Improved Sparrow Search Algorithm-Based Recurrent Neural Network for Short-Term Generation Load Forecasting of Hydropower Stations

Liyuan Sun[1], Yilun Dong[2]*, Junwei Yang[3]
Measurement Center, Yunnan Power Grid Co., Ltd., Kunming, China[1]
Faculty of Transportation Engineering, Kunming University of Science & Technology, Kunming, China[2]
LongShine Technology Group Co., Ltd., Wuxi, China[2, 3]

***Abstract*—To address the challenges of low accuracy and high randomness in short-term hydroelectric load forecasting within Multi-energy Coupled Virtual Power Plants (MC-VPPs), this study proposes a hybrid model integrating Variational Mode Decomposition (VMD), Long Short-Term Memory (LSTM) networks, and an Improved Sparrow Search Algorithm (ISSA). Traditional methods, such as exponential smoothing and multiple linear regression, often fail to capture nonlinear dynamics and external disturbances. The proposed framework first decomposes raw load data into four intrinsic mode functions (IMFs) via VMD to extract multi-scale features, including long-term trends, seasonal cycles, and short-term fluctuations. LSTM networks are then applied to model the temporal dependencies of each IMF. To enhance optimization, ISSA introduces a bidirectional sine-cosine search strategy, balancing global exploration and local exploitation to avoid premature convergence. Validated on 1,247 daily load records from a hydropower station in southwestern China, the ISSA-VMD-LSTM model achieves a 30.2% improvement in R^2, with reductions of 47.2% in RMSE, 47.8% in MAE, and 63.3% in MAPE, outperforming benchmarks like PSO-LSTM and SSA-VMD-LSTM. This demonstrates its robustness in handling nonlinearity and stochasticity. The model enhances MC-VPPs' operational efficiency by enabling intelligent scheduling and renewable energy integration, with future applications extending to real-time forecasting and other renewable energy systems.**

Keywords*—Power plant; load forecasting; Mode Decomposition; Long Short-Term Memory; Sparrow Search Algorithm

I. Introduction

The Multi-energy Coupled Virtual Power Plant [1]-[4] plays an essential role in optimizing short-term load forecasting for hydroelectric power. It exemplifies the deepening integration of Energy Internet technology. By incorporating Long Short-Term Memory networks, MC-VPP significantly improves the accuracy of predicting power output from hydropower stations and load demand. This algorithmic ensemble not only captures the time-series characteristics of electricity loads but also effectively addresses the nonlinear impacts of hydro-meteorological factors, providing more accurate and reliable forecast results. As a core component of intelligent scheduling and optimized operation within MC-VPPs, short-term load forecasting for hydroelectric power offers robust technical support to mitigate the intermittency and uncertainty associated with renewable energy. It enhances the flexibility and efficiency of power systems and introduces new methods for managing distributed energy resources with complex and diverse characteristics. This is significant for promoting the widespread adoption of clean energy and fostering the healthy development of power markets.

Current research on short-term load forecasting for hydroelectric power has widely applied methods such as Exponential Smoothing, Combined Forecasting, and Grey Prediction [5]-[8]. However, these approaches are data-dependent and constrained by underlying assumptions. As the aggregated resources within MC-VPPs expand, short-term hydroelectric load trends become increasingly complex. Techniques based on Multiple Linear Regression Models [9]-[12] have gained prominence due to their computational efficiency and clear representation of relationships between independent and dependent variables. Nevertheless, these models assume a linear relationship between variables and are sensitive to outliers, making it challenging to directly address the dynamic characteristics [13] of time series and the influence of external events. When using multiple linear regression for load forecasting, it is necessary to combine other techniques to overcome these limitations. Exponential Smoothing assigns different weights to historical data, allowing predictions to quickly adapt to recent changes. This method is suitable for relatively stable load patterns but struggles with capturing long-term dependencies and complex nonlinear patterns, leading to a significant decrease in prediction accuracy during extreme weather conditions. In contrast, Combined Forecasting integrates several prediction models to leverage each one's strengths, improving the accuracy and robustness of forecasts. This method addresses the complexity and nonlinearity of hydroelectric load data [14-17], but its construction is complex, requiring the establishment of sub-models and the determination of weights, which increases computational costs and model tuning challenges.

The models used for optimizing short-term load forecasting in hydroelectric power all face issues like strict applicability conditions or an inability to handle complex nonlinear relationships, limiting their responsiveness to real-world forecasting scenarios. In [18], the authors addressed the low prediction accuracy and poor stability of LSTM neural networks by proposing a short-term power load forecasting

*Corresponding author.

model optimized using a multi-strategy improved Golden Jackal Algorithm [19]-[20]. This approach effectively enhances the training efficiency and performance of LSTM [21]-[24] models, avoiding overfitting risks and achieving higher prediction accuracy and stability compared to traditional LSTM models.

In summary, existing studies on short-term load forecasting for hydroelectric power have been impacted by limited usage conditions and the inability to resolve complex relationships, affecting prediction accuracy. To enhance this, we propose a hybrid forecasting method based on Variational Mode Decomposition, an improved Sparrow Search Algorithm [25]-[27], and Long Short-Term Memory networks. The proposed method begins by adaptively decomposing the original load data into multiple intrinsic mode functions using Variational Mode Decomposition. This process isolates distinct temporal features such as long-term trends, seasonal cycles, and short-term fluctuations, effectively mitigating mode mixing and enhancing adaptability to nonlinear time-series data. Next, each IMF is modeled using Long Short-Term Memory networks, which leverage their gated mechanisms to capture complex temporal dependencies and address the vanishing gradient problem inherent in traditional recurrent neural networks. To optimize model performance, an Improved Sparrow Search Algorithm is employed to fine-tune LSTM hyperparameters. This enhanced algorithm incorporates a bidirectional sine-cosine search strategy to balance global exploration and local exploitation, thereby avoiding premature convergence. The decomposition capability of VMD enables LSTM to focus on features at specific time scales, while ISSA further refines parameter adaptability. Together, these components form a cohesive framework of "decomposition-modeling-optimization", significantly enhancing prediction accuracy and stability. This approach effectively addresses the challenges of modeling nonlinearity, stochasticity, and multi-scale characteristics in multi-energy coupled environments.

