{worst} is the current position with the lowest adaptation; Q denotes a $1 \times d$ -dimensional matrix in which each element is randomly assigned a value of 1 or -1.

When an individual within the population senses danger, those situated on the periphery swiftly relocate to a safer zone, whereas sparrows previously positioned in the center of the group will relocate randomly to establish a new population.

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta |X_{i,j}^t - X_{best}^t| & \text{if } f_i > f_g \\ X_{i,j}^t + K \frac{X_{i,j}^t - X_{best}^t}{(f_i - f_w) + \varepsilon} & \text{if } f_i = f_g \end{cases} \quad (15)$$

Where: X_{best}^t is the global optimal position of the sparrow population; β is the step control parameter, which is a random number that obeys a normal distribution with mean 0 variance 1; K is a random number of $[-1,1]$; f_i is the current individual fitness value; f_g and f_w are the current global optimal fitness value and global worst fitness value respectively; ε is the smallest constant.

The SSA boasts strong optimization capabilities, rapid convergence speed, and minimal parameter adjustment requirements. However, it is difficult to obtain a better search speed because the standard sparrow search algorithm uses the same position update formula for all the explorers' and followers' positions, which does not allow the most appropriate search behavior according to the characteristics of each individual. During the iteration of the algorithm, individual positions gradually cluster, making it difficult to maintain good population diversity and prone to the issue of being trapped in local extrema, which is hard to escape. Considering the analysis presented above, this paper introduces the following enhancements to the sparrow search algorithm:

1) A sinusoidal search strategy is added, which not only elevates the significance of promising individuals but also bolsters the algorithm's convergence rate while harmonizing its global and local search efficacies; 2) With the concept

of aggregation degree in biology, Cauchy variation is added to enhance the algorithm's population diversity when the aggregation degree is high, thereby aiding the algorithm in escaping from local optima.

1) *Sine Search Strategy*: Considering that only two individual types are distinguished in the standard SSA, explorers and followers, and in both types all individuals are updated according to the same position update formula regardless of their positional superiority, it is not possible to give a suitable update strategy based on their own position. To address this problem, this paper introduces a sinusoidal search strategy, which can be assigned weights with different values depending on the location of the individual. The sinusoidal search strategy dynamically adjusts the step size and direction of the search process through the periodic and oscillatory properties of the sinusoidal function, thus enhancing the algorithm's global exploration capability and local exploitation. The formula of sinusoidal search strategy is as follows:

$$w = w_{\min} + (w_{\max} - w_{\min}) \left(\sin \left(\left(\frac{f_i^t - f_{best}^t}{f_{worst}^t - f_{best}^t} + 1 \right) \frac{\pi}{2} + \pi \right) + 1 \right) \quad (16)$$

Where: w_{\min}, w_{\max} are the minimum and maximum values of the weight variation range, respectively; f_i^t is the fitness value of the i th sparrow in the t th iteration population; f_{best} is the optimal fitness value of the t th iteration population, and f_{worst}^t is the worst fitness value of the t th iteration population.

As the individual fitness value f_i^t approaches the optimal fitness value f_{best} , the weight W is relatively small and the algorithm continues to search near the current individual position. As the individual fitness value f_i^t approaches the worst fitness value f_{worst} , the weight W gradually increase to w_{\max} , and the algorithm starts searching far away from the current individual position.

By incorporating the W value derived from the sine search strategy into SSA, the updated formulae for the explorers and followers of the ISSA are as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \times \exp \left(-\frac{1}{\alpha \times it} \right) & \text{if } R_2 < ST \\ X_{i,j}^t + wQL & \text{if } R_2 \geq ST \end{cases} \quad (17)$$

$$X_{i,j}^{t+1} = \begin{cases} Q \times \exp \left(\frac{X_{worst}^t - X_{i,j}^t}{i^2} \right) & \text{if } i > \frac{n}{2} \\ X_p^{t+1} + w |X_{i,j}^t - X_p^{t+1} + 1| A^+ L & \text{if } i \leq \frac{n}{2} \end{cases} \quad (18)$$

$$A^+ = A^T (AA^T)^{-1}$$

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + w\beta |X_{i,j}^t - X_{best}^t| & \text{if } f_i > f_g \\ X_{i,j}^t + wK \frac{|X_{i,j}^t - X_{best}^t|}{(f_i - f_w) + \varepsilon} & \text{if } f_i = f_g \end{cases} \quad (19)$$

As can be seen from Fig. 1, the SSA with the addition of the sinusoidal search strategy approaches the global optimal solution faster in the early iterations and finds lower fitness values in multiple runs, with a final average optimal fitness value of 74.8031, which is a significant improvement over the average optimal fitness value of 419.387 for the conventional SSA.

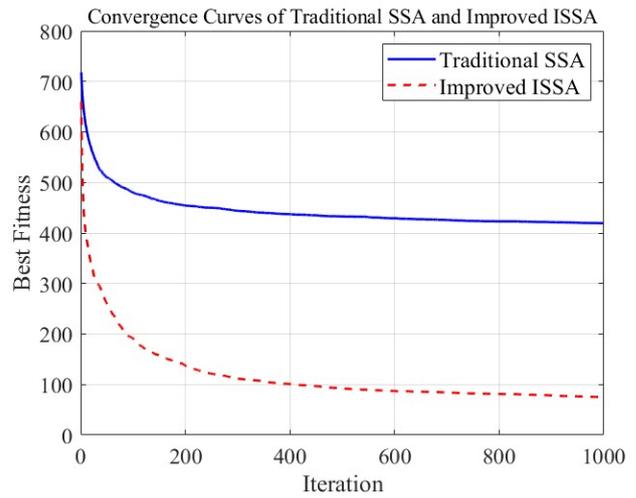


Fig. 1. Average Convergence Curves (1)

2) *Kersey Variation Strategy*: During the iterative convergence of the algorithm, individuals are highly prone to clustering excessively at a specific location, leading to a reduction in the algorithm's population diversity and subsequently trapping it in a local optimum. In order to address this issue, this paper proposes a metric that quantifies population aggregation in biology A :

$$A = \frac{\delta - \bar{x}}{\bar{x}^2} \quad (20)$$

Where: δ denotes the variance of sparrow population fitness; \bar{x} denotes the mean of sparrow population fitness. When $A \gg 0$, the population exhibits an aggregated state; when A tends to 0, the population exhibits a stochastic state. In order to avoid the emergence of an aggregated state at the beginning of the iteration, this paper uses the Cauchy variant for the population.

The global optimal solution is mutated using (21) when $t \leq \frac{it}{2}$ and if it is larger than a predefined threshold.

$$X = X_{best} + X_{best} \cdot \text{Cauchy}(0, 1) \quad (21)$$

The Cauchy variation strategy enhances the algorithm's capability to explore a broader range by boosting the population's diversity. This approach facilitates the discovery of more potential optimal solution areas during the initial iterations, thereby preventing the algorithm from overlooking the global optimal solution. The experimental results as shown in Fig. 2, show that the SSA with the introduction of the diversity variance treatment approaches the global optimal solution faster in the early iterations and finds lower fitness values in multiple runs, with the final average optimal fitness value of 417.6718 compared to 423.7532 for the conventional SSA, which is an improvement in the global search capability.

