Universe Scientific Publishing

Clean Energy Science and Technology 2025, 3(1), 297.
https://doi.org/10.18686/cest297

Article

Short-term forecasting of building heating load based on MVMD-SSA-LSTM

Bo Zhou[1,*], Yueyi Zhang[1], Chaoyang Fei[1], Xiuming Li[2], Zhigang Xie[1], Daohe Li[1]

[1] School of Architecture and Civil Engineering, Shenyang University of Technology, Shenyang 110870, China
[2] School of Metallurgy, Northeastern University, Shenyang 110819, China
*** Corresponding author:** Bo Zhou, liguodapple@sina.com

CITATION

Zhou B, Zhang Y, Fei C, et al. Short-term forecasting of building heating load based on MVMD-SSA-LSTM. Clean Energy Science and Technology. 2025; 3(1): 297. https://doi.org/10.18686/cest297

ARTICLE INFO

Received: 17 December 2024
Accepted: 31 January 2025
Available online: 5 March 2025

COPYRIGHT

Copyright © 2025 by author(s). *Clean Energy Science and Technology* is published by Universe Scientific Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: A short-term heating load forecast for buildings is a critical step in the subsequent control of energy systems, directly impacting system energy consumption. However, given that heating load and its influencing factors constitute volatile time series data, noise interference within the data significantly limits prediction accuracy and stability. To address this issue, this paper proposes a novel MVMD-SSA-LSTM model for building heating load forecasts, which integrates Multivariate Variational Mode Decomposition (MVMD), Sparrow Search Algorithm (SSA), and Long Short-Term Memory (LSTM) neural networks. Initially, a correlation analysis of the factors influencing building heating load is conducted to identify the key determinants. Subsequently, MVMD is employed to decompose the multidimensional dataset into several modes. A correlation analysis is then performed on these decomposed modes to extract supplementary features, which are combined with the original data to form a new dataset, thereby reducing feature redundancy. Finally, an LSTM neural network is utilized as the core predictive model, with the SSA algorithm optimizing three critical parameters: The maximum training iterations, the number of hidden units, and the initial learning rate. The predicted outputs of each heating load mode are aggregated to obtain the final forecast. Results demonstrate that the MVMD-SSA-LSTM model effectively mitigates the uncertainty in heating load sequence forecasts, overcoming noise disturbances and exhibiting superior performance compared to other commonly used models, with significantly higher accuracy and stability.

Keywords: building heating load forecast; multivariate variational mode decomposition; long short-term memory neural networks; sparrow search algorithm

1. Introduction

Building load forecasting is a critical foundation for the regulation and control of central air conditioning systems, with forecasting results at varying temporal granularities playing distinct roles in system planning. Based on the forecasting horizon, building load predictions can be broadly classified into short-term (1 h to 1 week), medium-term (1 week to several weeks), and long-term (several months to years) [1]. Building energy systems inherently exhibit significant thermal inertia due to the accumulation and release of heat, which introduces substantial time delays in system responses to load changes. This time lag is particularly pronounced in winter compared to summer [2–4]. Therefore, it is necessary to regulate in advance to ensure the efficiency and stability of operation. High-precision short-term heating load forecasting is instrumental in optimizing control parameters for various components within energy systems, playing a crucial role in ensuring stable and energy-efficient operation. Specifically, short-term load forecasting results can provide essential data support for daily system startup and shutdown schedules, as well as for the formulation of control strategies for various units. The accuracy of building load forecasts directly influences the actual energy consumption of the system and impacts occupant comfort [5,6]. In addition, real-time building heating load data cannot be readily obtained solely from current environmental measurements, making accurate forecasting a vital approach for addressing this challenge effectively.

In recent years, deep learning has emerged as the most widely utilized and rapidly advancing approach for addressing building load forecasting problems in both domestic and international research [7–9]. For instance, Pan et al. [10] applied a Back Propagation Neural Network (BPNN) to predict building energy loads, while Yildiz et al. [11] employed Support Vector Regression (SVR) for building energy consumption forecasting. Additionally, Wang et al. [12] leveraged Long Short-Term Memory (LSTM) networks to predict heating loads in buildings. Given the unique architecture of LSTM models, they have shown exceptional performance in short-term forecasting of heating and cooling loads for buildings. For example, Zhou et al. [13,14] developed an air-conditioning system load forecasting model for a public building using LSTM on an hourly basis and compared the results with other deep learning algorithms, such as Autoregressive Moving Average (ARMA) models and BPNN. The findings demonstrated that the LSTM model produced more accurate predictions. Furthermore, Chalapathy et al. [15] applied the LSTM neural network model to the prediction of cooling loads, with experimental results consistently outperforming those of the SVR algorithm. These studies indicate that LSTM is widely recognized as an effective model for such applications. Leveraging historical data, deep learning methods can effectively uncover the hidden coupling relationships among input parameters and capture temporal correlations, thereby accurately reflecting power fluctuation characteristics. Compared to traditional physical models, deep learning not only saves computational resources but also yields superior prediction accuracy [16–18].

Moreover, the parameters of neural networks are typically determined based on empirical computation. In recent years, numerous swarm intelligence algorithms have been developed and applied to optimize key parameters of LSTM networks, aiming to enhance prediction accuracy. The optimization principle of swarm intelligence algorithms often draws inspiration from bionics, simulating the behavioral characteristics of social animal groups to search for optimal solutions. Among these algorithms, the Sparrow Search Algorithm (SSA), introduced by Xue et al. [19] in 2020, has demonstrated remarkable performance in multi-dimensional parameter optimization problems. Compared to other algorithms such as Particle Swarm Optimization (PSO) and genetic algorithms, SSA features fewer control parameters, higher efficiency, a simpler structure, and stronger global search capability [20,21]. SSA-LSTM prediction models have been successfully employed across various research domains. For instance, Jiedeerbieke et al. [22] utilized SSA for predicting deformation in concrete dams, Huang et al. [23] applied it to short-term photovoltaic power forecasting, and Gao et al. [24] adopted the model for short-term wind power prediction. In a study by Han et al. [25], PSO and SSA were used to optimize BP, LSTM, and RNN networks for electric load forecasting, demonstrating that SSA outperformed other algorithms in terms of solving speed, stability, and convergence accuracy. LSTM networks are highly effective in addressing various building load prediction tasks. When employing SSA to optimize LSTM hyperparameters, the objective function of SSA is defined as the prediction error of an untrained LSTM model on training data. SSA refines the hyperparameters through an iterative search process until the parameters meet the configuration requirements or convergence criteria.

$$\int_{\tau_0}^{\tau_1} Q_n(\tau)d\tau = \int_{\tau_0}^{\tau_1} \{-[KF\alpha(\beta+N)+0.278C_pV\rho_w]\times t_z(\tau)+[KF\alpha(\beta+N)+0.278C_pV\rho_w]\times t_n(\tau)\}d\tau \tag{1}$$

Despite the notable predictive performance of neural networks, practical applications reveal that predictions from a single neural network model often exhibit variability, characterized by significant differences in accuracy across experimental trials. This uncertainty primarily arises from factors such as random initialization during training, variations in gradient descent paths, and the influence of random noise in the training samples. These observations suggest that single neural networks are particularly sensitive to input noise, which exacerbates prediction uncertainty to some extent [26,27].

Furthermore, based on the calculation formula for building heating load, as presented in Equation (1), the primary factors influencing building heating load include indoor and outdoor temperatures, envelope structures, wind speed, wind direction, and solar radiation. However, external humidity indirectly increases the heating load demand by introducing latent heating requirements, and occupants contribute to both sensible and latent heat gains. Additional random factors, such as equipment usage patterns, occupant movements, and transient changes in material properties, also contribute to stochastic noise in heating load data [28,29].

To mitigate the impact of noise and improve the accuracy of predictive models, it is essential to account for the volatility and time-series nature of building load data and its influencing factors. Signal decomposition methods are effective in reducing noise interference in building load data. By applying modal decomposition techniques, feature data can be finely decomposed into Z components. These components are then analyzed to extract key features for prediction, leading to enhanced model accuracy. This approach has been shown to effectively reduce noise and improve the precision of load forecasting models [30,31].

Traditional approaches to decomposing predictive data commonly employ algorithms such as wavelet transform (WT), empirical mode decomposition (EMD), and variational mode decomposition (VMD). WT is known for its excellent time-frequency localization properties; however, its denoising performance diminishes when applied to signals with high noise levels. As a result, WT is often less effective for multidimensional data influenced by multiple factors, where substantial noise is inherently present in the data [32,33]. In contrast, EMD requires minimal prior knowledge and parameter settings, making it a straightforward technique to decompose signals into intrinsic mode functions (IMFs). However, EMD is highly sensitive to noise, with minor perturbations in the signal potentially causing significant inaccuracies in the decomposition results [34]. Furthermore, the lack of a well-defined mathematical framework in EMD increases the likelihood of mode mixing, particularly when dealing with complex signals.