The key contributions of this study are as follows:

- Combines Variational Mode Decomposition, Long Short-Term Memory networks, and an Improved Sparrow Search Algorithm, significantly enhancing the accuracy and stability of short-term hydroelectric load forecasting.
- Introduces a bidirectional sine-cosine search strategy and a Good Lattice Points-based initialization strategy to strengthen global exploration, avoid premature convergence, and optimize LSTM hyperparameters.
- Decomposes raw load data into intrinsic mode functions to isolate long-term trends, seasonal cycles, and short-term fluctuations, improving the model's adaptability to complex nonlinear time-series data.
- Validated using real hydropower station data, the model achieves improvements of 30.2% in R^2, 47.2% in RMSE, 47.8% in MAE, and 63.3% in MAPE, outperforming benchmarks such as PSO-LSTM [28]-[29] and SSA-VMD-LSTM [30].
- Supports intelligent scheduling and multi-energy coordination in MC-VPPs, enhancing flexibility for renewable energy integration and grid efficiency.

II. Hydroelectric Time Series Prediction Model

A. Working Principle of the Variational Mode Decomposition Algorithm

To deeply investigate short-term load forecasting for hydroelectric power in a multi-energy coupled virtual power plant environment, it is essential to employ the Variational Mode Decomposition algorithm to decompose the raw data into multiple intrinsic mode functions with distinct central frequencies and limited bandwidths, thereby achieving modal separation of the data. The core principle of VMD lies in its adaptive signal decomposition framework that concurrently optimizes multiple intrinsic mode functions through a variational approach, effectively balancing modal bandwidth constraints and reconstruction accuracy. By constructing a constrained variational problem, VMD iteratively extracts compact IMFs with specific sparsity properties in the frequency domain, where each mode is designed to concentrate around a central frequency while maintaining minimal bandwidth. This process continuously updates the mode and center frequency until convergence, ensuring that the decomposed components exhibit orthogonality in both the time and frequency domains. The resulting IMFs demonstrate distinct spectral separation characteristics, enabling the isolation of underlying patterns such as seasonal variations, random fluctuations, and operational trends from the original hydropower load data. This decomposition mechanism significantly enhances subsequent forecasting models by mitigating mode mixing issues inherent in traditional signal processing methods, while preserving critical temporal

$$\min_{\{u_k\},\{w_k\}} \left\{ \sum_k^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \quad (1)$$

dependencies essential for accurate predictions in virtual power plant environments characterized by strong multi-energy coupling effects.

The variational problem model of this method is subject to the following constraints:

$$s.t. \sum_k^K u_k(t) = x(t) \quad (2)$$

In the formula, $u_k = \{u_1, u_2, \cdots, u_k\}$ represents the intrinsic mode function, $w_k = \{w_1, w_2, \cdots, w_k\}$ represents the central frequency; $\delta(t)$ is the Dirac delta function, which is used in the computation of the Hilbert transform; K determining the number of decomposed IMF components; ∂_t addressing gradient dizziness; $\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t)$ performing the Hilbert transform.

Redefine the Intrinsic Mode Function with a stricter finite bandwidth constraint (BIMF)

Defined as:

$$u_k(t) = A_k(t)\cos(\phi_k(t)) \quad (3)$$

The phase function $\phi_k(t)$ is non-monotonically decreasing, which means $\phi_k'(t) \geq 0$, and the amplitude $A_k(t) \geq 0$, with the instantaneous amplitude $A_k(t)$ and the instantaneous frequency $\omega_t(t) = \phi_k'(t) = \frac{d\phi_k(t)}{d_t}$ varying slowly relative to $\phi_k(t)$, within the interval $[t-\tau, t+\tau]$, where $\tau = \frac{2\pi}{\phi_k'(t)}, u_k(t)$ can be considered as a harmonic signal with an amplitude of $A_k(t)$ and a frequency of $\omega_t(t)$.

The expression for the Hilbert transform is the result of convolving the original signal with a specific kernel. After applying the Hilbert transform to the signal, the positive frequency components are multiplied by -j, which means that under the condition of maintaining the same amplitude, the phase is shifted by $-\frac{\pi}{2}$. For the negative frequency components, the phase is shifted by $\frac{\pi}{2}$.

Through spectral analysis, we validate the frequency separation characteristics of individual Intrinsic Mode Functions. Quantitative evaluation using statistical measures like sample entropy reveals distinct complexity patterns - lower entropy values indicate stable trend components while higher entropy corresponds to stochastic fluctuations. Temporal correlation analysis between IMFs and established physical mechanisms/external drivers enables identification of specific modal functions demonstrating significant associations with extreme weather events. The orthogonal validation further confirms the temporal-frequency independence among decomposed components, effectively preventing mode mixing. This rigorous decomposition process ensures both mathematical validity and physical interpretability, providing reliable inputs for subsequent LSTM modeling.

B. LSTM Model

LSTM is a variant of the Recurrent Neural Network (RNN) commonly used for handling tasks related to time series data. When training a typical neural network model, the computation is usually represented as $s = f(W^T X + b)$, where W represents the weights, X is the input, and b is the bias term. A common LSTM structure is shown in Fig. 1.

x_t represents the input data at each time step, h_t is the output at each time step, and the intermediate c_t is the long-term memory between cells.

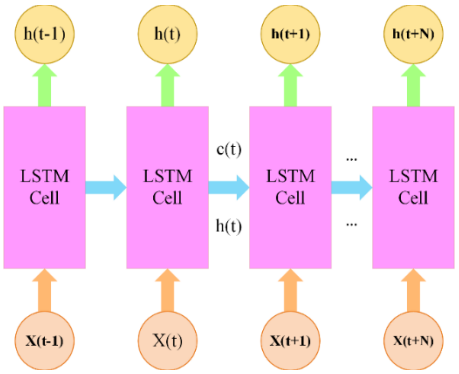

Fig. 1. LSTM structure diagram.