3) *ISSA Algorithm Flow*: The steps of ISSA incorporating the sinusoidal search strategy and the Kersey variation strategy are as follows:

- (1) Initialize population parameters such as number of iterations, population size, and so on;
- (2) Divide the population into explorers and followers according to appropriate proportions, set the threshold for

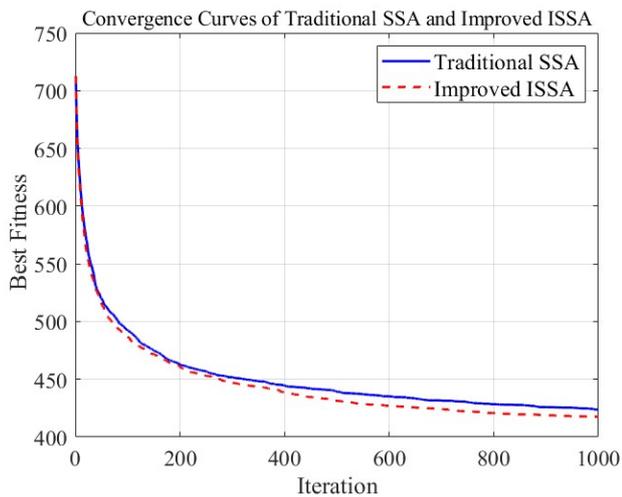


Fig. 2. Average Convergence Curves (2)

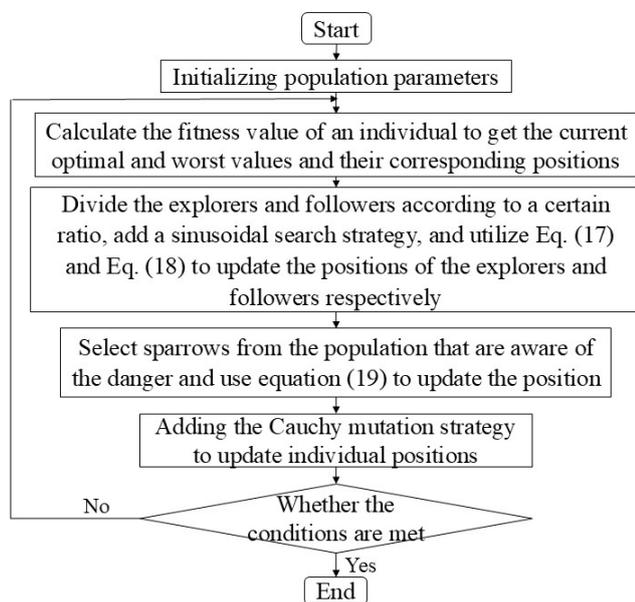


Fig. 3. Flowchart of ISSA Algorithm

sparrows to sound an alarm, and update the positions of individuals in the explorer and follower populations according to equations (17) and (18), respectively;

(3) Determine the number of individual sparrows in the population that are aware of the danger and update the position according to equation (19);

(4) Adding the Cauchy variation strategy, which utilizes Cauchy variation to disrupt the optimal position whenever the aggregation level of individuals attains a predefined threshold;

(5) Repeat steps (2) through (4) iteratively until the maximum number of iterations is reached or the optimization result meets the specified condition.

4) *Comparative Analysis of SSA and ISSA:* In this experiment, we test the convergence speed and stability of the standard SSA and the ISSA by using the Sphere function as the objective function, and compare and analyse the effect of the two algorithms through the generated parameter space plot and target space plot. The comparison chart is shown in Fig. 4.

The parameter space plot shows the distribution of the Sphere function in two dimensions. As illustrated in the figure, the Sphere function presents a bowl-shaped surface with its global minimum located at the origin (0,0). The target space plot shows the trend of the best solution found by SSA and ISSA in each iteration. By comparing the two curves, it becomes evident how the two algorithms differ in their convergence process and overall performance.

(1) Regarding the convergence speed, the ISSA curve drops rapidly in the early iterations, indicating that ISSA is able to quickly find regions close to the global optimal solution. This indicates that ISSA is more efficient in exploring the search space and can find better solutions in a shorter time.

(2) Regarding the quality of the final solution, the ISSA curves flatten out in the later stages and eventually converge to a very low value which is very close to the global optimum. This shows that ISSA not only converges faster but also finds results closer to the global optimal solution. Even though the SSA curve exhibits stabilization in the later phases, its ultimate convergence point lies notably above that of ISSA. This reveals that SSA falls short of ISSA in optimization precision, leading to a final solution of comparatively inferior quality.

In summary, by comparing the graphs of the parameter space and the target space, it is evident that ISSA not only converges more swiftly, enabling it to quickly locate regions close to the global optimal solution during early iterations, but also yields a final solution of higher quality that is able to approximate the global optimal solution more closely.

D. Long Short-Term Memory networks

Long Short-Term Memory networks (LSTM) are enhanced algorithms derived from Recurrent Neural Networks (RNNs) designed to address the issues of gradient vanishing and gradient explosion in RNNs. They are more proficient at capturing long-term dependencies in sequential data and are better suited for nonlinear prediction tasks. The fundamental principle of the LSTM algorithm hinges on the design of three critical components: the input gate, the forget gate, and the output gate, as schematically illustrated in Fig. 5.

In Fig. 5, C_{t-1} and C_t are the memory units of the $t - 1$ and t moment, respectively. The specific operation process is shown in (22)-(25). (1) The forget gate is the part responsible for controlling the transmission of information, mainly responsible for extracting some effective information from the input of the outside world.

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (22)$$

Where: f_t is the input of the forget gate; δ is the activation function; W_f is the weight of the input of the forget gate; h_{t-1} is the value of the output state in the previous moment; X_t is the input quantity; b_f is the deviation of the forget gate.

(2) The input gate determines whether new information is added to memory.

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \quad (23)$$

$$g_t = \tanh(W_c[h_{t-1}, X_t] + b_c) \quad (24)$$

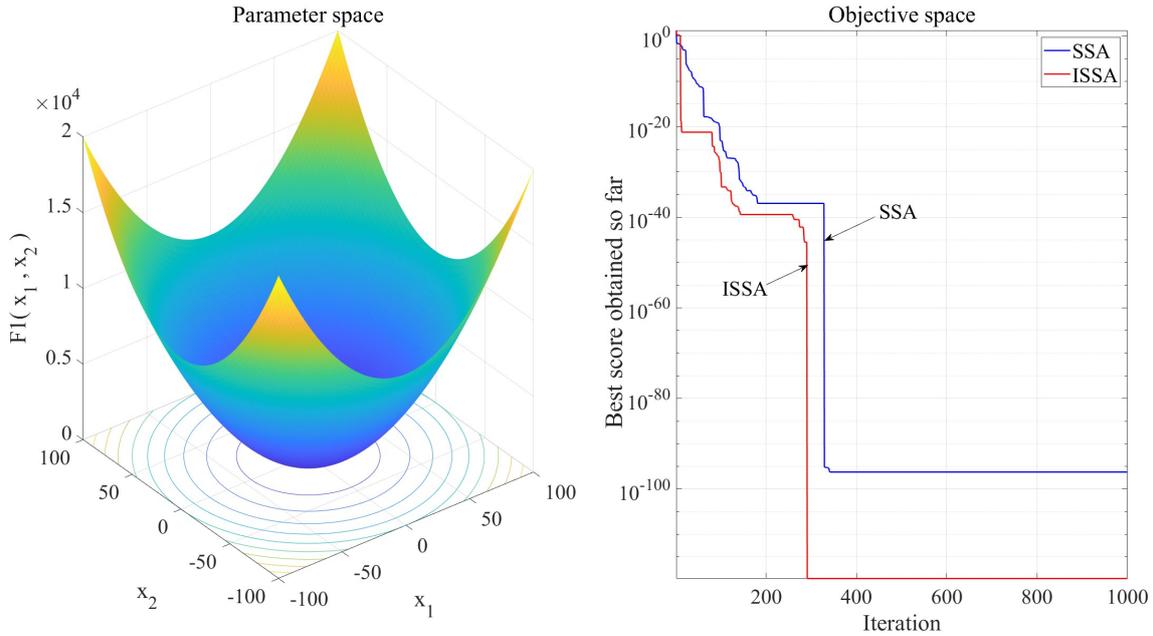


Fig. 4. Comparison of SSA and ISSA

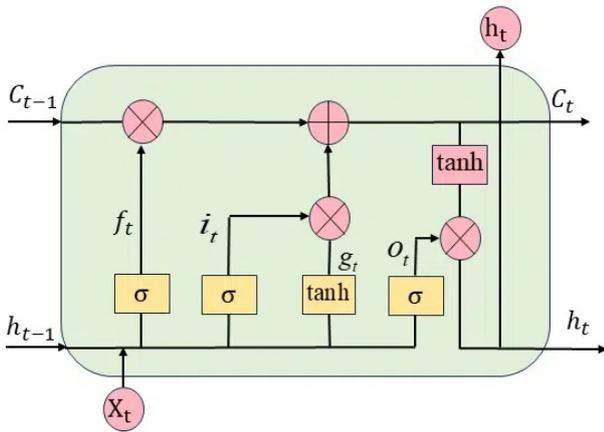


Fig. 5. LSTM Structure

Where: i_t is the input of the input gate; W_i is the input gate weight; b_i is the input gate deviation; g_t is the candidate cell state; W_c is the memory cell weight; b_c is the memory cell deviation.