VMD, introduced by Dragomiretskiy et al. [35] in 2014, is an adaptive, non-iterative signal processing method designed to address mode mixing issues. By optimizing the center frequency and bandwidth of each mode, VMD ensures non-overlapping frequency components across modes, leading to clearer and more stable decomposition results [36,37]. The robust mathematical foundation of VMD further enhances its reliability in signal decomposition tasks. Despite these advantages, applying VMD to prediction problems with multiple influencing factors often necessitates separate decomposition of data from individual channels. This process can overlook the coupling relationships between heating load and other variables in the time and frequency domains. Additionally, separate decomposition may result in frequency mismatch across sub-modes derived from different channels. Consequently, VMD in its standard form is less suited for handling multichannel data that includes heating load and their influencing factors. To address these limitations, VMD requires enhancement to support multichannel data processing, leading to the development of Multivariate Variational Mode Decomposition (MVMD), which offers a more comprehensive solution for multichannel signal analysis.

In order to address the above problem of heating load forecasting with the coupled influence of multiple factors and the accompanying noise interference, this paper proposes a combined MVMD-SSA-LSTM short-term heating load forecasting model for buildings, which firstly uses MVMD to modal decompose the multidimensional dataset, conducts correlation analysis, constitutes a new dataset with the original data, and then uses the SSA optimization algorithm combined with the LSTM neural network to make a forecasting, and compared with other commonly used basic forecasting models, according to the simulation results, the forecasting accuracy of this hybrid model proposed in this paper is better than other comparison models.

2. MVMD-SSA-LSTM model for short-term building heating load forecasting

In the present investigation, a Multivariate Variational Mode Decomposition (MVMD) technique is combined with a Long Short-Term Memory (LSTM) neural network to develop a model for short-term building heating load forecasting. Additionally, the Sparrow Search Algorithm (SSA) is incorporated to optimize the forecasting process. The framework of the proposed model is presented in **Figure 1**.

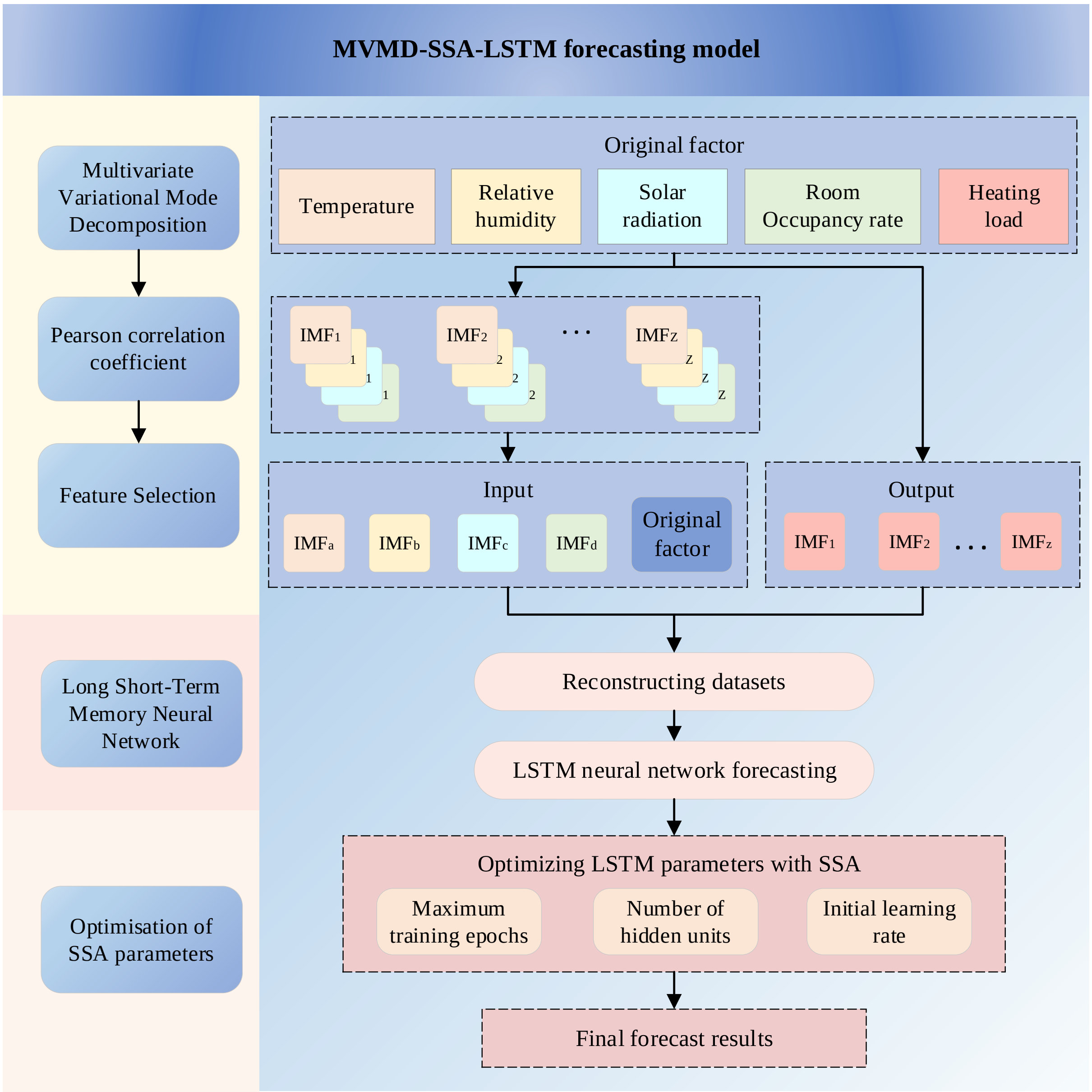


Figure 1. Forecasting structure of the MVMD-SSA-LSTM model.

2.1. Multivariate variational mode decomposition (MVMD)

To effectively extract data features and reduce the instability of time-series data, this section proposes a Multivariate Variational Mode Decomposition (MVMD) framework model based on the VMD algorithm. The framework is designed to decompose building load data and its influencing factors. Its principle involves decomposing a complex signal into sub-signals with fixed bandwidths, referred to as intrinsic mode functions (IMFs) [38]. Additionally, to identify features with strong correlations to the load data, each IMF is subjected to Pearson correlation analysis to compute the correlation coefficients of the decomposed components. This step minimizes feature redundancy, enabling the extraction of input features for load prediction. The features selected through modal decomposition, combined with the original influencing factors, are used as inputs for load prediction. As a generalized extension of the original VMD algorithm, MVMD processes input data containing C

data channels, which can be denoted by $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_C(t)]$, and further decomposes them into Z predefined multivariate modulated oscillations $\mathrm{u}_z(t)$.

$$\mathrm{x}(t) = \sum_{z=1}^{Z} \mathrm{u}_z\,(t)a \tag{2}$$

where, $\mathrm{u}_z(t) = [\mathrm{u}_1(\mathrm{t}), \mathrm{u}_2(\mathrm{t}), \ldots \mathrm{u}_C(\mathrm{t})]$.

From Equation (2), the multivariate modulated oscillation set $\{\mathrm{u}_z(t)\}_{z=1}^{Z}$ is obtained. The Hilbert transform is applied to each component of $\mathrm{u}_z(t)$, resulting in the vector analytic signal $\mathrm{u}_+^z(t)$. For the multivariate modulated oscillation, the model assumes a common component ω_z exists across all data channels. To ensure that the frequency components of each modal function are concentrated around the predetermined center frequency ω_z, harmonic mixing is performed by multiplying with a complex exponential $e^{-j\omega_z t}$ based on the Fourier transform principle. The bandwidth of the multivariate oscillation is then calculated using the Frobenius norm of the resulting matrix, as expressed in Equation (3):

$$f = \sum_{z}\sum_{c}\left\|\partial_t\left[u_+^{z,c}(t)e^{-j\omega_z t}\right]\right\|_2^2 \tag{3}$$

where, $u_+^{z,c}(t)$ represents the analytic signal vector corresponding to the c-th channel and the z-th mode.

The constraints of MVMD are as follows: The total bandwidth of all modes is minimized; the decomposed modes can accurately reconstruct the original data signal. The constrained optimization problem can be expressed as shown in Equation (4):

$$\underset{\{u_{z,c}\},\{\omega_z\}}{\text{minimize}}\left\{\sum_{z}\sum_{c}\left\|\partial_t\left[u_+^{z,c}(t)e^{-j\omega_z t}\right]\right\|_2^2\right\}\text{subject to}\sum_{z} u_{z,c}\,(t) = x_c(t), c = 1,2,\ldots,C \tag{4}$$

For the aforementioned multiple linear equality constraints, to facilitate solving, the original optimization problem is reformulated into an unconstrained problem by augmenting the Lagrangian function based on the number of channels. This reformulation simplifies the solution process, with the augmented Lagrangian function expressed in Equation (5):

$$\mathrm{L}\left(\{u_{z,c}\},\{\omega_z\},\lambda_c\right) = \alpha\sum_{z}\sum_{c}\left\|\partial_t\left[u_+^{z,c}(t)e^{-j\omega_z t}\right]\right\|_2^2 + \sum_{c}\left\|x_c(t) - \sum_{z}u_{z,c}(t)\right\|_2^2 + \sum_{c}\left\langle\lambda_c(t), x_c(t) - \sum_{z}u_{z,c}(t)\right\rangle \tag{5}$$

For the resulting unconstrained optimization problem, the alternate direction method of multipliers (ADMM) is employed to update both the modes and the center frequencies. Mode updating involves iteratively refining the modal functions in each step, as shown in Equation (6):

$$\hat{u}_{z,c}^{z+1}(\omega) = \frac{\hat{x}_c(\omega) - \sum_{i\neq z}\hat{u}_{i,c}(\omega) + \frac{\hat{\lambda}_c(\omega)}{2}}{1 + 2\alpha(\omega - \omega_z)^2} \tag{6}$$

The center frequency update adjusts the frequency characteristics of the modal functions to better adapt to the frequency-domain features of the original signal, thereby improving the accuracy of decomposition. The function for updating center frequencies is provided in Equation (7):

$$\omega_z^{n+1} = \frac{\sum_c \int_0^\infty \omega \, |\hat{u}_{z,c}(\omega)|^2 d\omega}{\sum_c \int_0^\infty |\, \hat{u}_{z,c}(\omega)|^2 d\omega} \tag{7}$$

Following the aforementioned operational steps, the data for each factor (data from *C* channels) can be decomposed into *Z* modes (*Z* IMFs). The modal frequency scales within the same layer are identical across all channels, ensuring frequency consistency between different channels.

