energies

MDPI

Article

Multi-Level Decomposition and Interpretability-Enhanced Air Conditioning Load Forecasting Study

Xinting Yang [1], Ling Zhang [2], Hong Zhao [2], Wenhua Zhang [2], Chuan Long [1], Gang Wu [1], Junhao Zhao [3] and Xiaodong Shen [3,*]

[1] State Grid Sichuan Economic Research Institute, Chengdu 610041, China; vividlove6@163.com (X.Y.); longchuan123@126.com (C.L.); scu_gw@163.com (G.W.)
[2] State Grid Sichuan Electric Power Company, Chengdu 610041, China; zhangl003x@sc.sgcc.com.cn (L.Z.); zhaoh0025a@sc.sgcc.com.cn (H.Z.); zwh3702@163.com (W.Z.)
[3] College of Electrical Engineering, Sichuan University, Chengdu 610065, China; zjh02080616@163.com
* Correspondence: shengxd@scu.edu.cn

Abstract: This study seeks to improve the accuracy of air conditioning load forecasting to address the challenges of load management in power systems during high-temperature periods in the summer. Given the limitations of traditional forecasting models in capturing different frequency components and noise within complex load sequences, this paper proposes a multi-level decomposition forecasting model using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), sample entropy (SE), variational mode decomposition (VMD), and long short-term memory (LSTM). First, CEEMDAN is used for the preliminary decomposition of the raw air-conditioning load series, with modal components aggregated by sample entropy to generate high-, medium-, and low-frequency subsequences. VMD then performs a secondary decomposition on the high-frequency subsequence to reduce its complexity, while LSTM is applied to each subsequence for prediction. The final prediction result of the air-conditioning load is obtained through reconstruction. To validate model performance, this paper uses air-conditioning load data from Nanchong City and Sichuan Province, for experimental analysis. Results show that the proposed method significantly outperforms the LSTM model without decomposition and other benchmark models in prediction accuracy, with the Root Mean Square Error (RMSE) reductions ranging from 40.26% to 74.18% and the Modified Mean Absolute Percentage Error (MMAPE) reductions from 37.75% to 73.41%. By employing the SHAP (Shapley additive explanations) method for both global and local interpretability, the model reveals the influence of key factors, such as historical load and temperature, on load forecasting. The decomposition and aggregation approach introduced in this paper substantially enhances forecasting accuracy, providing a scientific foundation for power system load management and dispatch.

Keywords: air conditioning; load forecasting; fully adaptive noise empirical mode decomposition; variational mode decomposition; long short-term memory network; sample entropy; SHAP interpretability

Citation: Yang, X.; Zhang, L.; Zhao, H.; Zhang, W.; Long, C.; Wu, G.; Zhao, J.; Shen, X. Multi-Level Decomposition and Interpretability-Enhanced Air Conditioning Load Forecasting Study. *Energies* **2024**, *17*, 5881. https://doi.org/10.3390/en17235881

Academic Editors: Antonio Gabaldón, María Carmen Ruiz-Abellón and Luis Alfredo Fernández-Jiménez

Received: 4 November 2024
Revised: 19 November 2024
Accepted: 20 November 2024
Published: 23 November 2024

1. Introduction

With the intensification of global warming and the increasing frequency of extreme weather events, the usage frequency and intensity of residential air conditioning has significantly risen [1]. This change directly impacts the load curve of the power system, especially during high temperatures in the summer, where air conditioning loads often constitute a substantial portion of the total electricity load, resulting in a sharp increase in the load over a short period. This poses a significant challenge to the safe and stable operation of the power grid [2]. Therefore, accurately forecasting residential air conditioning loads is crucial, as it helps power authorities to develop proactive measures to ensure the safe and stable operation of the grid during peak electricity usage.

Air conditioning load forecasting methods can be primarily divided into physical model-driven and data-driven approaches. The physical model-driven methods simulate

and predict air conditioning loads by establishing equipment operation models and user behavior models, effectively capturing the dynamic characteristics of the system and the interactions between devices. However, these methods often rely on complex data and a large number of physical parameters, requiring specialized software tools (e.g., Energy Plus [3], TRNSYS [4]) for in-depth analysis. The main limitation lies in the high complexity of model construction and the long simulation time required, while the high uncertainty of user behavior and environmental conditions makes the accurate construction of physical models particularly challenging.