(3) The output gate is responsible for deciding what information to output.

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad (25)$$

where: o_t is the state of the output gate; W_o is the weight of the output gate; b_o is the deviation of the output gate.

E. GWO-VMD-ISSA-LSTM Model

The algorithmic flow of the GWO-VMD-ISSA-LSTM model is shown in Fig. 6.

In this paper, we propose a multi-algorithm coupling model that leverages LSTM networks for efficiently learning and training time series data to extract temporal features.

The model includes three main stages: First, the Grey Wolf Optimizer (GWO) algorithm is utilized to determine the optimal values for two key parameters, K and α , of the Variational Mode Decomposition (VMD). Then, the GWO-optimized VMD algorithm is utilized to decompose the original power sequences into k components. In the second stage, the Improved Sparrow Search Algorithm (ISSA) is employed to optimize the three critical parameters of LSTM, leading to the establishment of a coupled model integrating the Improved Sparrow Search Algorithm with LSTM, termed ISSA-LSTM. This step constructs the ISSA-LSTM coupled model. Finally, in the third stage, each component is individually input into the ISSA-LSTM model for prediction, and the predicted values of all components are summed to obtain the final prediction result.

III. EXPERIMENT AND RESULT ANALYSIS

A. Experimental Data

To assess the feasibility of the GWO-VMD-ISSA-LSTM forecasting model, this paper employs the electricity load dataset for Panama, which is publicly available on the Kaggle platform. The dataset is available in the form of hourly records and includes electricity load data for Panama from January 2015 to June 2020 as well as weather data for three major cities. In this paper, a total of 8,760 data points from 0:00 January 1, 2019 to 23:00 December 30, 2019 were selected for the study, and we allocated the first 70% of the data to the training set, reserving the final 30% for the test set, aiming to study short-term electricity load forecasting at 1-hour intervals. Weather data is taken as average of three cities.

In this paper, we analyze the data for 2019 electricity data, the raw data is shown in Fig. 7(a). The time span of the data is too small, which leads to a relatively strong data oscillation and it is not easy to see the trend of the data, so in order to analyze the long term trend of

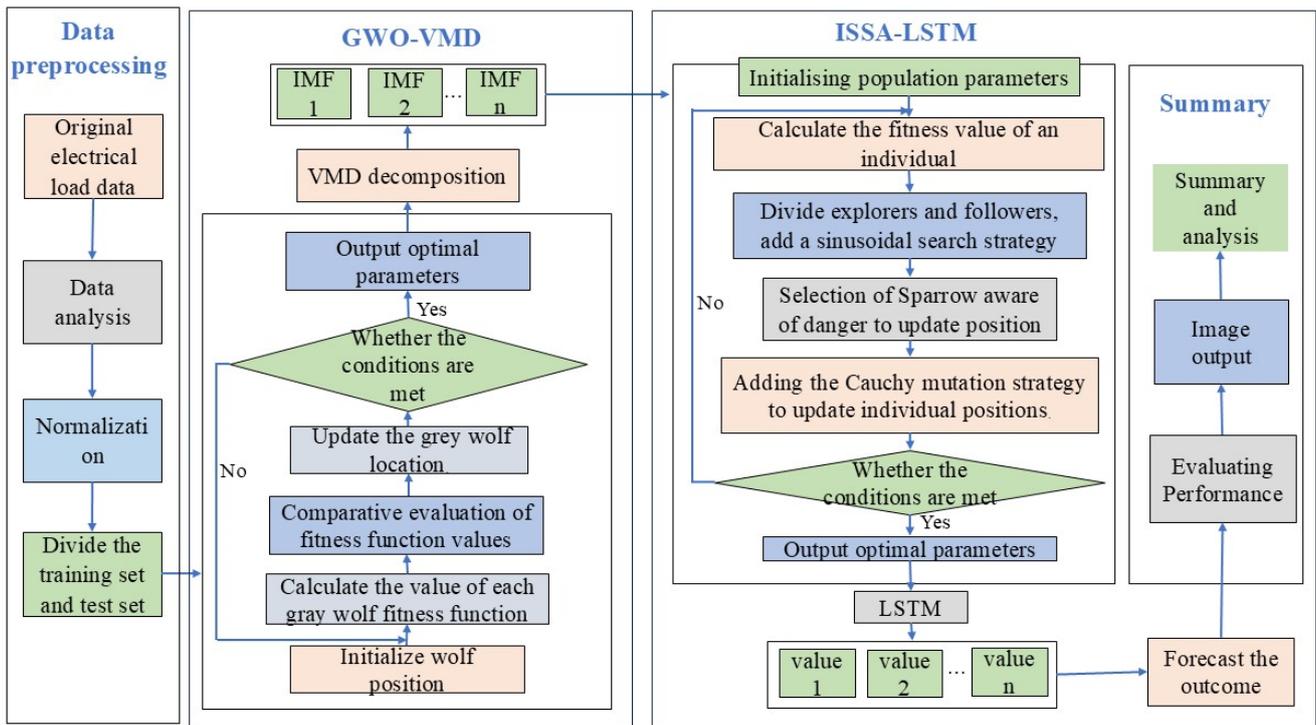


Fig. 6. Flowchart of GWO-VMD-ISSA-LSTM Algorithm

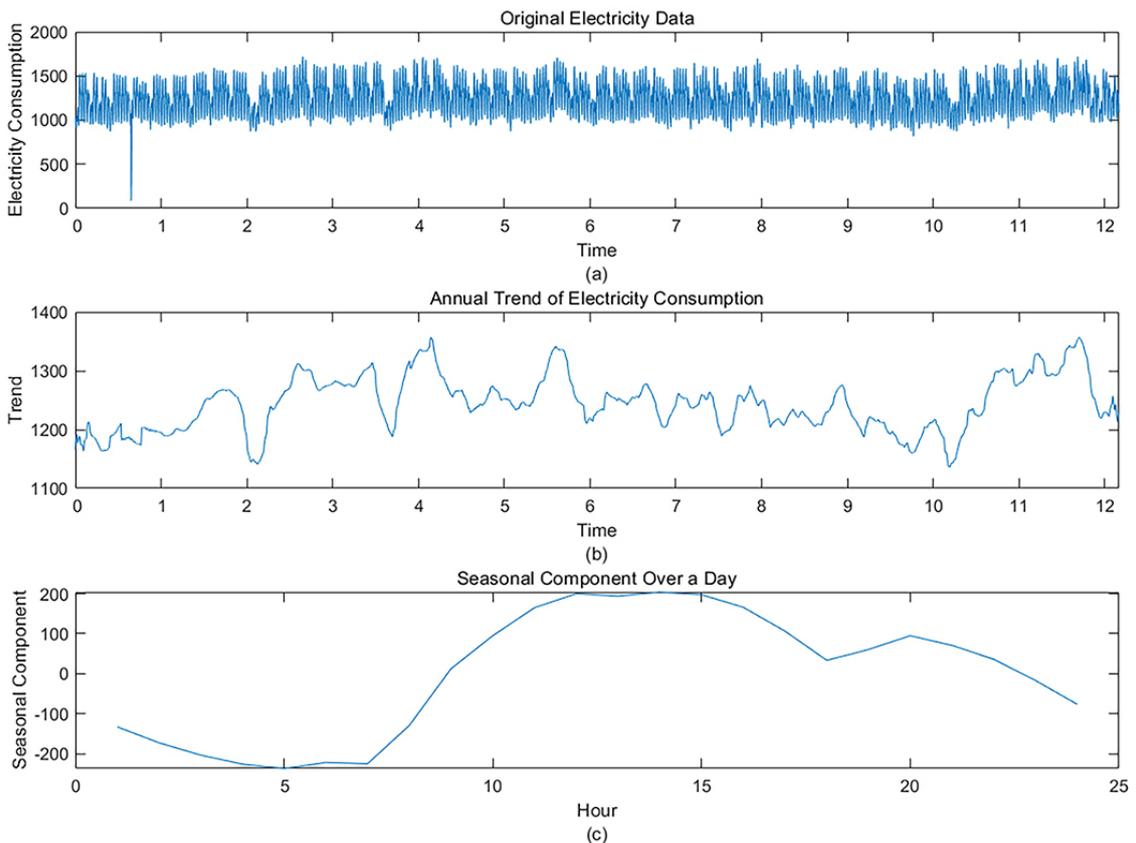


Fig. 7. Data Analysis Chart

the electricity load consumption, we use a 7-day moving average method to calculate the trend components. The trend