To enhance the predictive accuracy of the model, the decomposed modes are subjected to Pearson correlation coefficient analysis. This analysis identifies the mode with the highest correlation coefficient for each channel (temperature, humidity, solar radiation, etc.) as a supplementary feature for heating load prediction inputs. The formula for calculating the Pearson correlation coefficient (PCC) [39,40] is provided in Equation (8):

$$\gamma_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{8}$$

where vectors $\overline{X}$ and $\overline{Y}$ represent the mean values of vectors X and Y, respectively.

The Pearson correlation coefficient is a method used to measure the linear correlation between two variables, X and Y, with a range of [−1, 1]. When the value approaches 1 or −1, the correlation between X and Y becomes stronger. Conversely, when the value approaches 0, the correlation between the two variables becomes weaker.

2.2. Long short-term memory neural network (LSTM)

Building heating load prediction refers to forecasting the heating demand required by a system at a future time during the building's operational phase. This prediction facilitates optimizing the control of air conditioning systems by determining optimal setpoints or operational conditions, thereby enabling the formulation of the best operation strategies for air conditioning systems. The LSTM neural network demonstrates exceptional performance in handling time-series data; it consists of an input layer, a hidden layer, and an output layer. The hidden layer is composed of specialized LSTM units, each of which includes an independent memory cell and three gates—input gate, output gate, and forget gate. Each gate is equipped with its own weights and biases, which collectively determine the storage and flow of information.

Compared to other commonly used neural networks, the unique architecture of LSTM overcomes the gradient vanishing and explosion issues often observed during the training of traditional RNNs and other single neural networks [41]. It also enables long-term information retention, making it particularly well-suited for time-series data prediction.

The internal mechanism of LSTM is illustrated in **Figure 2**.

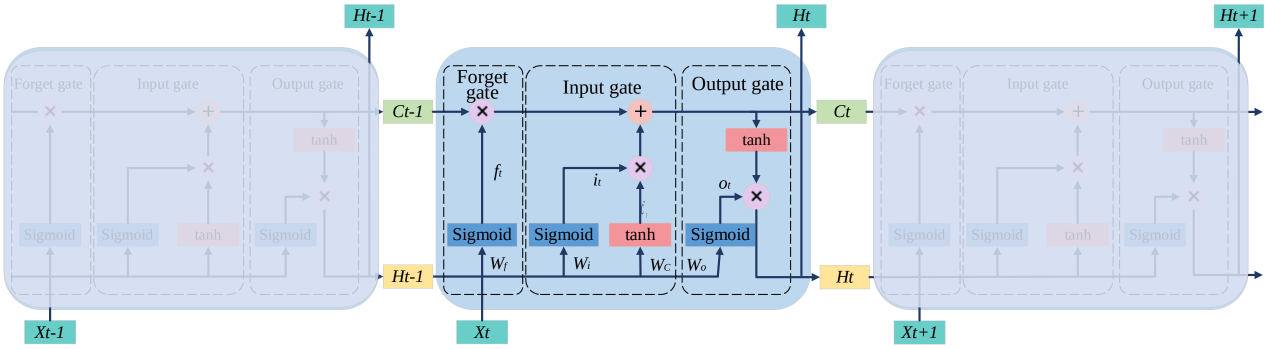


Figure 2. Internal mechanism of the LSTM model.

Where h_t represents the hidden state at time step t, capturing the contextual representation of the current moment influenced by the global information from time steps 0 to t. x_t denotes the input vector at time step t, while C_t represents the memory at time step t, which essentially encodes the global information from time steps 0 to t. Based on the input x_t and the hidden state information h_{t-1} from the previous time step, three gating mechanisms and a candidate memory are generated: The forget gate f_t, input gate i_t, output gate o_t, and candidate memory $\tilde{C}_t$.

The forget gate determines how much of the cell state from the previous time step should be retained in the current cell state, effectively deciding which information to discard from the cell state. The calculation formula is shown in Equation (9):

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \tag{9}$$

In this context, σ represents the sigmoid activation function, W_f denotes the weight matrix of the forget gate, and b_f is the bias term associated with the forget gate. This gate processes h_{t-1} and x_t, passing them through the sigmoid layer to produce an output f_t, which is a value between 0 and 1. The output f_t is then multiplied element-wise with each value in the cell state C_{t-1}. A value of 0 for f_t indicates complete forgetting, while a value of 1 signifies full retention.

The input gate determines the proportion of the current input to be retained in the cell state at the current time step. It consists of two components: The sigmoid layer, which decides which values to update, and the tanh layer, which generates the information to be added to the cell state. The tanh layer creates a new cell state vector, $\tilde{C}_t$, which is incorporated into the state, as shown in Equations (10) and (11):

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{10}$$

$$\tilde{C}_t = tanh(W_C \times [h_{t-1}, x_t] + b_C) \tag{11}$$

In these equations, W_i and W_C represent the weight matrices of the input gate and memory module, respectively, while b_i and b_C are their corresponding bias terms. The cell state C_{t-1} is updated to C_t by multiplying the old state with f_t, thereby discarding the designated information, and then adding i_t*C_t, where "*" represents element-wise multiplication. This process completes the cell state update, as described in Equation (12):

$$C_t = f_t{}^{*}C_{t-1} + i_t{}^{*}\tilde{C}_t \tag{12}$$

The output gate determines which portion of the cell state will be output through a sigmoid layer. The cell state is processed by a tanh function, yielding values within the range of −1 to 1, which are then multiplied with the output of the sigmoid gate. This operation ensures that only the designated portion of the state is output. The computations are described in Equations (13) and (14):

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{13}$$

$$h_t = o_t{}^{*}\, tanh(\,C_t) \tag{14}$$

where W_o represents the weight matrix of the output gate, and b_o denotes the bias term associated with the output gate.

2.3. Parameter optimization of Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) mimics the foraging behavior of sparrows. This optimization algorithm is based on a discoverer-follower model, augmented by a scouting and early-warning mechanism. In this model, the sparrow individual that first locates food acts as the discoverer, while the others are designated as followers. Additionally, a certain proportion of individuals are selected to act as scouts for early warning. If danger is detected, the food is abandoned [42]. The position matrix of sparrows is represented in Equation (15):

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix} \tag{15}$$

where *n* represents the number of sparrows, and *d* denotes the dimension of the variables in the optimization problem. The fitness values of all sparrows are computed using Equation (16):

$$F_x = \begin{bmatrix} f([x_{1,1} & x_{1,2} & \cdots & x_{1,d}]) \\ f([x_{2,1} & x_{2,2} & \cdots & x_{2,d}]) \\ \vdots & & & \\ \vdots & & & \\ f([x_{n,1} & x_{n,2} & \cdots & x_{n,d}]) \end{bmatrix} \tag{16}$$

During each iteration, the position update for discoverers can be described by Equation (17):

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t}\, exp[-i/a\mathrm{iter}_{\mathrm{max}}] & R_2 < S_T \\ X_{i,j}^{t} + QL & R_2 \geqslant S_T \end{cases} \tag{17}$$

where *t* is the current iteration number, *j* = 1, 2, …, *d*. $iter_{\mathrm{max}}$ is the maximum number of iterations (a constant); $X_{i,j}^{t}$ represents the position information of the *i*-th sparrow in the *j*-th dimension; $\alpha \in [0, 1]$ is a random number; $R_2 \in [0, 1]$ indicates the alertness value; $S_T \in [0.5, 1]$ represents the safety threshold; Q is a random number following a normal distribution; and L is a $1 \times d$ matrix with all elements equal to 1.

The position updates for joiners can be described by Equation (18):

$$X_{i,j}^{t+1} = \begin{cases} Q\, exp[\,(X_{\text{worst}}^{t} - X_{i,j}^{t})/i^2] & i > n/2 \\ X_{p}^{t+1} + |X_{i,j}^{t} - X_{p}^{t+1}|A^{+}L & \text{otherwise} \end{cases} \tag{18}$$

$$A^{+} = A^{T}(AA^{T})^{-1} \tag{19}$$

The term “exp” represents the current globally optimal position of the discoverer, while X_{worst}^{t} denotes the globally worst position. A is a $1 \times d$ matrix, with each element randomly assigned a value of either 1 or −1. When $i > n/2$, it indicates that the i-th joiner has a relatively low fitness value and has failed to obtain food, signifying a state of hunger and necessitating foraging in other locations.