Unlike physical model-driven methods, data-driven approaches predict loads by analyzing historical load data and environmental variables. These methods offer advantages such as low cost, model simplicity, and high computational efficiency, without the need for in-depth knowledge of the physical domain. In terms of statistical methods, improved exponential smoothing [5], autoregressive models combined with weather forecasts [6], and multivariate linear regression [7] have all been applied to air conditioning load forecasting for various types of buildings, such as residential and commercial spaces. However, these methods face limitations in capturing complex nonlinear relationships and long-term dependencies, which are crucial in scenarios with highly variable loads and diverse environmental influences [8].

To address these limitations, machine learning techniques have become popular in data-driven air conditioning load forecasting models. For example, Gao et al. [9] proposed a hybrid method for cooling load forecasting in large commercial buildings, integrating an Extreme Learning Machine (ELM) with Random Forest (RF) and an Improved Parallel Whale Optimization Algorithm (IPWOA). This model addresses the demand fluctuations in commercial spaces with high accuracy, achieving Mean Absolute Percentage Error (MAPE) reductions of up to 95.79% and making it suitable for real-time applications in energy management systems. Zhou et al. [10] developed a hybrid ISSA-LSTM model combining an Improved Sparrow Search Algorithm (ISSA) with LSTM to predict building air conditioning heat load. Tested on a university laboratory dataset, the model achieved a high value R^2 of 0.9971 and a significant reduction in RMSE, showing potential for reducing operational costs and energy usage in building energy management. Wang et al. [11] applied a Wavelet Neural Network (WNN) to forecast short-term cooling, heating, and electrical loads in typical buildings, including hotels, hospitals, shopping malls, offices, and residential complexes. With a maximum MAPE of 1.8%, this method demonstrated high accuracy and proved effective for rapid, reliable load forecasting across various building types. Wang et al. [12] proposed a dynamic forecasting model for building cooling loads, combining an Artificial Neural Network (ANN) with an ensemble approach that integrated data-driven methods and physical laws to capture the complex nonlinear relationships within HVAC system cooling loads. This hybrid approach significantly improved accuracy, achieving an R^2 of 0.96, providing a more precise load forecasting solution for HVAC systems compared to traditional models.

Recent studies have shown that, beyond improving model architectures, data decomposition techniques such as Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and Variational Mode Decomposition (VMD) can further enhance prediction accuracy by extracting underlying patterns and trends in the data [13,14]. For instance, Huang et al. [15] developed an EMD-LSTM-Markov model for cooling load forecasting in commercial office buildings. This hybrid approach addressed nonlinear, seasonal load variations and achieved significant accuracy improvements, with RMSE reductions of 40–94% and MAPE reductions of 70–96%, providing an effective solution for real-time HVAC load forecasting. Karijadi and Chou [16] proposed a hybrid RF-LSTM model with CEEMDAN for improved building energy consumption forecasting. Applied to various building types, including dormitories and offices, this model reduced MAPE by over 40% compared to the standard models, effectively handling non-stationary, nonlinear data for enhanced energy management. Li et al. [17] proposed a hybrid short-term load forecasting model combining the VMD and LSTM

networks, optimized using Bayesian Optimization. Designed for power load prediction in regions like Hubei Province, the model accounts for complex, nonlinear factors such as temperature and the time of day. Compared to LSTM alone, this approach improved RMSE prediction performance across four seasons by 63.8% to 79.7%.

Although existing methods have improved the accuracy of air conditioning load forecasting to some extent through better model structures or decomposition techniques, the following issues remain: (1) the high-frequency fluctuation characteristics of air conditioning loads have not been effectively addressed, as single decomposition methods cannot fully utilize the hidden information in the data; (2) decomposition methods break down the raw data into multiple Intrinsic Mode Functions (IMFs), requiring each component to be predicted separately, which increases computational burden; (3) existing forecasting models, while improving accuracy, often lack sufficient interpretability, which is crucial for ensuring operational transparency and decision credibility in power grid operations.