component was obtained by calculating the average of the data points every 168 hours (7 days). Fig. 7(b) shows the

trend components, which shows the overall trend of electrical energy consumption throughout 2019. As can be seen from the graph, Panama's electricity consumption figures show a very stable overall performance throughout the year, a phenomenon that is mainly attributed to the country's tropical climate, where temperatures are relatively stable throughout the year, with small temperature differences, so that air-conditioning use in the summer and heating needs in the winter are not as pronounced as they would be in a temperate country. Slightly higher figures were recorded for the March to July and November to December periods, a phenomenon that is mainly due to a combination of several factors: 1) the hot weather during the Panamanian summer months of March to July leads to a significant increase in air conditioning use, which pushes up the demand for electricity; 2) November to December is the peak seasons for tourism, especially during the important festivals of Christmas and New Year's Day, when the number of tourists rises and the number of hotels, restaurants and other tourism-related facilities; 3) electricity consumption in residential households also increases during these two time periods due to summer cooling and winter heating demand. The combination of these factors has led to higher electricity consumption figures for Panama during these two periods than the average for the year as a whole.

In order to analyze the cyclical variations in electrical energy consumption, we calculated the seasonal component for 24 hours per day. The seasonal component was obtained by calculating the average of the difference between the data points and the trend component for each time period. Fig. 7(c) is a seasonal component plot, which shows the pattern of variation in electrical energy consumption over a 24-hour period within a day. By seasonally decomposing Panama's hourly electricity consumption data for 2019, we can see that the seasonal component curves are relatively smooth, indicating that Panama's daily electricity consumption pattern is relatively stable and does not fluctuate dramatically. Specifically, the seasonal component value from 8:00 to around 22:00 is positive, indicating that the actual electricity consumption is higher than the trend component during this period, due to the fact that this is the time when the most residential and commercial activities take place, including activities such as going to work, going to school, shopping and entertainment. However, after 22:00 and before 5:00 a.m. when most of the residents are at rest, the demand for electricity decreases and the value of the seasonal component turns negative. The trend of the curve shows that Panama's daily electricity consumption pattern is consistent with the distribution of people's daily work and rest schedules.

B. Indicators for Data Preprocessing and Model Evaluation

As the dataset is devoid of any missing values or outliers, we proceed directly to normalize the data using the equation provided below.

$$x_{i,s}^{norm} = \frac{x_{i,s} - x_s^{\min}}{x_s^{\max} - x_s^{\min}} \quad (26)$$

where: $x_{i,s}^{norm}$ denotes the normalized value of the s -th component of the eigenvector of the i -th sample, $x_{i,s}$ is the value before normalization, x_s^{\max} and x_s^{\min} are the maximum and minimum values of the first component of the eigenvector, respectively.

TABLE I
OPTIMAL PARAMETERS FOR VMD

VMD	K	α
value	10	3000

In this paper, the predictive ability of the model is evaluated using a variety of assessment metrics, including root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE) and mean absolute percentage error (MAPE). These metrics are calculated using the following formulas:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (u_{actual}^t - u_{predict}^t)^2} \quad (27)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |u_{actual}^t - u_{predict}^t| \quad (28)$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (u_{actual}^t - u_{predict}^t)^2 \quad (29)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{u_{actual}^t - u_{predict}^t}{u_{actual}^t} \right| \quad (30)$$

A lower error value signifies that the predicted value is closer to the actual value, indicating a higher prediction accuracy for the model.

C. The GWO-VMD Model

To guarantee the quality of the input data and mitigate the impact of noise on predictions, this paper employs Variational Mode Decomposition (VMD) to decompose the electricity load input sequence. Given that applying VMD to decompose power load data necessitates setting multiple parameters, this paper utilizes the Grey Wolf Optimizer (GWO) algorithm to identify the optimal VMD parameters. After optimization, the optimal penalty factor α and the optimal number of modal component decompositions K are obtained, which are then used to decompose the original signal through VMD. In this paper, the minimum envelope entropy is selected as the fitness function for the GWO algorithm, and the resulting optimal parameter values are presented in Table I.

Fig. 8 displays the time-domain waveforms of the ten Intrinsic Mode Functions (IMFs) decomposed by the GWO-optimized VMD. It is evident from the figure that the eigenfrequencies of the IMFs across different time domains are distinctly differentiated and exhibit a more regular distribution. This effectively circumvents the issue of modal aliasing.

D. ISSA-LSTM Model

This paper utilizes the Improved Sparrow Search Algorithm (ISSA) to determine the optimal values for three key LSTM parameters: the number of hidden units, the maximum number of training period, and the initial learning rate. To ensure fairness, the key parameters such as initial population size and iteration number are kept the same for SSA and