A subset of sparrows capable of detecting danger, referred to as sentinels, is introduced. These sparrows are initially placed at random positions within the population. When encountering threats, sentinels quickly relocate to safe areas. The position update for a sentinel, $X_{i,j}^{t+1}$, is described by Equation (20):

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^{t} + \beta \mid X_{i,j}^{t} - X_{\text{best}}^{t} \mid & f_i > f_g \\ X_{i,j}^{t} + K\left(\dfrac{\mid X_{i,j}^{t} - X_{\text{worst}}^{t} \mid}{(f_i - f_w) + \varepsilon}\right) & f_i = f_g \end{cases} \tag{20}$$

where X_{best}^{t} represents the current globally optimal position, β is the step-size control parameter following a normal distribution with a mean of 0 and variance of 1, and $K \in [-1, 1]$ is a random number. Additionally, f_i denotes the fitness value of the iii-th sparrow, f_w is the globally worst fitness value, f_g is the globally best fitness value, and ε is a small constant to prevent division by zero.

Based on the above mathematical model, the flowchart of SSA optimizing LSTM parameters is shown in **Figure 3**. The fundamental operational steps can be summarized as follows:

1) Initialize the population, iteration count, and other parameters;
2) Rank the fitness values and identify the individuals with the best and worst fitness in the current iteration;
3) Update the positions of discoverers, followers, and randomly selected sparrows (those acting as sentinels);
4) Transmit parameters to the LSTM model and configure the LSTM neural network structure and training parameters based on the updated parameter set;
5) Calculate the fitness values based on the feedback from LSTM training results;
6) Define the best fitness value, update the position if the new position outperforms the previous one;
7) Repeat steps 3–6;
8) Output the optimal fitness value and the corresponding sparrow individual.

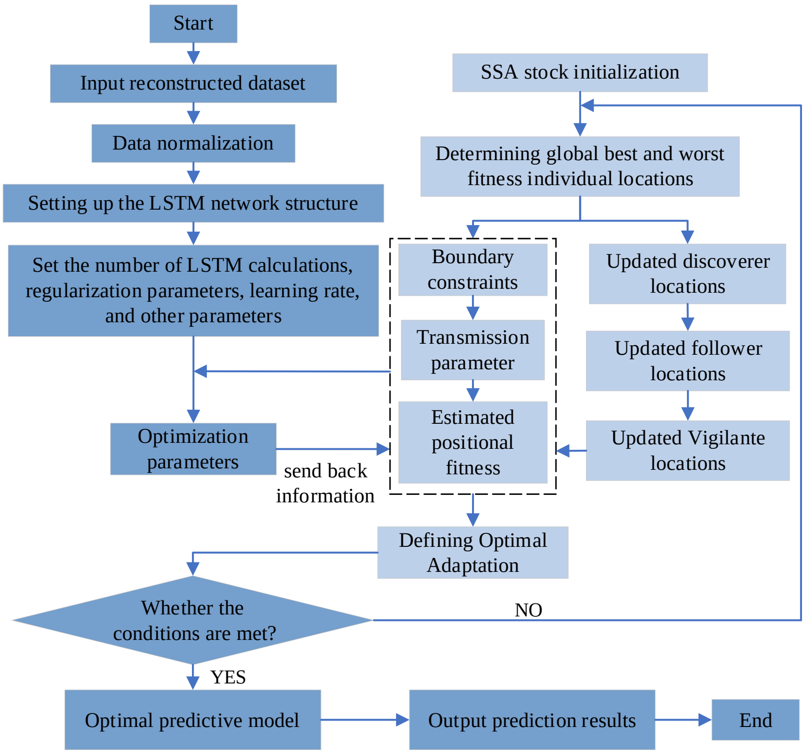


Figure 3. Forecasting workflow.

3. Evaluation indicators

To assess the forecasting accuracy of the models, four evaluation metrics were utilized: Root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and variance (VAR). RMSE and MAE reflect the deviation between predicted and actual values, with smaller values indicating better predictive performance. The R^2 value evaluates the goodness-of-fit, with values closer to 1 representing higher fitting accuracy. VAR assesses the stability of the forecasting model, where smaller values indicate greater model stability. The calculation formulas for these four metrics are shown in Equations (21)–(24):

$$X_{\mathrm{RMSE}} = \sqrt{\frac{1}{n_{\mathrm{test}}}\sum_{i=1}^{n_{\mathrm{test}}}\left(Y_{\mathrm{test}} - \hat{Y}_{\mathrm{test}}\right)^2} \tag{21}$$

$$X_{\mathrm{MAE}} = \left(\frac{1}{n_{\mathrm{test}}}\sum_{i=1}^{n_{\mathrm{test}}}\left|Y_{\mathrm{test}} - \hat{Y}_{\mathrm{test}}\right|\right) \tag{22}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{\mathrm{test}}}\left(Y_{\mathrm{test}} - \hat{Y}_{\mathrm{test}}\right)^2}{\sum_{i=1}^{n_{\mathrm{test}}}\left(Y_{\mathrm{test}} - \bar{Y}\right)^2} \tag{23}$$

$$X_{\mathrm{VAR}} = \frac{1}{n_{\mathrm{test}}}\sum_{i=1}^{n_{\mathrm{test}}}\left(Y_{\mathrm{test}} - \hat{Y}_{\mathrm{test}} - \mu\right) \tag{24}$$

In the equations, Y_{test} represents the actual measured values of the building heating load in the test set, $\hat{Y}_{\text{test}}$ denotes the predicted values of heating load by the model, $\bar{Y}$ is the sample mean calculated from the actual measured heating load values, μ refers to the sample mean of the prediction error sequence, and n_{test} indicates the number of samples in the test set.

4. Case simulation and analysis

4.1. Engineering overview

This study focuses on the energy system of an office building located in Shenyang, China. The building has an effective heating area of approximately 80,000 square meters during the winter season. The air-conditioning system operates daily from 7:00 to 19:00, with data recorded at 1-hour intervals. Environmental data and heating load data from the heating seasons (November to March of the following year) in 2021 and 2022 are used as the prediction dataset. The dataset is divided into a training set and a testing set in a 7:3 ratio.

4.2. Correlation analysis of influencing factors

As mentioned in the introduction, building heating loads are influenced by multiple coupled factors, such as the building envelope, outdoor temperature, outdoor relative humidity, solar radiation, room occupancy rate, wind speed, and wind direction. Among these, the building envelope is determined during the design phase and can therefore be treated as a fixed parameter for existing buildings. Although it may undergo slight variations due to environmental changes, these variations are challenging to capture and can be considered a form of noise interference.

The degree of influence of various factors on building heating load varies. When a factor has a low impact, incorporating it into the prediction model may reduce prediction accuracy and increase model complexity. Therefore, Pearson correlation coefficients (r) were initially employed to assess the relationships between influencing factors and heating load. The correlation results are shown in **Figure 4**.

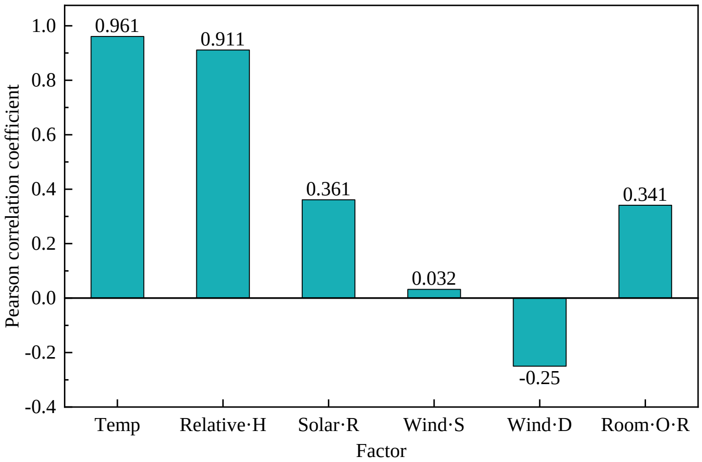


Figure 4. Pearson correlation coefficients of influencing factors.

To reduce subsequent computational complexity while ensuring prediction accuracy, it is necessary to select influencing factors with a certain degree of correlation. Additionally, the original factors associated with strongly correlated modes after modal decomposition should be retained as much as possible. According to the Pearson correlation coefficient strength table [43], it is shown in **Table 1** that two weakly correlated factors—wind direction and wind speed ($0.1 \leq |r| < 0.3$)—were excluded to prevent feature redundancy from adversely affecting the prediction results.