An LSTM includes a forget gate, an input gate, and an output gate. Each gate requires a computation similar to the one mentioned above, in addition to calculating the current state. Compared to the structure of an RNN, an LSTM cell is more complex, incorporating fully connected layers with activation functions in its gate structures. These gates output values between 0 and 1, indicating the degree to which feature information should be retained (0 means retain, 1 means discard). This gated structure enables LSTMs to maintain a stable gradient flow across extended time horizons while adaptively managing sequential information retention and suppression. The output of the forget gate is defined as f_t, the output of the input gate as i_t, the candidate value for memory update as $\bar{c}_t$, and the output of the output gate as o_t.

The forget gate f_t selects information from the previous cell state c_{t-1}:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

The input gate determines how much of the current network input x_t is to be saved in the cell state c_t:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

Input c_t:

$$\bar{c}_t = \tan h(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

The output gate controls how much of the cell state is output to the current LSTM output value h_t:

$$O_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_0) \quad (7)$$

W_f, W_i, W_c, W_0 are weight matrices, and b_f, b_i, b_c, b_0 are bias terms. $[h_{t-1}, x_t]$ denotes the concatenation of the hidden

state h_{t-1} and the input state x_t along the horizontal axis. σ represents the sigmoid function, and tanh represents the hyperbolic tangent function.

The LSTM model updates its weights using the backpropagation algorithm to minimize prediction error. The backpropagation process for LSTM mainly includes the following steps:

1) Definition of the loss function: Define a loss function based on the task type (e.g., classification or regression). Common loss functions include cross-entropy for classification tasks and mean squared error for regression tasks. This function measures the prediction error of the model.

2) Backpropagation to compute gradients: Use the Backpropagation Through Time (BPTT) algorithm to propagate the error gradient backward from the last time step to the first. Each gate (forget gate, input gate, output gate) in the LSTM network participates in the error propagation, gradually computing the gradient for each parameter. BPTT is suitable for sequential data, and due to the gating structure of LSTM, the vanishing gradient problem is alleviated to some extent.

3) Gradient clipping: Due to the potential for gradient explosion in long sequences, gradient clipping techniques limit the size of the gradients, ensuring they do not exceed a certain threshold. This prevents overly large gradients from causing unstable training.

4) Weight update: LSTM uses gradient descent to update the weight matrices, optimizing the model parameters to reduce the value of the loss function.

In practical applications, an LSTM learn effective patterns in sequence data through backpropagation and gradient updates. This training process enables LSTM to achieve good performance on various sequence prediction tasks, particularly excelling in handling long-term dependencies. By decoupling memory retention from hidden state updates, LSTMs achieve superior performance in learning complex temporal dynamics compared to conventional RNNs, particularly when handling intermittently sampled or irregularly spaced time series data common in multi-energy virtual power plant environments.

C. Sparrow Search Algorithm

The Improved Sparrow Search Algorithm introduces two main upgrades to the traditional SSA. First, it adopts a Good Lattice Points strategy for population initialization, replacing random placement to ensure sparrow positions are evenly distributed across the search space. This adjustment accelerates convergence and reduces the likelihood of solutions becoming trapped in local optima due to uneven initialization. Second, ISSA incorporates a Bidirectional Sine-Cosine Search Strategy to update discoverer positions. By alternating between forward and reverse search phases, the algorithm maintains a balance between broad exploration and refined exploitation. The forward phase steers individuals toward current optimal solutions, while the reverse phase drives exploration into uncharted regions, effectively curbing premature convergence. Together, these enhancements strengthen the algorithm's ability to tackle complex nonlinear challenges in load forecasting, delivering higher optimization accuracy and adaptability.

The Traditional Sparrow Search Algorithm initializes the sparrow population using a random distribution method, which results in randomness in the population distribution and can easily lead to SSA getting trapped in local optima. To address this issue, this study introduces an initialization strategy based on good lattice points for optimizing the initial positions of the sparrows, thereby improving the convergence accuracy and speed of the algorithm. The principle behind this approach is as follows: In the D-dimensional Euclidean space G_D , where $r \in G_D$, the good lattice point set $P_n(k) = \left(\left\{r_1^{(n)} \cdot k\right\}, \left\{r_2^{(n)} \cdot k\right\}, \ldots, \left\{r_D^{(n)} \cdot k\right\}\right)$ for $1 \le k \le n$,the discrepancy is $\varphi(n) = C(r,\varepsilon) n^{-1+\varepsilon}$, where $C(r,\varepsilon)$ is a constant related to r and ε , $P_n(k)$ represents the set of good lattice points. Additionally, in this study, $r = 2cos\left(\frac{2k\pi}{p}\right), 1 \le k \le n$, where p is the smallest prime number satisfying $\frac{(p-3)}{2} \ge D$ .Therefore, the initialization strategy based on good lattice points can be expressed as follows:

$$x_i(k) = \left(up_j - lp_j\right) \cdot \{P_n(k)\} + lp_j \tag{7}$$

The SSA is prone to premature convergence during the position update process, which can result in solutions getting trapped in local optima rather than finding the global optimum. To address this issue, a bidirectional sine cosine search strategy has been introduced. This strategy enhances exploration by defining both forward and reverse searches, thereby expanding the local search methods. In the forward search, the current individual is guided toward the current optimal individual, emphasizing the swarm's ability to autonomously approach the global optimum as iterations progress. Conversely, the reverse search encourages the swarm to explore regions beyond the current optimal area, demonstrating a stronger selectivity for uncharted territories. The initialization strategy based on Good Lattice Points ensures that the initial population is uniformly distributed in the high-dimensional search space, mitigating the risk of region omission caused by random initialization. This global coverage characteristic, combined with dynamic search direction adjustments, significantly reduces the algorithm's sensitivity to local optima and enhances global optimization efficiency in complex nonlinear problems. The enhanced position update formula for the discoverer is defined as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \left|\sin r_1\right| + r_2 \cdot \sin r_1 \cdot \left|x_1 \cdot x_{best}^t - x_2 \cdot x_{i,j}^t\right| \\ \qquad , R_2 < ST, R_1 \ge 0.5 \\ x_{i,j}^t \cdot \left|\cos r_3\right| - r_4 \cdot \cos r_1 \cdot \left|x_1 \cdot x_{best}^t - x_2 \cdot x_{i,j}^t\right| \\ \qquad , R_2 \ge ST, R_1 \ge 0.5 \\ x_{i,j}^t + Q \cdot L, \ldots R_2 \ge ST \end{cases} \tag{9}$$