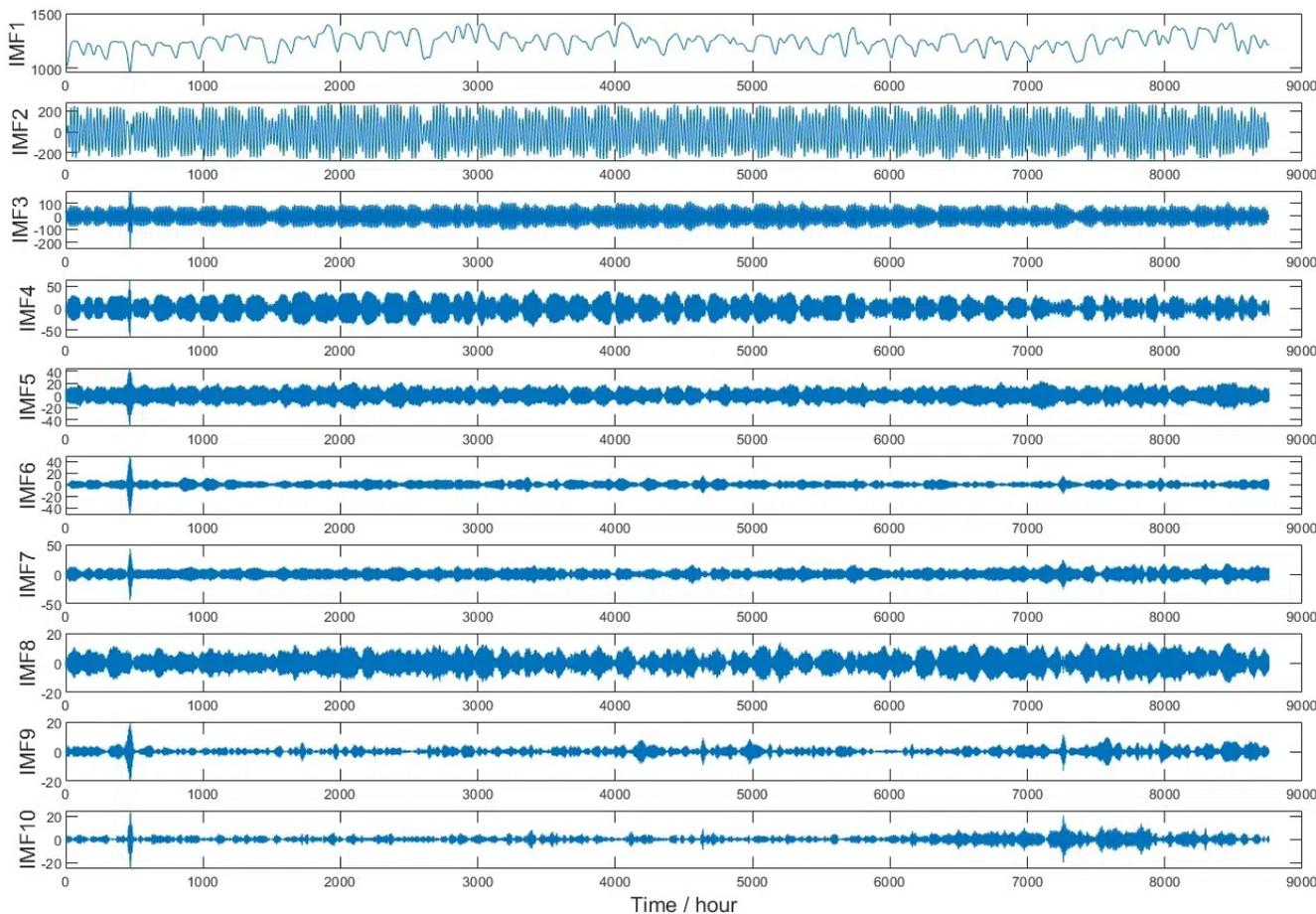


Fig. 8. GWO-VMD Frequency Domain Analysis

TABLE II
OPTIMAL PARAMETERS OF LSTM

LSTM	Number of hidden units	Maximum training period	Initial learning rate
value	181	300	0.0027838

ISSA in this paper. Fig. 9 shows the evolution curve graphs obtained by SSA and ISSA on the parameters of LSTM for optimization search, and it can be clearly seen that ISSA is able to find the suitable hyperparameter combinations faster during the training process relative to SSA to decrease the training time, and the best fitness value of SSA-LSTM is 0.0302, and the best fitness value of ISSA-LSTM is 0.0291, the lower adaptation value indicates the higher quality of its final solution.

The final optimal parameter values obtained are shown in Table II.

E. Forecasting Results

1) *Single Model Prediction Results:* To evaluate the predictive performance of various models, we have forecasted Panama’s 2019 electricity data using five models, LSTM, VMD-LSTM, VMD-SSA-LSTM, VMD-ISSA-LSTM and GWO-VMD-ISSA-LSTM, and plotted line graphs of the forecasts of each model against the original data.

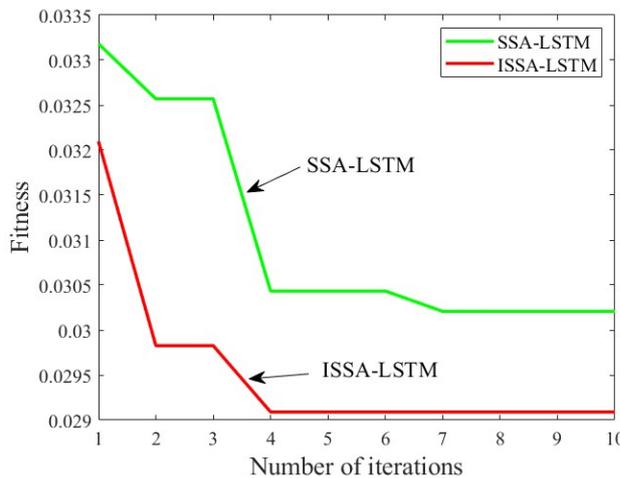


Fig. 9. Evolutionary Curve

From Fig. 10(a) to Fig. 10(e), as the model complexity rises, it becomes evident that the prediction accuracy progressively improves. Among the five models, the prediction results of the single LSTM model differed the most from the original data, mainly due to the difficulty of the LSTM model in handling complex nonlinear relationships. In contrast, the VMD-LSTM model introduces the VMD to preprocess the raw data, which adequately separates the modes at different