Table 1. Strength levels of pearson correlation coefficients.

r Value	Correlation Strength
$0.1 \leq \lvert r\rvert < 0.3$	Weak correlation
$0.3 \leq \lvert r\rvert < 0.5$	Moderate correlation
$0.5 \leq \lvert r\rvert < 0.7$	Fairly strong correlation
$0.7 \leq \lvert r\rvert < 0.9$	Strong correlation
$0.9 \leq \lvert r\rvert \leq 1$	Very strong correlation

The correlation coefficient analysis reveals that, although wind direction and wind speed theoretically influence building loads to some extent, their effects are largely diminished when the building has good airtightness and insulation. This is particularly evident under conditions where low temperatures coincide with low wind speeds and high temperatures with high wind speeds, resulting in the impact of temperature and other coupled factors on building loads overshadowing the influence of wind direction and wind speed. Consequently, four factors with higher correlations were selected based on the prescribed criteria: Outdoor temperature, outdoor relative humidity, solar radiation, and room occupancy rate.

4.3. MVMD parameter configuration and results

To further extract complementary features for prediction, MVMD decomposition was performed on the four selected influencing factors and heating load data identified in the previous section.

4.3.1. MVMD parameter configuration

After decomposition, the original data exhibits greater regularity. The parameter settings for MVMD are presented in **Table 2**.

Table 2. Main parameter settings of MVMD.

Penalty facto (alpha)	Noise tolerance (tau)	Initialization mode (init)	Convergence value (tol)	Number of modes (Z)
2500	0	1	1×10^{-7}	2–10

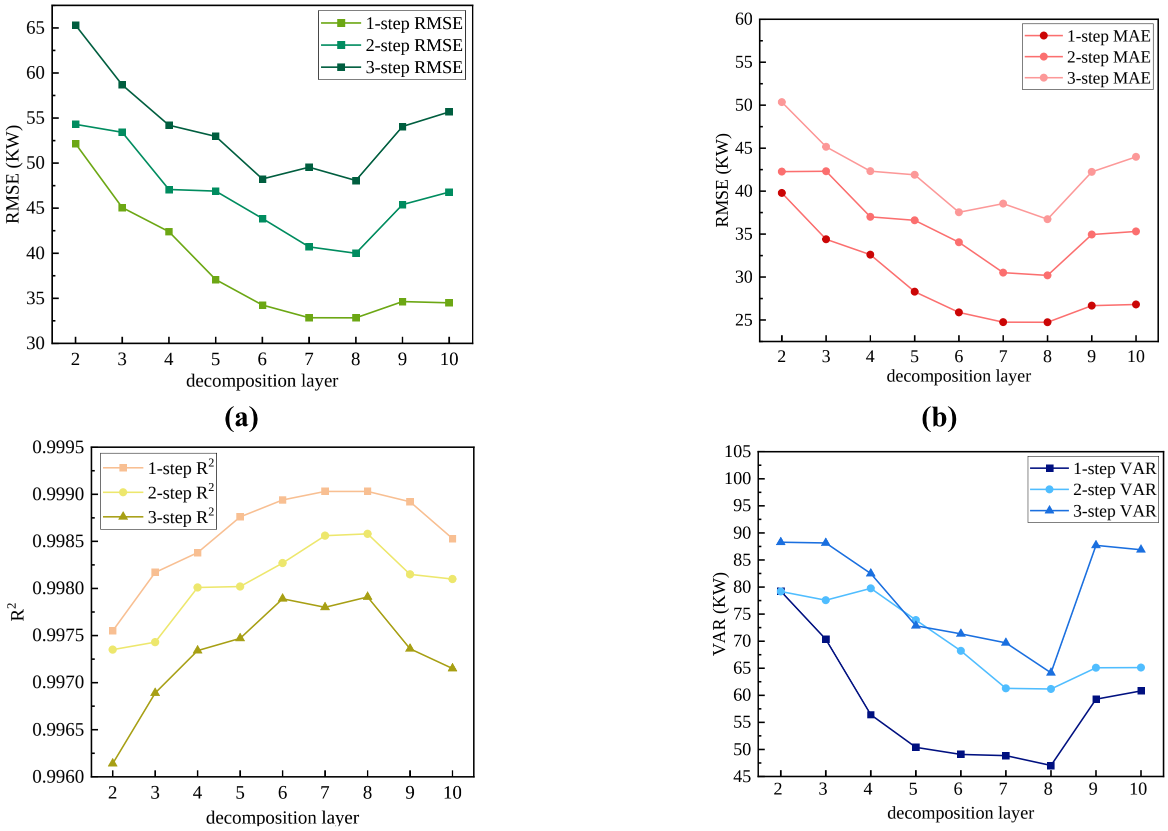
In **Table 2**, the penalty factor controls the smoothness of the modal decomposition, with higher values leading to smoother modes. The noise tolerance allows for a certain level of noise, where a value of 0 indicates no strict fidelity requirement. For initialization mode selection, a value of 1 represents uniform initialization. The convergence value determines the stopping criterion for the iteration process, halting when the change falls below this value. The number of modes specifies the number of sub-signals into which the signal is decomposed.

4.3.2. Determination of MVMD decomposition layers

MVMD effectively captures signal characteristics; however, an inappropriate number of decomposed modes, whether too many or too few, can adversely affect prediction outcomes. When the number of modes is too small, there is a risk of information loss, and critical frequency components and signal features may not be adequately captured. Conversely, an excessive number of modes can capture noise and random fluctuations within the data, reducing the model's generalization ability and leading to poor performance on new datasets. Additionally, higher mode counts increase computational complexity, potentially diminishing prediction accuracy.

To determine the optimal number of decomposition modes for the MVMD model, the mode count was empirically set between 2 and 10. Simulations were conducted sequentially with different step sizes to evaluate performance.

Figure 5 illustrates the errors of the MVMD-LSTM prediction model under varying numbers of decomposition modes for step sizes of 1, 2, and 3. As shown in the figure, when the number of decomposition modes is set to 8, all four error metrics reach their minimum values, indicating the highest prediction accuracy, best performance, and optimal model stability. Additionally, all four error metrics increase as the step size grows. Therefore, in the subsequent model analysis, the number of decomposition modes for MVMD is uniformly set to 8.



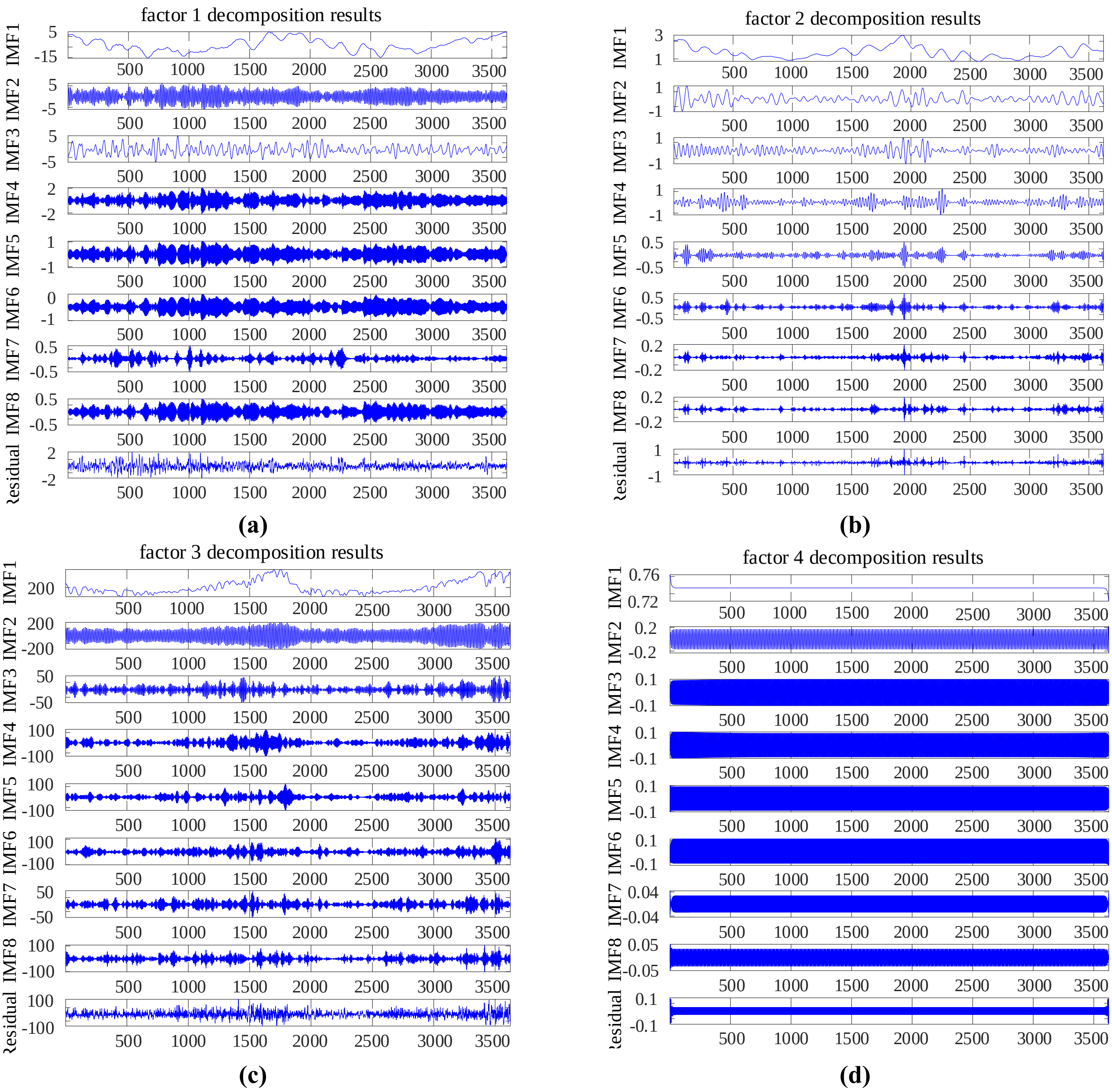
(a) **(b)**

(c) **(d)**

Figure 5. Forecasting errors under different step sizes and decomposition levels, **(a)** RMSE error comparison; **(b)** MAE error comparison; **(c)** R^2 error comparison; **(d)** VAR error comparison.