In this algorithm, $x_{i,j}^{t}$ represents the position of the current generation sparrow, x_{best}^{t} while denotes the position of the best sparrow found so far at iteration t. $R_1, R_2 \in [0,1]$, where R_2 is the vigilance factor, and $ST \in [0.5,1]$ is the safety threshold. Q is a random number following a normal distribution, and L is a unit vector with $r_1, r_3 \in [0,2\pi]$ and $r_2, r_4 \in [0,\pi]$, which are used to determine the movement distance and search direction for the next generation of sparrows. The expressions $x_1 = -\pi + 2\pi(1-\tau)$ and $x_2 = -\pi + 2\pi$ define specific positions, where τ is the golden ratio coefficient. The optimization of the safety threshold (ST) and vigilance factor is grounded in multi-objective trade-offs and sensitivity analysis. The ST is determined through statistical analysis of historical experimental data, with its value balancing population convergence speed and escape capability: an excessively high ST would suppress the exploratory nature of reverse search, while an overly low ST may induce ineffective perturbations. The vigilance factor is configured based on problem dimensionality and solution space characteristics. Its dynamic range is adjusted through grid search and cross-validation to prioritize global exploration during early iterations and progressively strengthen local exploitation in later stages. The parameter optimization process integrates convergence validation and error sensitivity testing, ultimately yielding a parameter combination that ensures stable generalization capability across both training and validation sets.

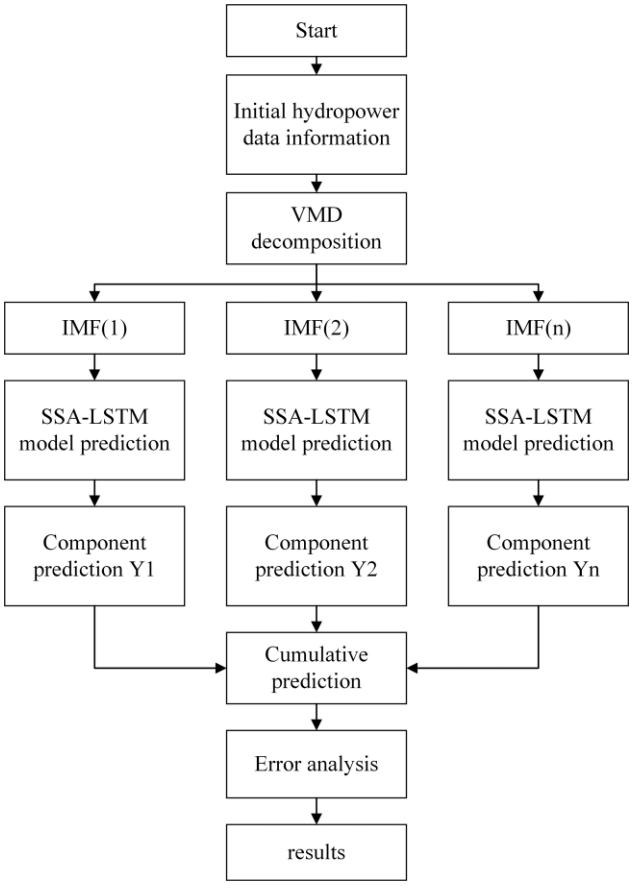

Fig. 2. Prediction process of the ISSA-VMD-LSTM model.

D. Short-Term Load Forecasting Framework for Hydropower

The forecasting framework proposed in this study aims to utilize historical hydropower data, achieving the goal of short-term load forecasting by conducting a comprehensive analysis of the correlations within these data. Fig. 2 illustrates the overall implementation flowchart of the hybrid forecasting model in load forecasting.

III. Case Study Analysis

A. Case Study Setup

The experiment was conducted on NVIDIA GeForce RTX 4060 Laptop GPU and implemented with Python 3.8. Data spanning from January 1, 2021, to May 1, 2024, was selected for analysis, with a sampling interval of one day. After integrating and screening the data, a comprehensive dataset consisting of 1,247 data points was obtained. To address noise and outliers in raw data, a sliding window technique integrated with the 3 principles is implemented for anomaly identification. Data points deviating beyond the mean ±3 standard deviations range are rectified using linear interpolation.

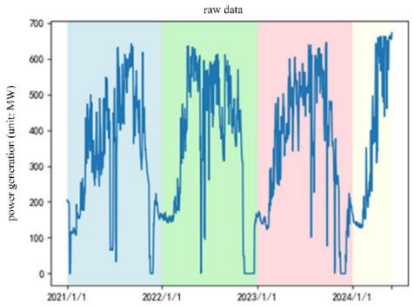

Fig. 3. Original hydropower load data.

Furthermore, consecutive data gaps resulting from equipment failures are systematically removed to ensure data integrity. Of these, 80% were allocated to the training set, while the remaining 20% were reserved for the test set to validate the final prediction outcomes of each model. In the VMD process, it is crucial to balance the number of decomposed modes to avoid excessive computational burden and time costs without compromising the signal's characteristic representation. Therefore, the number of modes, the K value, was set to 4. Additionally, the penalty factor α was configured to 2000, the frequency distribution initialization was set to 1, and the convergence tolerance was established at 1e-7. Fig. 3 illustrates the plot of the original data.

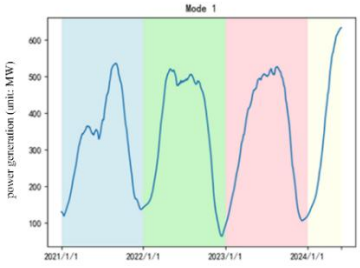

Fig. 4. Component of hydropower load decomposition of mode 1.