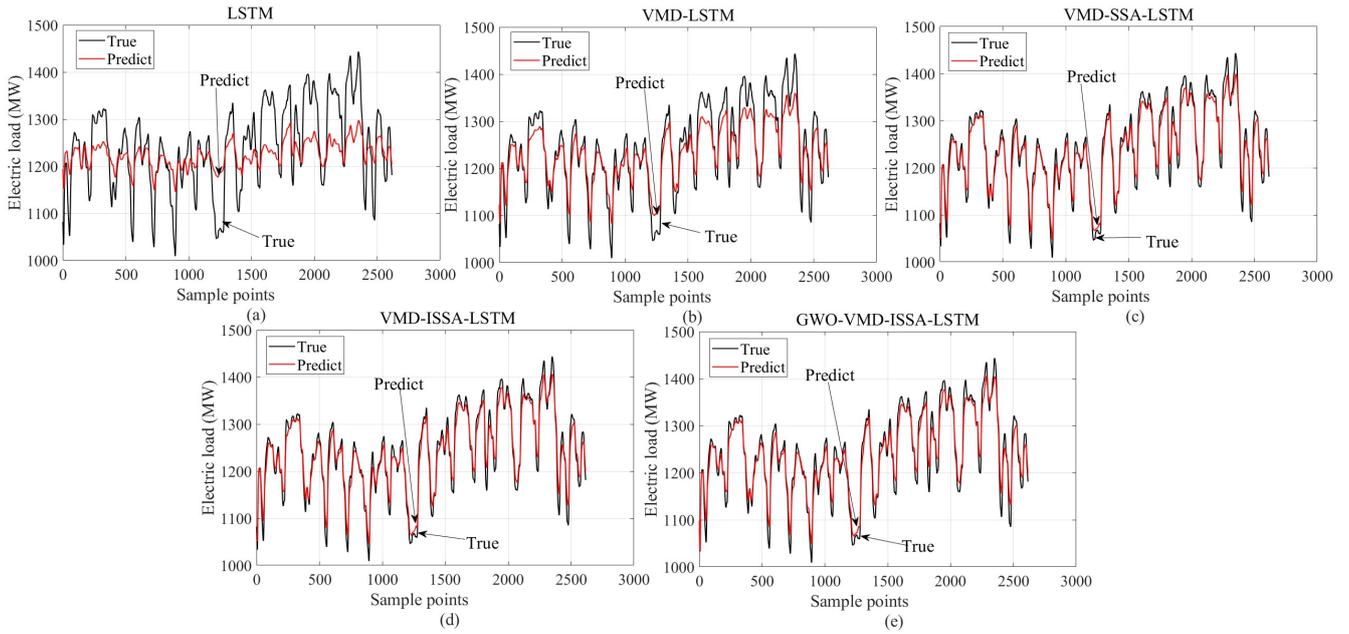


Fig. 10. Single Model Prediction

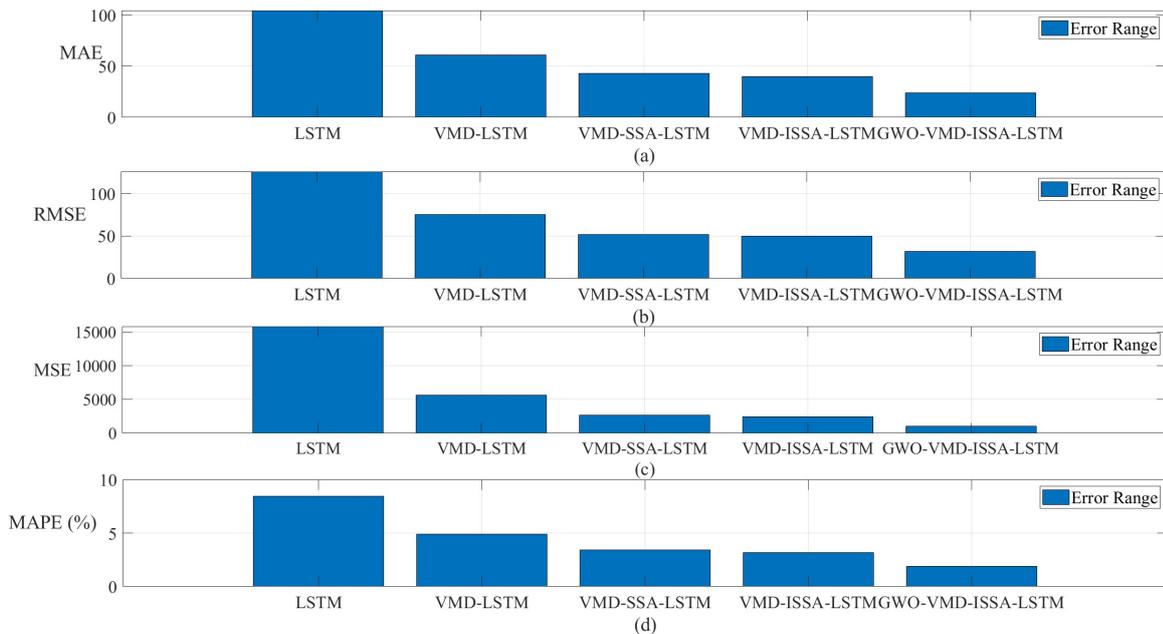


Fig. 11. Comparison of Model Prediction Errors

frequencies and makes the model perform better in dealing with high-frequency fluctuations. Further, the VMD-SSA-LSTM model in turn incorporates the SSA algorithm, which not only separates the modes at different frequencies, but also optimizes the key parameters of the LSTM model, resulting in a more stable performance in predicting long-period trends. The VMD-ISSA-LSTM model replaces the original SSA with the improved ISSA. By incorporating the sinusoidal search strategy and the Cauchy mutation strategy, it balances the global and local search performance. Additionally, when the aggregation degree is high, it aids the algorithm in escaping from local optima, resulting in a

stronger prediction capability for the overall model. Finally, the GWO-VMD-ISSA-LSTM model introduces GWO to optimize the two key parameters of VMD, which avoids the subjectivity of manual parameter tuning, and achieves the highest accuracy prediction, almost eliminating all the significant prediction errors.

The prediction error values for the five models are shown in Table III.

Among the five models, the GWO-VMD-ISSA-LSTM model exhibits the lowest values for MAE, RMSE, MSE and MAPE, which are 23.8017, 32.1823, 1035.6991 and 1.9095% respectively, indicating its superior performance in terms

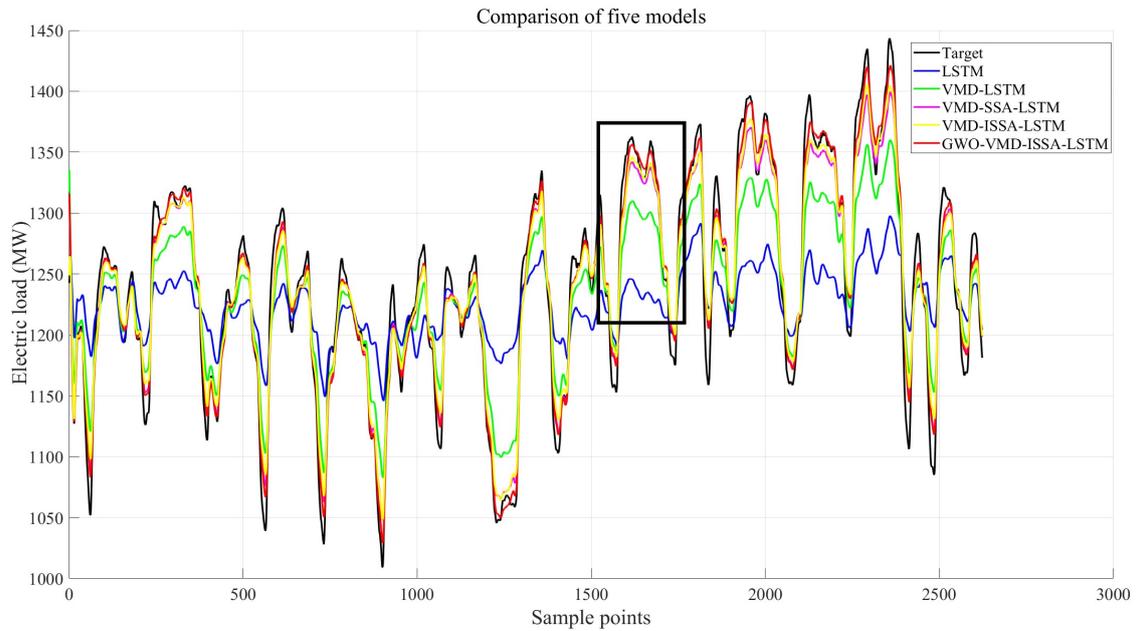


Fig. 12. Comparison of Model Prediction Errors

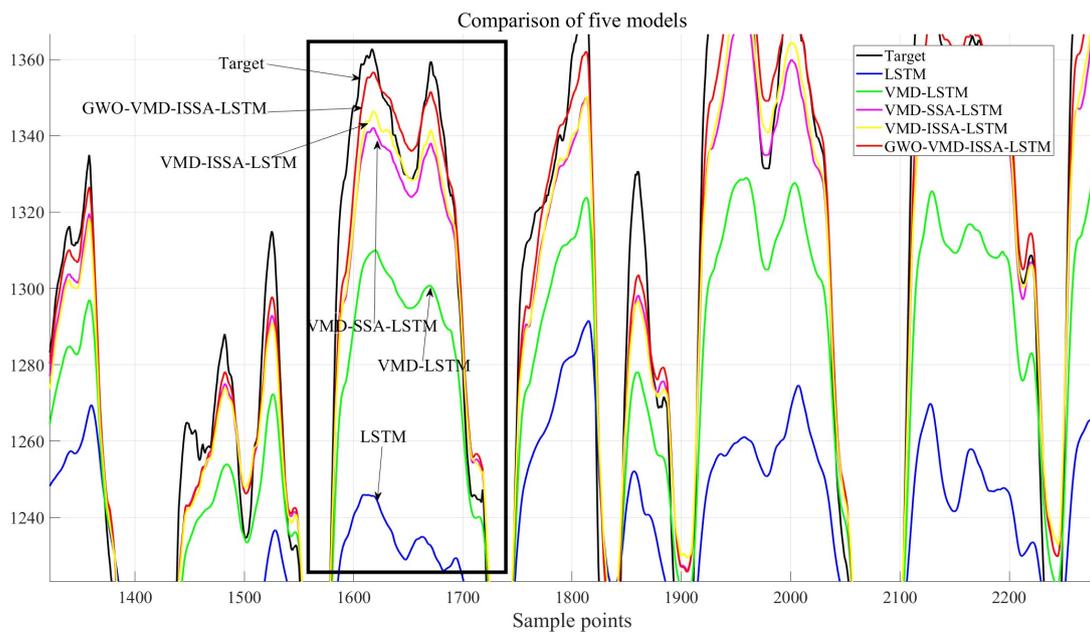


Fig. 13. Detail View

of prediction accuracy and stability. This model offers fresh perspectives and methodologies for forecasting complex time

series data.