4.3.3. MVMD decomposition results

The decomposition results based on the mode number determined in the previous section are presented in **Figure 6**.


factor 1 decomposition results

factor 2 decomposition results

factor 3 decomposition results

factor 4 decomposition results

(a) **(b)** **(c)** **(d)**

Figure 6. Modal decomposition results of heat load influencing factors, **(a)** modal decomposition of temperature; **(b)** modal decomposition of relative humidity; **(c)** modal decomposition of solar radiation; **(d)** modal decomposition of room occupancy rate.

The Pearson correlation coefficients between each mode obtained from the decomposition of the four influencing factors and the original heating load data are presented in **Table 3**.

Table 3. Pearson correlation coefficients between each mode and heating load.

Mode	Temperature	Relative humidity	Solar radiation	Room occupancy rate
1	0.8316	0.6630	0.4168	0.01224
2	0.3437	0.6300	0.7913	0.4069
3	0.4919	0.5122	0.2087	0.7597
4	0.1136	0.3428	0.2164	0.7596
5	0.0751	0.2348	0.1863	0.2805
6	0.0581	0.1736	0.1750	0.3079
7	0.0396	0.0670	0.1742	0.2490
8	0.0504	0.0554	0.188	0.1412

To select modes with stronger regularity as supplementary features to the original influencing factors for heating load prediction, the mode with the highest correlation coefficient for each influencing factor is identified and combined with the original factors as input data for prediction. Based on the results in **Table 3**, the following modes are selected: The first sub-mode for temperature, the first sub-mode for relative humidity, the second sub-mode for solar radiation, and the third sub-mode for room occupancy rate.

Each mode of the decomposed load data is used as a separate output for prediction. The decomposition results are illustrated in **Figure 7**, where each of the k modes is sequentially treated as an output. Finally, the predicted outputs are aggregated to reconstruct the original form of the data, enabling subsequent error analysis.

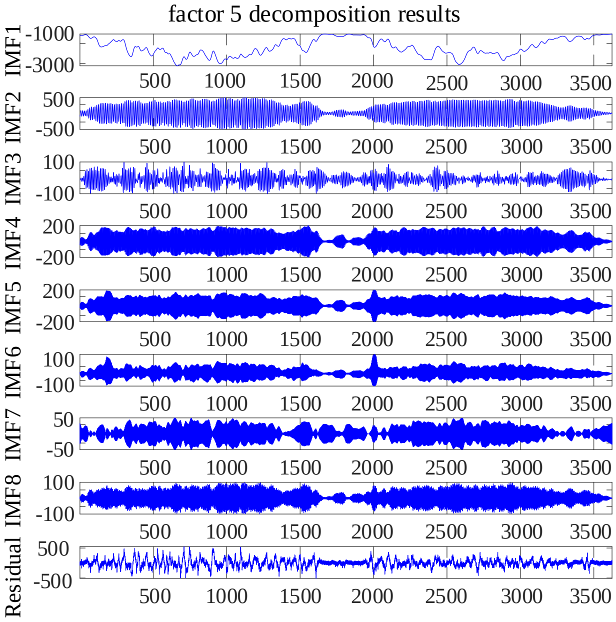


Figure 7. Heating load modal decomposition results.

4.4. LSTM network architecture and parameter configuration

The structure of the Long Short-Term Memory (LSTM) neural network comprises an input layer (which accepts time-series data as input), an LSTM layer with 60 hidden units (to process sequential input data and capture long- and short-term dependencies), a ReLU activation layer (to introduce non-linearity and enhance the network's representational power), a fully connected layer, and a regression layer (to compute the loss and guide network training and optimization).

The training process utilizes the Adam optimization algorithm with a maximum of 200 training epochs, a batch size of 15 for each iteration, and an initial learning rate of 0.02. The learning rate is adaptively adjusted during training to enhance performance. These configurations enable the LSTM network to effectively capture the intricate characteristics of time-series data and produce accurate predictions. Additional parameters are configured as shown in **Table 4**.

Table 4. Main parameter settings of the LSTM neural network.

Maximum Training Epochs	**Batch Size**	**Gradient Threshold**	**Initial Learning Rate**	**Learning Rate Decay Period**	**Learning Rate Decay Factor**	**Regularization Parameter**	**Training Environment**
200	15	0.15	0.02	60	0.25	1×10^{-6}	CPU

4.5. SSA parameter configuration and optimization results

The Sparrow Search Algorithm (SSA) was employed to optimize the LSTM parameters, significantly enhancing the model's predictive performance and generalization ability. SSA performs a comprehensive search across the parameter space to identify the global optimum, effectively mitigating the uncertainties associated with empirical parameter selection. The primary parameter settings are presented in **Table 5**.

Table 5. Main parameter settings of SSA.

Population Size	**Maximum Iterations**	**Optimization Parameter Dimensions**	**Scouting Threshold**	**Proportion of Discoverers**	**Proportion of Sparrows Aware of Danger**
30	50	3	0.6	0.7	0.2

The optimization targets include the maximum training epochs, the number of hidden units, and the initial learning rate. Each of these parameters plays a critical role. Maximum training epochs determine the adequacy of model training. The number of hidden units controls the network's memory capacity and complexity. The initial learning rate affects both convergence speed and optimization quality. These parameters require careful tuning to balance computational efficiency and prediction accuracy. Excessive values lead to increased computation time and resource waste, whereas insufficient values may result in suboptimal predictive performance. Based on computational experience, the following constraints were established for parameter optimization: Lower bounds lb = [50, 100, 0.001] and upper bounds ub = [200, 250, 0.01]. The SSA iteratively searches within these defined dimensions and ranges to determine the optimal LSTM hyperparameter configuration. Results indicate that the optimal combination [180, 216, 0.0085] yields the lowest prediction error.

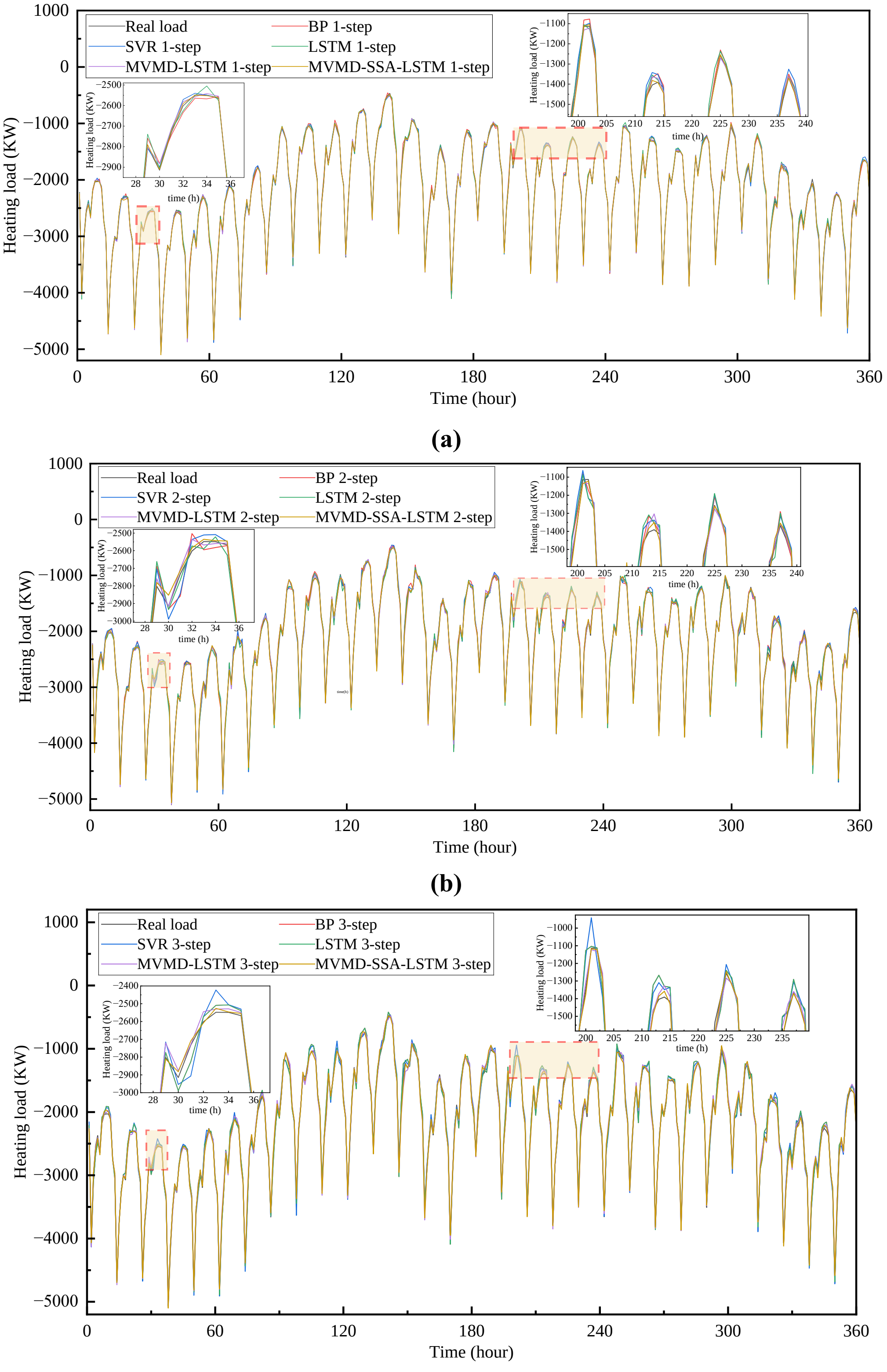
5. Comparison and analysis of forecasting results