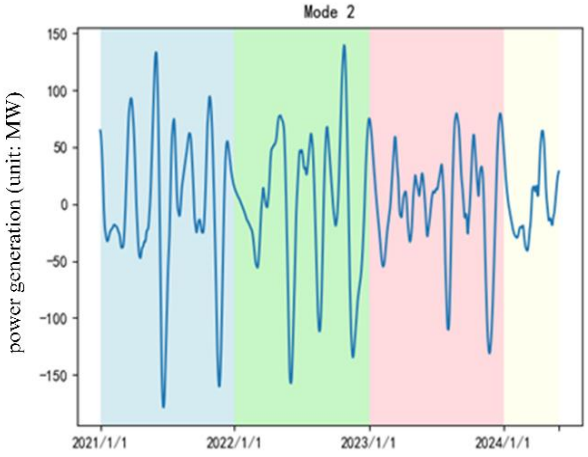

Fig. 5. Component of hydropower load decomposition of mode 2.

Fig. 4 represents the primary waveform trend, which shows an initial increase followed by a decrease in the hydropower load from 2021 to 2023. For the period from January to May 2024, the trend exhibits a growing pattern. Fig. 5 displays a more regular periodic variation, better reflecting the changes on a monthly and quarterly basis. The fluctuations in Fig. 6 are more intense, primarily capturing short-term variations. This mode reflects short-term demand changes or the impact of random events. The fluctuations in Fig. 7 are the most intense, capturing instantaneous changes.

B. Case Study Metrics Description

To comprehensively evaluate the performance of the proposed forecasting model, we utilize a combination of four metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). RMSE calculates the square root of the average of the squared differences between the predicted and actual values, which makes it sensitive to larger errors and provides a measure of error magnitude. MAE measures the average of the absolute differences between predictions and actual observations, offering a robust estimate of overall error without heavily penalizing large discrepancies. MAPE expresses the average of the absolute percentage errors, making it suitable for comparing datasets of different scales, although caution is needed due to its instability when actual values are close to zero. Finally, R^2 reflects the proportion of the variance in the dependent variable that is predictable from the independent variables, indicating how well the model fits the data compared to a simple mean prediction. These metrics together provide a thorough assessment of the model's accuracy and reliability.

In the formulas, y_g denotes the actual value of the g-th observation, $\hat{y}_g$ denotes the predicted value of the g-th observation, and G represents the total number of observations.

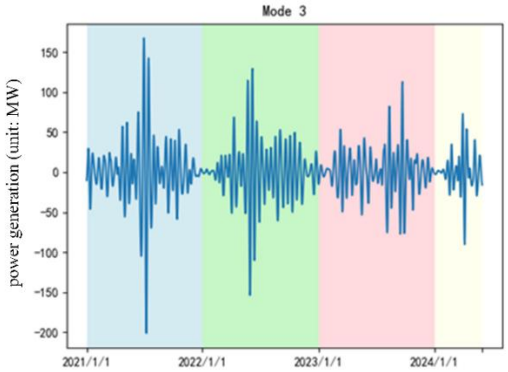

Fig. 6. Component of hydropower load decomposition of mode 3.

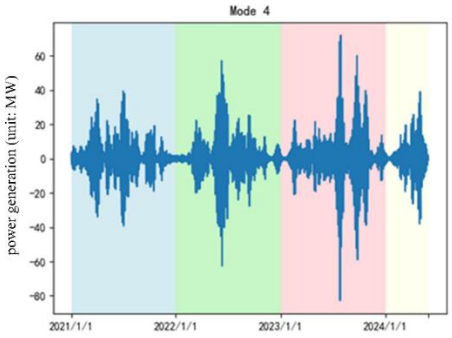

Fig. 7. Component of hydropower load decomposition of mode 4.

C. Result Analysis

Analysis of the data in Table I shows that the ISSA-VMD-LSTM algorithm outperforms all other algorithms across all evaluation metrics, achieving the highest R^2 and the lowest RMSE, MAE, and MAPE. This indicates that the ISSA-VMD-LSTM algorithm, as mentioned in this study, exhibits superior prediction performance with greater accuracy and stability compared to other algorithms.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

Algorithm	R2	RMSE	MAE	MAPE	Time Complexity
LSTM	0.652	0.235	0.201	0.049	O(T*N2)
PSO-LSTM	0.752	0.215	0.183	0.042	O(P*I*T*N2)
PSO-VMD-LSTM	0.851	0.176	0.133	0.031	O(K*N2+P*I*T*N2)
SSA-VMD-LSTM	0.802	0.183	0.142	0.035	O(K*N2+S*I*T*N2)
ISSA-VMD-LSTM	0.934	0.124	0.105	0.018	O(K*N2+S*I*T*N2+C)

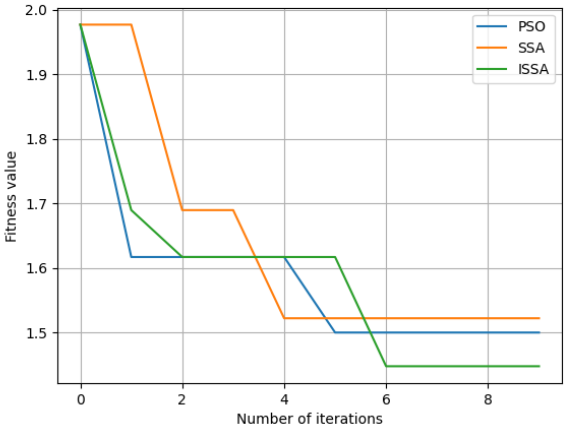

Fig. 8. Fitness curve of algorithms.

Table II lists the optimal parameters for each mode.