TABLE III
MODEL PREDICTION ERRORS

Model	MAE	RMSE	MSE	MAPE
LSTM	104.5225	125.9036	15851.7244	8.4250%
VMD-LSTM	61.2528	75.3523	5677.9719	4.8933%
VMD-SSA-LSTM	42.7666	51.9493	2698.7259	3.4268%
VMD-ISSA-LSTM	39.8720	49.6835	2468.4520	3.1770%
GWO-VMD-ISSA-LSTM	23.8017	32.1823	1035.6991	1.9095%

Fig. 11 presents the error bar chart generated based on the prediction results of various models, from which it is evident that the error values of the five models exhibit a gradual decrease from the LSTM model to the GWO-VMD-ISSA-LSTM model. This indicates that with the increase in model complexity and the introduction of improved algorithms, the prediction performance has been significantly enhanced. Firstly, the LSTM model, serving as the baseline model, exhibits the poorest performance across all error metrics. Subsequently, the introduction of VMD significantly enhances the model's capability to process raw data by decom-

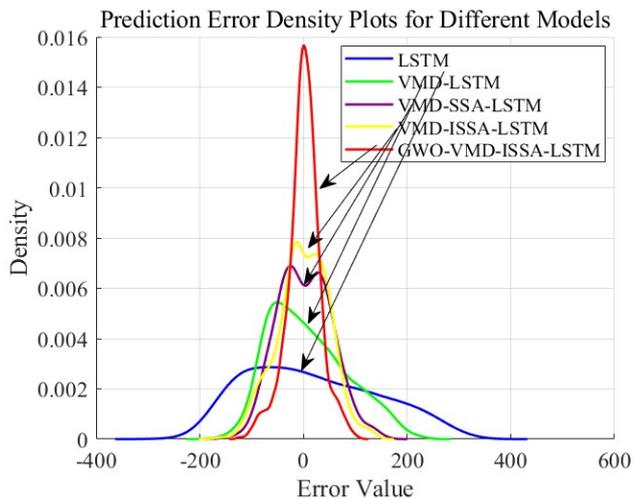


Fig. 14. Model Prediction Error Density Chart

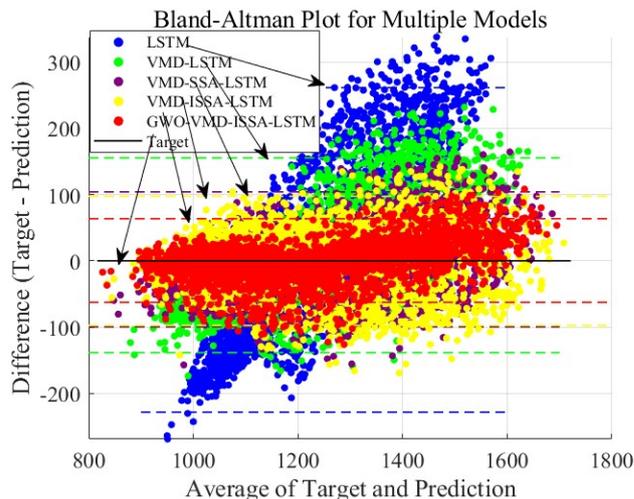


Fig. 15. Bland-Altman Plot for Multiple Models

posing complex signals into multiple frequency components, rendering each sub-signal more manageable for modeling and thereby improving prediction accuracy. Furthermore, the incorporation of SSA and ISSA further elevates the model’s performance, with ISSA demonstrating a superior ability in finding optimal solutions compared to SSA, resulting in relatively lower prediction errors. Finally, the integration of GWO further optimizes the selection of model parameters, ensuring that the ultimate GWO-VMD-ISSA-LSTM model achieves the best performance across all error metrics. This provides reliable technical support for time series forecasting in practical applications.

2) *Multi-model Comparison:* To provide a more intuitive comparison of the performance of the five models mentioned above, we plot the predictive value curves of all models on a single graph for comparison. Fig. 12 shows the comparison between the predicted value curves of the five models and the original data value curves, in which the black line indicates the original data value, and the blue, green, purple, yellow and red lines indicate the predicted result values of the different models, respectively, and it can be clearly seen that GWO-VMD-ISSA-LSTM outperforms the other four models in terms of performance. To better visualize the details, Fig. 13 provides a detail view of the region enclosed by the black rectangle in Fig. 12. This magnified section highlights the detailed prediction curves of the models. Fig. 14 shows the density plot of the prediction error of different models, the density plot shows the distribution of the prediction error of different models, by observing the density curve of each model, we can understand the concentration of the error, the distribution pattern and whether there are obvious outliers. As can be seen from Fig. 14, the prediction errors of the GWO-VMD-ISSA-LSTM model are mainly concentrated near zero, and the maximum point of the density curve is higher, evidencing that most of the prediction errors are smaller; and the density curve shows a more symmetrical shape, indicating that the distribution of positive and negative errors is more uniform and the prediction bias of the model is smaller; the shorter tails of the density curves indicate fewer extreme errors and better robustness of the model.

To more comprehensively evaluate the consistency and systematic bias of predictive models, as well as to identify

potential outliers and anomalies, this paper incorporates Bland-Altman plots to compare the prediction results of five different models. The Bland-Altman plot not only uncovers systematic bias issues that arise as measured values change; it also demonstrates the error range for 95% of the data points through the Limits of Agreement (LoA). This allows us to clearly see which predictions fall outside the expected error interval. From Fig. 15, it is evident that the GWO-VMD-ISSA-LSTM model exhibits significant advantages compared to other models. Firstly, in terms of scatter plot distribution, the points for the GWO-VMD-ISSA-LSTM model are most densely clustered around the zero-difference line (Target line), indicating that the deviations between the model’s predictions and actual values are minimal, thereby demonstrating high accuracy. Furthermore, the Limits of Agreement (LoA) for the GWO-VMD-ISSA-LSTM model are the narrowest among all models, reflecting a high degree of consistency in its predictions and low random variation, which attests to the model’s reliability and stability.

Additionally, the optimized model has the smallest number of outliers, suggesting its excellent performance in handling extreme situations and its ability to effectively avoid abnormal predictions. This characteristic is crucial for ensuring the robustness of the model in practical applications. Lastly, in contrast to the obvious upward trends exhibited by other models, the scatter plot distribution of the optimized model is closer to a horizontal line, implying that the prediction error of this model remains stable as the measured values change, without significant systematic bias occurring.

In summary, the GWO-VMD-ISSA-LSTM model distinguishes itself among all tested models due to its high accuracy and consistency, excellent robustness, and stable prediction performance. These advantages are not only attributed to the optimizations made by ISSA to the traditional SSA but also to the enhancement provided by GWO to VMD, collectively driving the overall performance of the model to new heights.

IV. CONCLUSION

As society and the economy advance, and people’s energy demands increase, the stability and reliability of power systems are facing unprecedented challenges. Especially in

recent times, the integration of large-scale renewable energy sources like wind and solar into the grid has led to a substantial increase in the power system's uncertainty. Against this background, accurate short-term power forecasts not only help power companies to rationalize their power generation schedules and reduce unnecessary energy wastage, but also promote the effective integration of renewable energy sources, therefore improving the operational efficiency and economic benefits of the entire power system. In this paper, a combined model combining the Grey Wolf Optimizer (GWO) algorithm, Variational Mode Decomposition (VMD), Improved Sparrow Search Algorithm (ISSA), and Long Short-Term Memory networks (LSTM) is proposed to predict the Panama's electricity data in conjunction with the four environmental factors: temperature, humidity, precipitation, and wind speed, and a high prediction accuracy is achieved. The key findings are outlined below:

(1) The VMD decomposition algorithm effectively mitigates noise and eliminates the non-smoothness inherent in original electricity load data. To address the challenge of selecting the optimal values for the quadratic penalty sum factor and modal decomposition number in VMD, we introduce the GWO-VMD model. This model utilizes the GWO algorithm as the optimization algorithm to determine the best parameters for VMD. The minimum envelope entropy serves as the fitness function, allowing us to evaluate the quality of the features extracted through VMD. Using the GWO algorithm in combination with the VMD to decompose the power series overcomes the limitation of the traditional VMD that requires manual parameter tuning, and can significantly enhance the predictive capability of the model.

(2) In this paper, the traditional SSA is enhanced by incorporating a sinusoidal search strategy. This strategy boosts the weights of promising individuals, accelerates the algorithm's convergence speed, and achieves a balance between global and local search performance. Additionally, a Cauchy variation strategy is introduced to enhance the algorithm's population diversity when the aggregation degree is high, thereby assisting the algorithm in escaping from local optima. ISSA is able to be faster than SSA during the training process to find a suitable combination of hyperparameters and reduce the training time and resource consumption, and the best fitness value of SSA-LSTM is lower than that of ISSA-LSTM, which indicates that the quality of its final solution is higher, and verifies the effectiveness of the improvement of the traditional sparrow search algorithm.