To evaluate the accuracy and stability of the MVMD-SSA-LSTM model, this section conducts an error comparison and analysis of commonly used baseline models, including BP, SVR, LSTM, and the optimized MVMD-LSTM model. All models were tested for 1-step, 2-step, and 3-step predictions, corresponding to forecasts made 1 h, 2 h, and 3 h in advance, respectively. **Table 6** summarizes the error comparisons across different models and prediction horizons. The forecasting result curves for the primary models are illustrated in **Figure 8**, and the residual plots are shown in **Figure 9**, with the figures presenting the forecasting results for a single month.

Table 6. Multi-step forecasting errors of different models.

Step	Model	RMSE/kW	MAE/kW	R^2	VAR/kW
1	BP	62.31	43.43	0.99683	87.90
	SVR	58.65	43.36	0.99690	92.97
	LSTM	52.35	36.35	0.99753	60.46
	VMD-LSTM	33.65	26.43	0.99898	48.96
	VMD-SSA-LSTM	29.94	23.87	0.99916	47.71
	MVMD-LSTM	32.84	24.74	0.99903	47.04
	MVMD-SSA-LSTM	28.57	21.66	0.99927	46.89
2	BP	84.65	60.57	0.99555	145.34
	SVR	85.67	59.82	0.99340	136.421
	LSTM	69.94	51.05	0.99560	102.26
	VMD-LSTM	42.87	34.17	0.99807	70.13
	VMD-SSA-LSTM	40.40	31.89	0.99833	68.92
	MVMD-LSTM	39.99	30.20	0.99858	64.15
	MVMD-SSA-LSTM	33.38	25.25	0.99900	47.65
3	BP	90.90	68.22	0.99448	169.74
	SVR	101.98	70.55	0.99061	182.55
	LSTM	87.99	67.29	0.99499	142.95
	VMD-LSTM	52.73	42.38	0.99603	83.62
	VMD-SSA-LSTM	49.86	38.91	0.99790	70.64
	MVMD-LSTM	48.06	36.73	0.99791	61.16
	MVMD-SSA-LSTM	41.34	32.36	0.99843	54.90

To minimize prediction errors and enhance model accuracy, the LSTM network, which exhibited the best performance among BP, SVR, and LSTM models, was selected as the baseline model. As shown in **Table 6**, MVMD decomposition significantly improves the performance of the LSTM prediction model. Compared to the single baseline LSTM model, the MVMD-LSTM model achieves substantial reductions in prediction errors for 1-step, 2-step, and 3-step forecasts. Specifically, RMSE decreases by 37.27%, 42.82%, and 45.38%, respectively, while MAE decreases by 31.94%, 40.84%, and 45.41%. Additionally, R^2 improves by 0.0015, 0.0030, and 0.0029, and VAR decreases by 22.20%, 37.27%, and 57.22%. These results indicate a significant reduction in errors, demonstrating that optimizing with the MVMD algorithm greatly enhances prediction accuracy and model stability, with more pronounced benefits as the forecasting horizon increases.

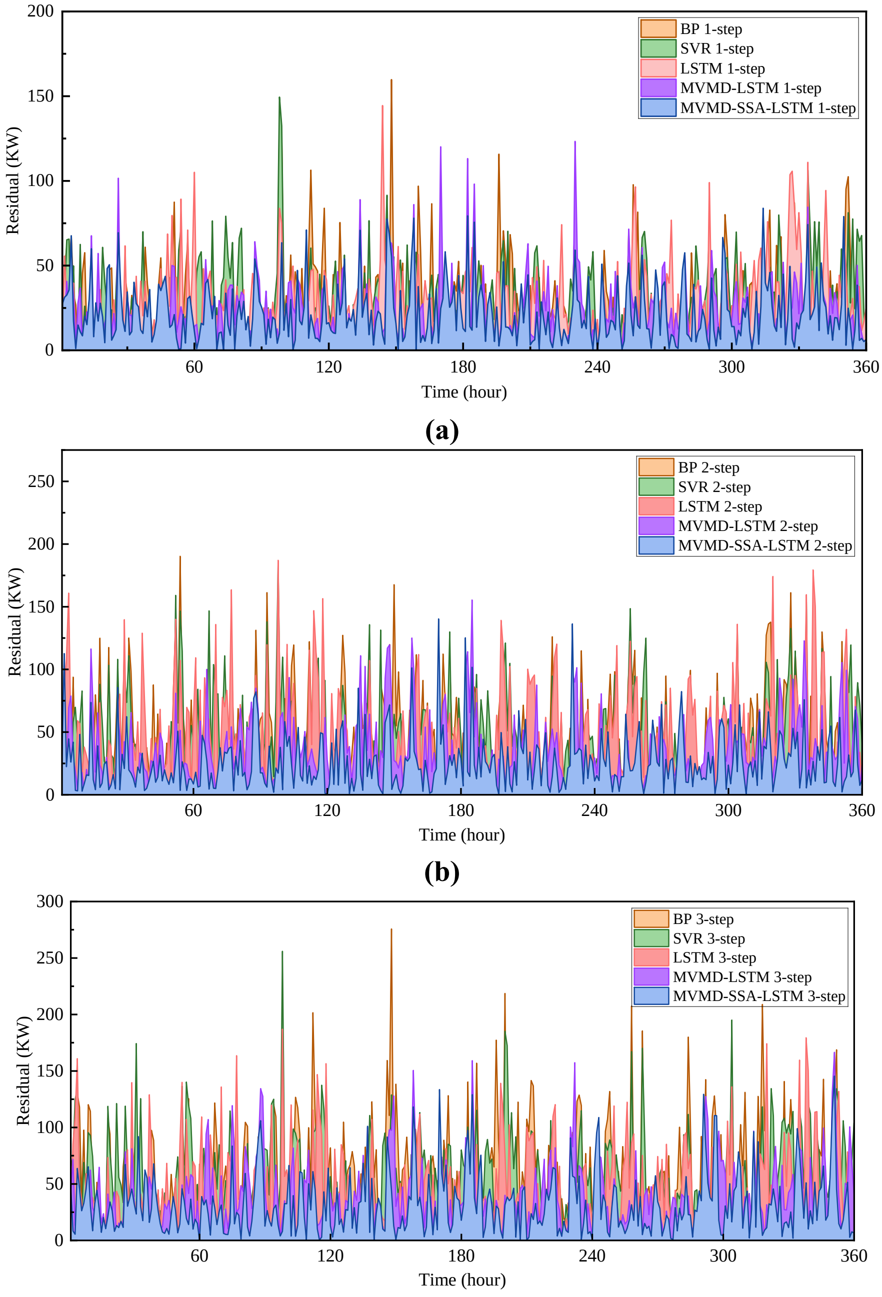


(a)

(b)

(c)

Figure 8. The primary model's multi-step forecasting results, **(a)** 1-step forecasting results; **(b)** 2-step forecasting results; **(c)** 3-step forecasting results.



(a)

(b)

(c)

Figure 9. The primary model's multi-step forecasting residuals, **(a)** 1-step forecasting residuals; **(b)** 2-step forecasting residuals; **(c)** 3-step forecasting residuals.

Building on the MVMD-LSTM model, the MVMD-SSA-LSTM introduces further optimization. Comparisons of their prediction results for 1-step, 2-step, and 3-step forecasts reveal that RMSE decreases by 13.00%, 16.53%, and 13.98%, respectively, while MAE decreases by 12.45%, 16.39%, and 11.90%. R^2 improves by 0.00024, 0.00042, and 0.00052, and VAR decreases by 0.32%, 25.72%, and 10.24%. Residual plots further illustrate that the MVMD-SSA-LSTM model consistently achieves the smallest residual values across all prediction horizons. These findings confirm that optimizing key LSTM parameters using the Sparrow Search Algorithm (SSA) further enhances the model's accuracy and stability. Moreover, the results indicate that while the VDM-SSA-LSTM model performs slightly worse than the MVMD-SSA-LSTM model for 1-step predictions, the performance gap widens as the prediction horizon increases.