TABLE II. OPTIMAL PARAMETERS FOR MODES

Number of Modes	epochs	batch size	LSTM 1 of layer neurons	LSTM 2 layer of neurons	Dense Layer of neurons	learning rate	Number of Modes
IMF1	47	22	54	53	43	0.0021	IMF1
IMF2	20	12	46	67	87	0.0060	IMF2
IMF3	46	43	65	84	47	0.0050	IMF3
IMF4	14	56	77	63	79	0.0055	IMF4

Comparing the individual algorithms, as shown in Fig. 8, it can be observed that the ISSA achieves the lowest final fitness value, indicating its superior optimization performance. In contrast, PSO has a relatively higher final fitness value, while SSA exhibits the highest final fitness value. Additionally, we note that ISSA converges at the fastest rate. Although PSO and SSA also exhibit rapid convergence in the early stages, their final fitness values are higher, suggesting that their optimization effectiveness is not as good as that of ISSA. The optimization process involves tuning a total of 24 parameters across four modes, including the number of iterations, batch size, the number of output neurons in the first LSTM layer, the number of output neurons in the second LSTM layer, the number of neurons in the fully connected layer, and the learning rate. The parameter optimization iteration process is illustrated in Fig. 9.

To address potential instability in MAPE metrics when actual load values approach zero - a scenario that may cause abnormal error amplification and distort model performance evaluation - we implemented a threshold adjustment strategy. Specifically, observations with actual values below 1% of the dataset's average load are excluded from MAPE calculation. This approach maintains the metric's interpretability while eliminating outlier distortion caused by near-zero measurements. For enhanced robustness, we complement the evaluation with SMAPE (Symmetric Mean Absolute Percentage Error), which constrains results within a 0%-200% range through symmetrical error calculation, effectively mitigating sensitivity to low-value measurements. Cross-validation confirmed consistent results from both metrics, ensuring reliable model performance assessment.

The ISSA-VMD-LSTM model demonstrated strong predictive performance with a validated R^2 value of 0.934. To assess estimation stability, we conducted 1,000 bootstrap iterations, yielding a narrow 95% confidence interval of [0.921, 0.943] for the R^2 metric. Furthermore, 5-fold cross-validation was implemented to evaluate model generalization, producing a mean R^2 of 0.927 ±0.006 - statistically consistent with the test set performance. These rigorous validation procedures confirm both the model's predictive accuracy and its robustness against overfitting risks.

The ISSA-VMD-LSTM algorithm proposed in this study demonstrates significant improvements over the traditional LSTM algorithm, with an increase of 30.2% inR^2, a reduction of 47.2% in RMSE, a decrease of 47.8% in MAE, and a lowering of 63.3% in MAPE. In terms of optimization strategies, the ISSA introduces a bidirectional search mechanism and initialization improvements. Compared with traditional optimization algorithms like PSO and SSA, the proposed algorithm demonstrates enhanced balancing capability between global exploration and local exploitation, effectively avoiding local optima traps while accelerating convergence. When compared with Transformer models, our method exhibits superior performance in large-scale data scenarios and achieves high-precision prediction in small-scale data through decomposition and optimization strategies, making it more applicable to data-constrained environments. In contrast to lightweight GRU models, this approach reduces hyperparameter sensitivity through signal decomposition mechanisms and demonstrates stronger prediction stability in multi-energy coupled environments. The algorithm presented in this study substantially enhances all evaluation metrics, markedly improving model fit while further reducing errors and increasing prediction accuracy.

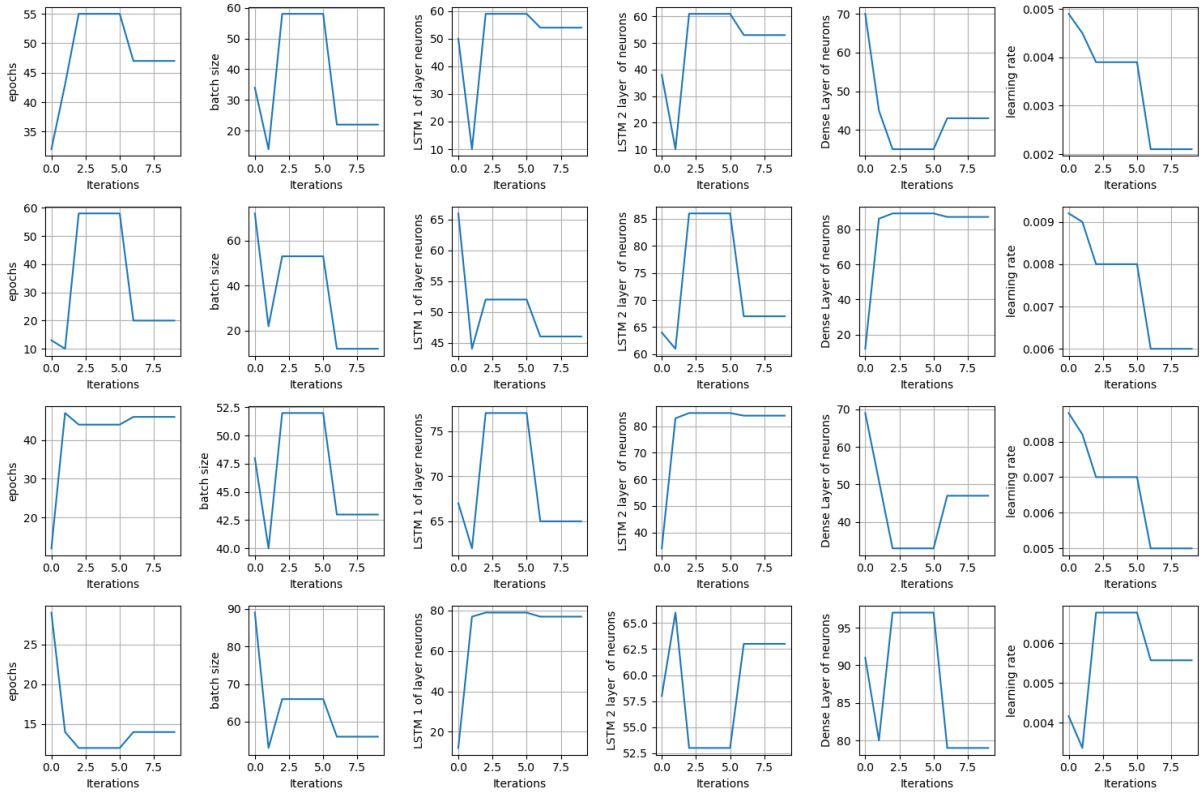

Fig. 9. Parameter optimization iteration process.