(3) The GWO-VMD-ISSA-LSTM coupled model proposed in this paper outperforms traditional LSTM and other benchmark models on actual power datasets, indicating that this model has significant practical value and broad promotion potential.

V. OUR FURTHER WORK

Although the GWO-VMD-ISSA-LSTM model proposed in this paper has achieved high-accuracy short-term electricity load forecasting on actual datasets, the dataset used in this study only includes climate-related factors such as temperature. However, the factors influencing power load are diverse. Therefore, to further enhance the comprehensiveness and prediction accuracy of the model, various types of influencing factors can be incorporated in future work, such

as socioeconomic factors, policy and regulatory factors, and holiday effects. By introducing these non-climatic factors, not only can the forecasting model be enriched, but it can also be made more aligned with real-world scenarios, thereby providing more precise and valuable decision support for power sectors. In summary, future efforts will be dedicated to collecting relevant data and developing corresponding algorithms to integrate these new factors, with the aim of further improving the performance of short-term power load forecasting.

REFERENCES

- [1] Z. Su, J. Zeng, J. Liu *et al.*, "A short-term power load forecasting approach based on emd-ssa-lstm," in *Journal of Physics: Conference Series*, vol. 2774, no. 1. IOP Publishing, 2024, p. 012005.
- [2] Z. Chen, G. Chen, W. Wang *et al.*, "Short-term power load forecasting based on ceemdan-cnn-lstm hybrid modeling," in *2024 12th International Conference on Intelligent Control and Information Processing (ICICIP)*. IEEE, 2024, pp. 9–16.
- [3] N. T. N. Anh, N. N. Anh, T. N. Thang *et al.*, "Online SARIMA applied for short-term electricity load forecasting," *Applied Intelligence*, vol. 54, no. 1, pp. 1003–1019, 2024.
- [4] M. S. Bakare, A. Abdulkarim, A. N. Shuaibu *et al.*, "A hybrid long-term industrial electrical load forecasting model using optimized anfis with gene expression programming," *Energy Reports*, vol. 11, pp. 5831–5844, 2024.
- [5] C. Tarmanini, N. Sarma, C. Gezegein *et al.*, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, pp. 550–557, 2023.
- [6] S. Chaturvedi, E. Rajasekar, S. Natarajan *et al.*, "A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for india," *Energy Policy*, vol. 168, p. 113097, 2022.
- [7] S. Zhang, R. Chen, J. Cao *et al.*, "A CNN and LSTM-based multi-task learning architecture for short and medium-term electricity load forecasting," *Electric Power Systems Research*, vol. 222, p. 109507, 2023.
- [8] X. Dong, S. Deng, and D. Wang, "A short-term power load forecasting method based on k-means and svm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 11, pp. 5253–5267, 2022.
- [9] G. F. Fan, L. Z. Zhang, M. Yu *et al.*, "Applications of random forest in multivariable response surface for short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 139, p. 108073, 2022.
- [10] X. Q. Yu, X. R. Yao, and J. Gao, "Study on plant diseases and insect pests recognition based on deep learning," *IAENG International Journal of Computer Science*, vol. 52, no. 1, pp. 111–120, 2025.
- [11] X. Song, J. Man, S. Song *et al.*, "Gain-scheduled finite-time synchronization for reaction-diffusion memristive neural networks subject to inconsistent markov chains," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 2952–2964, 2020.
- [12] M. R. Chen, G. Q. Zeng, K. D. Lu *et al.*, "A two-layer nonlinear combination method for short-term wind speed prediction based on elm, enn, and lstm," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6997–7010, 2019.
- [13] H. Aprillia, H. T. Yang, and C. M. Huang, "Statistical load forecasting using optimal quantile regression random forest and risk assessment index," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1467–1480, 2020.
- [14] Z. Meng, X. Xie, Y. Vie, and J. Sun, "Federated learning for short-term load forecasting," *IAENG International Journal of Computer Science*, vol. 52, no. 1, pp. 52–58, 2025.
- [15] Z. Wang, Y. Ying, L. Kou *et al.*, "Ultra-short-term offshore wind power prediction based on pca-ssa-vmd and bilstm," *Sensors*, vol. 24, no. 2, 2024.
- [16] J. Wei, X. Wu, T. Yang *et al.*, "Ultra-short-term forecasting of wind power based on multi-task learning and lstm," *International Journal of Electrical Power & Energy Systems*, vol. 149, p. 109073, 2023.
- [17] G. Geng, Y. He, J. Zhang *et al.*, "Short-term power load forecasting based on PSO-optimized VMD-TCN attention mechanism," *Energies*, vol. 16, no. 12, p. 4616, 2023.
- [18] Y. S. Ham, K. B. Sonu, U. S. Paek *et al.*, "Comparison of LSTM network, neural network and support vector regression coupled with wavelet decomposition for drought forecasting in the western area of the DPRK," *Natural Hazards*, vol. 116, no. 2, pp. 2619–2643, 2023.

- [19] O. Abedinia, M. Lotfi, M. Bagheri *et al.*, "Improved EMD-based complex prediction model for wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2790–2802, 2020.
- [20] A. Meng, S. Chen, Z. Ou *et al.*, "A hybrid deep learning architecture for wind power prediction based on bi-attention mechanism and crisscross optimization," *Energy*, vol. 238, p. 121795, 2022.
- [21] X. Chen, Y. Yang, Z. Cui *et al.*, "Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy," *Energy*, vol. 174, pp. 1100–1109, 2019.
- [22] J. Li, Z. Song, X. Wang *et al.*, "A novel offshore wind farm typhoon wind speed prediction model based on PSO-Bi-LSTM improved by VMD," *Energy*, vol. 251, p. 123848, 2022.
- [23] J. H. Ye, J. Cao, L. Yang *et al.*, "Ultra short-term load forecasting of user level integrated energy system based on variational mode decomposition and multi-model fusion," *Power System Technology*, vol. 46, no. 7, pp. 2610–2618, 2022.
- [24] D. Yi, Y. Zhang, and W. Zhang, "Image enhancement CHPSO processing technology based on improved particle swarm optimization algorithm," *IAENG International Journal of Computer Science*, vol. 52, no. 1, pp. 130–142, 2025.
- [25] X. Liu, N. Shi, Y. Wang *et al.*, "Data-driven boolean network inference using a genetic algorithm with marker-based encoding," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 3, pp. 1558–1569, 2021.
- [26] Q. Han, X. Yang, H. Song *et al.*, "Multi-objective ship path planning using non-dominant relationship-based whale optimization algorithm in marine meteorological environment," *Ocean Engineering*, vol. 266, p. 112862, 2022.
- [27] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
- [28] H. U. Ahmed, R. R. Mostafa, A. Mohammed *et al.*, "Support vector regression (SVR) and grey wolf optimization (GWO) to predict the compressive strength of GGBFS-based geopolymer concrete," *Neural Computing and Applications*, vol. 35, no. 3, pp. 2909–2926, 2023.
- [29] Z. Wang, Y. Ying, L. Kou *et al.*, "Ultra-short-term offshore wind power prediction based on pca-ssa-vmd and bilstm," *Sensors*, vol. 24, no. 2, 2024.
- [30] M. A. Awadallah, M. A. Al-Betar, I. A. Doush *et al.*, "Recent versions and applications of sparrow search algorithm," *Archives of Computational Methods in Engineering*, vol. 30, no. 5, pp. 2831–2858, 2023.
- [31] D. Niu, L. Sun, M. Yu *et al.*, "Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model," *Energy*, vol. 254, p. 124384, 2022.
- [32] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [33] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2013.
- [34] J. Xue, "Research and application of a novel swarm intelligence optimization technique: Sparrow search algorithm," Ph.D. dissertation, Donghua University, 2020.