6. Conclusion

This study proposes a short-term building heating load forecasting framework based on Multivariate Variational Mode Decomposition (MVMD) and the Sparrow Search Algorithm-Long Short-Term Memory (SSA-LSTM) model. In the MVMD-SSA-LSTM architecture, MVMD is first applied to decompose the original time series data of temperature, relative humidity, solar radiation, room occupancy rate, and building heating load. Supplementary features derived from the decomposition, combined with the original features, serve as the input for forecasting. Each mode of the building heating load is predicted individually, and the final forecasting is obtained by aggregating the predicted results.

Experimental results demonstrate that the decomposition of raw data using MVMD, coupled with feature extraction, significantly enhances forecasting accuracy and model stability. This indicates that MVMD effectively identifies instabilities in time series data, reduces sequence complexity, and performs better than VMD. Compared with univariate decomposition methods, MVMD overcomes the frequency mismatch issue that arises from decomposing multi-factor input sequences. Furthermore, by conducting Pearson correlation coefficient analysis, MVMD effectively mitigates noise interference in building load data, reduces feature redundancy, and yields better predictive performance.

Additionally, the integration of SSA to optimize the parameter settings of the LSTM model reduces the biases associated with empirically determined parameters, further improving the precision and stability of the predictive model. Compared to other baseline forecasting models, the MVMD-SSA-LSTM model exhibits superior performance in both 1-step and multi-step forecasting, demonstrating significant application potential.

Author contributions: Conceptualization, YZ and BZ; methodology, YZ; software, YZ; validation, XL, YZ and BZ; formal analysis, XL and CF; investigation, YZ, CF and ZX; resources, BZ; data curation, DL; writing—original draft preparation, YZ; writing—review and editing, XL and BZ; visualization, YZ; supervision, CF; project administration, BZ; funding acquisition, BZ. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Ekonomou L, Christodoulou CA, Mladenov V. A short-term load forecasting method using artificial neural networks and wavelet analysis. Int. J. Power Syst. 2016; 1: 64–68.
2. Verbeke S, Audenaert A. Thermal inertia in buildings: A review of impacts across climate and building use. Renewable and Sustainable Energy Reviews. 2018; 82: 2300–2318.

3. Rugani R, Bigazzi M, Fantozzi F, et al. Impact of building thermal inertia in different climates using energy dynamic simulation through a simplified description model. In: Proceedings of the 17th IBPSA Conference; 1–3 September 2021; Bruges, Belgium. Volume 17, pp. 2578–2585.
4. Wu Q, Wang J, Meng X. Influence of wall thermal performance on the contribution efficiency of the Phase-Change Material (PCM) layer. Case Studies in Thermal Engineering. 2021; 28: 101398.
5. Chapaloglou S, Varagnolo D, Marra F, et al. Data-driven energy management of isolated power systems under rapidly varying operating conditions. Applied Energy. 2022; 314: 118906.
6. Luo XJ, Oyedele LO. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. Advanced Engineering Informatics. 2021; 50: 101357.
7. Hu J, Zou Y, Soltanov N. A multilevel optimization approach for daily scheduling of combined heat and power units with integrated electrical and thermal storage. Expert Systems with Applications. 2024; 250: 123729.
8. Wei S, Tien PW, Calautit JK, et al. Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method. Applied Energy. 2020; 277: 115506.
9. Sze V, Chen YH, Yang TJ, et al. Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE. 2017; 105(12): 2295–2329.
10. Pan L, Wang S, Wang J, et al. Research on central air conditioning systems and an intelligent prediction model of building energy load. Energies. 2022; 15(24): 9295.
11. Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews. 2017; 73: 1104–1122.
12. Wang Z, Hong T, Piette MA. Building thermal load prediction through shallow machine learning and deep learning. Applied Energy. 2020; 263: 114683.
13. Zhou C, Fang Z, Xu X, et al. Using long short-term memory networks to predict energy consumption of air-conditioning systems. Sustainable Cities and Society. 2020; 55: 102000.
14. Xu Y, Gao W, Qian F, et al. Potential analysis of the attention-based LSTM model in ultra-short-term forecasting of building HVAC energy consumption. Frontiers in Energy Research. 2021; 9: 730640.
15. Chalapathy R, Khoa NLD, Sethuvenkatraman S. Comparing multi-step ahead building cooling load prediction using shallow machine learning and deep learning models. Sustainable Energy, Grids and Networks. 2021; 28: 100543.
16. Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation. 2019; 31(7): 1235–1270.
17. Zhang L, Wen J, Li Y, et al. A review of machine learning in building load prediction. Applied Energy. 2021; 285: 116452.
18. Das A, Annaqeeb MK, Azar E, et al. Occupant-centric miscellaneous electric loads prediction in buildings using state-of-the-art deep learning methods. Applied Energy. 2020; 269: 115135.
19. Xue J, Shen B. A novel swarm intelligence optimization approach: Sparrow search algorithm. Systems Science & Control Engineering. 2020; 8(1): 22–34.
20. Saafan MM, El-Gendy EM. IWOSSA: An improved whale optimization salp swarm algorithm for solving optimization problems. Expert Systems with Applications. 2021; 176: 114901.
21. An W, Gao B, Liu J, et al. Predicting hourly heating load in residential buildings using a hybrid SSA–CNN–SVM approach. Case Studies in Thermal Engineering. 2024; 59: 104516.
22. Madiniyeti J, Chao Y, Li T, et al. Concrete dam deformation prediction model research based on SSA–LSTM. Applied Sciences. 2023; 13(13): 7375.
23. Huang Z, Huang J, Min J. SSA-LSTM: Short-Term photovoltaic power prediction based on feature matching. Energies. 2022; 15(20): 7806.
24. Gao X, Guo W, Mei C, et al. Short-term wind power forecasting based on SSA-VMD-LSTM. Energy Reports. 2023; 9: 335–344.
25. Han M, Zhong J, Sang P, et al. A combined model incorporating improved SSA and LSTM algorithms for short-term load forecasting. Electronics. 2022; 11(12): 1835.
26. Chen H, Zhang H, Wu S, et al. Numerical simulation and optimisation design for ventilation and heat dissipation in high-temperature and high-load indoor substations. Case Studies in Thermal Engineering. 2024; 59: 104502.
27. Cha YJ, Mostafavi A, Benipal SS. DNoiseNet: Deep learning-based feedback active noise control in various noisy environments. Engineering Applications of Artificial Intelligence. 2023; 121: 105971.

28. Genzel M, Macdonald J, März M. Solving inverse problems with deep neural networks–robustness included? IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022; 45(1): 1119–1134.
29. Leprince J, Madsen H, Miller C, et al. Fifty shades of grey: Automated stochastic model identification of building heat dynamics. Energy and Buildings. 2022; 266: 112095.
30. Bampoulas A, Pallonetto F, Mangina E, et al. A Bayesian deep-learning framework for assessing the energy flexibility of residential buildings with multicomponent energy systems. Applied Energy. 2023; 348: 121576.
31. Yeh JR, Shieh JS, Huang NE. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. Advances in Adaptive Data Analysis. 2010; 2(02): 135–156.
32. Wang H, Huang S, Li P, et al. An improved wavelet transform threshold method for denoising fault signals. Measurement Science and Technology. 2016; 27(3): 1–10.
33. Liu H, Wang W, Xiang C, et al. A de-noising method using the improved wavelet threshold function based on noise variance estimation. Mechanical Systems and Signal Processing. 2018; 99: 30–46.
34. Wardana ANI. A comparative study of EMD, EWT and VMD for detecting the oscillation in control loop. In: Proceedings of the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic); 05–06 August 2016; Semarang, Indonesia.
35. Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing. 2014; 62(3): 531–544.
36. Fu J, Cai F, Guo Y, et al. An Improved VMD-Based Denoising Method for Time Domain Load Signal Combining Wavelet with Singular Spectrum Analysis. Mathematical Problems in Engineering. 2020; 2020(1): 1485937.
37. Bagheri A, Ozbulut OE, Harris DK. Structural system identification based on variational mode decomposition. Journal of Sound and Vibration. 2018; 417: 182–197.
38. ur Rehman N, Aftab H. Multivariate variational mode decomposition. IEEE Transactions on signal processing. 2019; 67(23): 6039–6052.
39. Zhang L, Li J, Xu X, et al. High spatial granularity residential heating load forecast based on Dendrite net model. Energy. 2023; 269: 126787.
40. Salami BA, Abba SI, Adewumi AA, et al. Building energy loads prediction using bayesian-based metaheuristic optimized-explainable tree-based model. Case Studies in Construction Materials. 2023; 19: e02676.
41. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena. 2020; 404: 132306.
42. Awadallah M A, Al-Betar M A, Doush I A, et al. Recent versions and applications of sparrow search algorith. Archives of Computational Methods in Engineering, 2023, 30(5): 2831-2858.
43. Asuero AG, Sayago A, González AG. The correlation coefficient: An overview. Critical Reviews in Analytical Chemistry. 2006; 36(1): 41–59.