IV. Conclusion

This study introduces a short-term hydropower load forecasting method based on the ISSA-VMD-LSTM model. Given that hydropower load data typically exhibit complex nonlinear characteristics, VMD is employed to decompose the raw data into multiple intrinsic mode functions, effectively extracting features such as long-term trends and seasonal variations. This decomposition enhances the model's adaptability to intricate time-series data. To optimize the LSTM hyperparameters, an improved version of the Sparrow Search Algorithm is utilized, which incorporates a bidirectional cosine search strategy that combines forward and reverse searches. This approach expands the local search range, effectively preventing the traditional SSA from getting trapped in local optima, thereby significantly enhancing the optimization of LSTM hyperparameters. For the modal components obtained from VMD decomposition, the ISSA-optimized LSTM model further refines prediction accuracy. Experimental results demonstrate that the ISSA-VMD-LSTM model surpasses traditional LSTM models and other hybrid models across all evaluation metrics, showcasing superior forecasting performance.

While the proposed ISSA-VMD-LSTM model demonstrates significant advantages in short-term hydroelectric load forecasting, its practical application still faces limitations. First, the computational complexity of the VMD decomposition and ISSA optimization processes remains high, particularly in multi-modal data scenarios, which may prolong training and inference times, potentially hindering real-time requirements. Second, the model's sensitivity to hyperparameters necessitates reliance on empirical expertise or iterative parameter tuning, increasing deployment costs. Future improvements could focus on simplifying the algorithm and enhancing efficiency. For instance, adopting parallel computing or streamlined decomposition strategies to reduce computational overhead. Exploring adaptive parameter optimization mechanisms or integrating edge computing frameworks could improve real-time processing capabilities. Incorporating incremental learning or transfer learning techniques may further enhance the model's adaptability to dynamic multi-energy coupled environments, strengthening its generalization and practical utility.

Looking ahead, Multi-Energy Coupled Virtual Power Plants are expected to play a crucial role in short-term hydropower load forecasting. MC-VPPs represent the deep integration of Energy Internet technology, combining various energy resources through intelligent optimization and scheduling. They provide more accurate and reliable forecasting outcomes, enhancing the flexibility and efficiency of power systems. Moreover, MC-VPPs effectively address the intermittency and uncertainty associated with renewable energy sources, promoting the widespread adoption of clean energy and supporting the healthy development of electricity markets. As technological advancements continue and practical applications deepen, MC-VPPs will drive the energy system toward greater intelligence and efficiency, ensuring a more stable and sustainable power supply.

ACKNOWLEDGMENT

This work was supported by the Science and Technology Project of China Southern Power Grid Co., Ltd., under Grants YNKJXM20230530.

REFERENCES

[1] Guo H, Zhu R, Liu Z, et al. Distributed Robust Optimization Strategy for Multi-Energy Virtual Power Plant Clusters . Journal of Shaanxi University of Science & Technology, 2024, 42(06): 199-207. DOI: 10.19481/j.cnki.issn2096-398x.2024.06.023.

[2] Liu Jinpeng, Hu Guosong, Peng Jinchun, et al. Low-carbon-economic-robust optimal scheduling of multi-energy complementary virtual power plants . Proceedings of the CSEE, 2024, 44(24): 9718-9731.DOI:10.13334/j.0258-8013.pcsee.242431.

[3] Wang Qianchun, Du Xinhui, Wu Yingying, et al. Optimal scheduling strategy for multi-energy complementary virtual power plants considering carbon trading . Electrical Measurement & Instrumentation, 2024, 61(11): 22-30.DOI:10.19753/j.issn1001-1390.2024.11.003.

[4] Wei Xuan, Pan Zhaoguang, Wang Bin, et al. Review and Prospect of Resource Aggregation and Cooperative Regulation for Virtual Power Plant Under Cloud-Pipe-Edge-End Architecture [J]. Journal of Global Energy Interconnection, 2020, 3(6): 539-551. DOI: 10.19705/j.cnki.issn2096-5125.2020.06.001.

[5] Li Fangneng, Liang Yifeng, Xu Jiangning, et al. BDS satellite clock bias prediction based on robust dual-parameter exponential smoothing method . Journal of Chinese Inertial Technology, 2024, 32(07): 645-653.DOI:10.13695/j.cnki.12-1222/o3.2024.07.002.

[6] Zhang Yaxin, Ru Fangjie, Yan Jianpeng. Research on railway automatic control algorithm integrating improved PSO and grey prediction fuzzy PID [J]. Automation and Instrumentation, 2025, (11): 74-77+82. DOI: 10.14016/j.cnki.1001-9227.2025.11.074.

[7] Huang Haihong, Chen Zhao, Wang Haixin. Improved grey prediction sliding mode current control for fast control power supply of all-superconducting tokamak nuclear fusion power generation device [J]. Transactions of China Electrotechnical Society, 2025, 40(16): 5317-5329. DOI: 10.19595/j.cnki.1000-6753.tces.241221.

[8] Mao Cuiwei, Gou Xiaoyi, Tuo Yibo, et al. Prediction of development scale of higher vocational education in China based on real-order grey prediction model .Journal of Fudan University (Natural Science), 2025, 64(01): 14-23.DOI:10.15943/j.cnki.fdxb-jns.2025.01.010.

[9] Song Tengjiao, Liu Ying, Wang Yan, et al. Correlation of physical and mechanical parameters of cohesive soil based on multiple linear regression—A case study of the central urban area of Changchun [J]. Journal of Jilin University (Earth Science Edition), 2025, 55(4): 1225-1239. DOI: 10.13278/j.cnki.jjuese.20230321.

[10] Li N, Jiang T, Sui X, et al. A Combined Forecasting Method for Short-Term Electric Load Based on Time-Frequency Scale . Power System Protection and Control, 2024, 52(13): 47-58. DOI: 10.19783/j.cnki.pspc.231284.

[11] Li Mugong, Yan Chunguang, Sun Hairong, et al. Research on performance early warning method for EMU batteries based on multiple linear regression [J]. Electric Drive for Locomotives, 2025, (4): 130-136. DOI: 10.13890/j.issn.1000-128X.2025.04.016.

[12] Ji Changzheng, Wan Ren, Shi Zhaochong, et al. Prediction of speed of sound in ionic liquids based on multiple linear regression and back propagation artificial neural network [J]. Journal of East China University of Science and Technology (Natural Science Edition), 2025, 51(2): 158-165. DOI: 10.14135/j.cnki.1006-3080.20240417001.

[13] Ming Liang, Chen Jixin, Zhang Zhengtao, et al. Research on dynamic characteristics and control strategies of gas turbines under hydrogen-enriched conditions [J]. Journal of Engineering for Thermal Energy and Power, 2025, 40(12): 30-43+52. DOI: 10.16146/j.cnki.rndlgc.2025.12.003.

[14] Zhao Chenhui, He Shan, Liu Xianming, et al. Research progress on application of neural network-based optimization algorithms in EDA .Application Research of Computers, 2025, 42(01): 1-10.DOI:10.19734/j.issn.1001-3695.2024.05.0171.

[15] Zheng Zhou. Robust scheduling algorithm for WNCS based on mixed-integer nonlinear programming .Machinery Design & Manufacture, 2024, (12): 95-103.DOI:10.19356/j.cnki.1001-3997.20240516.022.

[16] Li Xiangdong, Sun Wanquan. Nonlinear coupled dynamics of hydroelectric unit shaft system based on fractional-order theory .Water Resources and Power, 2025, 43(01): 187-191.DOI:10.20040/j.cnki.1000-7709.2025.20240152.

[17] Luo Guanghao, Cui Mingjian, Han Yining, et al. Tensor Decomposition-Based Personalized Federated Learning Method for Ultra-Short-Term Power Load Prediction Considering Low Communication Costs [J/OL]. Transactions of China Electrotechnical Society, 1-19 [2026-01-19].DOI:10.19595/j.cnki.1000-6753.tces.251325.

[18] Wang Y, Cao Y, Sun J. Short-Term Electric Load Forecasting Based on Multi-Strategy Improved Golden Jackal Algorithm Optimized LSTM . Power System Protection and Control, 2024, 52(14): 95-102.DOI: 10.19783/j.cnki.pspc.231431.

[19] Peng Bo, Li Lijun, Gao Zicheng, et al. Trajectory planning of robotic arm for camellia fruit picking vehicle based on improved golden jackal optimization algorithm [J]. Journal of Mechanical Transmission, 2025, 49(12): 114-122. DOI: 10.16578/j.issn.1004.2539.2025.12.014.

[20] Li Dandan, Zhu Shilei, Li Zhongkang, et al. Improved golden jackal optimization algorithm based on multi-strategy fusion [J]. Transducer and Microsystem Technologies, 2025, 44(01): 127-130. DOI: 10.13873/J.1000-9787(2025)01-0127-04.

[21] Chang Ling, Wang Bo, Meng Bin, et al. Research on power generation prediction of wind-solar complementary system based on PSO-LSTM-Attention [J]. Journal of Engineering for Thermal Energy and Power, 2025, 40(12): 171-179. DOI: 10.16146/j.cnki.rndlgc.2025.12.016.

[22] Shi Xin, Tang Jia, Fan Zhirui, et al. Two-stage compensation control strategy based on overshoot prediction-optimization for lower limb exoskeleton motor [J/OL]. Chinese Journal of Scientific Instrument, 1-15 [2026-01-19]. DOI: 10.19650/j.cnki.cjsi.J2514346.

[23] Xi Tao, Wang Long, Wang Lijing. Fault Diagnosis of Variable Load Rolling Bearings Based on SABO-VMD-IWT and LSTM . Modular Machine Tool & Automatic Manufacturing Technique, 2025, (03): 192-198.DOI: 10.13462/j.cnki.mmtamt.2025.03.037.

[24] Guo Min. Research on Dam Deformation Monitoring Using LSTM Based on Genetic Algorithm . Shaanxi Water Resources, 2025, (03): 175-177+180.DOI: 10.16747/j.cnki.cn61-1109/tv.2025.03.007.

[25] Li Shengqing, Gao Zehua, Qiao Jingxiao, et al. Optimal Capacity Configuration of Microgrid Based on Improved Sparrow Search Algorithm . Acta Energiae Solaris Sinica, 1-10 [2025-04-01].DOI : 10.19912/j.0254-0096.tynxb.2024-0968.

[26] Wang Fanrong, Wang Junhan, Jiang Junjie. Transformer Fault Diagnosis Using BP Neural Network Optimized by Improved SSA . Modern Electronics Technique, 2025, 48(04): 145-150.DOI: 10.16652/j.issn.1004-373x.2025.04.025.

[27] Zhu Yifan, Qiu Xudong, Li Shuo. Path Planning for Quadrotor UAV Based on Improved Sparrow Search Algorithm . Sensor and Microsystem, 2025, 44(01): 141-145+151. DOI: 10.13873/J.1000-9787(2025)01-0141-05.

[28] Yang Yi, Zhao Jingtao, Fu Guoqiang. Temperature Prediction of Geothermal Reservoirs Using PSO-LSTM Model . Journal of Mining Science and Technology, 2024, 9(04): 538-548.DOI:10.19606/j.cnki.jmst.2024.04.006.\

[29] Zhou Jianxin, Wang Hongtao. Short-Term Wind Power Output Prediction Based on Similar Day Theory and Improved PSO-LSTM Model . Chinese Journal of Construction Machinery, 2024, 22(04): 447-451+457.DOI:10.15999/j.cnki.311926.2024.04.001.

[30] Wang Shuai, Shen Jiewen, Xu Bin, et al. Dynamic ampacity prediction method for overhead transmission lines based on SSA-VMD-LSTM [J]. Electronic Measurement Technology, 2025, 48(19): 115-125. DOI: 10.19651/j.cnki.emt.2518236.