To address these issues, this paper proposes an explainable forecasting framework based on CEEMDAN-SE-VMD-LSTM with SHAP (Shapley additive explanations). This framework first employed CEEMDAN to decompose air conditioning load data, evaluating the complexity of modal components using sample entropy to reorganize them into high-, mid-, and low-frequency subsequences. The high-frequency subsequence then underwent secondary decomposition with VMD to simplify its complexity and improve its predictability. Each subsequence was independently forecasted using LSTM, and the final prediction results were integrated. SHAP is used to explain the prediction outcomes, revealing key influencing factors and enhancing the forecast credibility.

The remainder of this paper is organized as follows: Section 2 provides a detailed description of the methods used in this study. Section 3 presents the case study analysis, including forecasting and interpretability assessments. Finally, Section 4 summarizes the main findings and introduces directions for future research.

2. Materials and Methods

This section introduces the various methods used in the proposed prediction framework, including CEEMDAN decomposition, sample entropy, VMD, LSTM neural networks, and SHAP explainability methods. The overall prediction framework is illustrated in Figure 1.

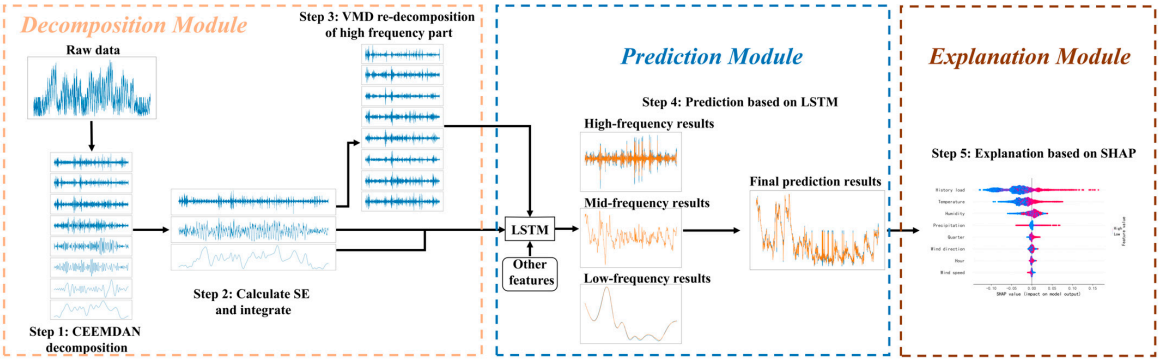

Figure 1. Overall prediction framework.

2.1. CEEMDAN

CEEMDAN is an improvement based on EMD, adaptively decomposing nonlinear and non-stationary time series into multiple IMFs of different frequencies [18]. By incorporating adaptive white noise alongside ensemble averaging, CEEMDAN effectively eliminates mode mixing while enhancing the stability and reliability of the decomposition. The steps are as follows:

(1) Add white noise to the original sequence $x(t)$ to generate a new sequence $x_i(t)$:

$$x_i(t) = x(t) + \varepsilon_0 \omega^i(t), i = 1, 2, \ldots n \quad (1)$$

where ε_0 represents the noise coefficient, which controls the amplitude of the added white noise; and $\omega^i(t)$ denotes the white noise sequence added during the i-th iteration.

(2) Decompose each $x_i(t)$ using EMD to obtain the first intrinsic mode function $IMF_1(t)$ and the first residual component $r_1(t)$:

$$IMF_1(t) = \frac{1}{n}\sum_{i=1}^{n} IMF_1^i(t) = \frac{1}{n} EMD_1(x_i(t)) \tag{2}$$

$$r_1(t) = x(t) - IMF_1(t) \tag{3}$$

(3) Continue adding white noise to the sequence to construct a new sequence for calculating the second intrinsic mode function $IMF_2(t)$ and the second residual component $r_2(t)$:

$$IMF_2(t) = \frac{1}{n}\sum_{i=1}^{n} EMD_1(r_1(t) + \varepsilon_1 EMD_1(\omega^i(t))) \tag{4}$$

$$r_2(t) = r_1(t) - IMF_2(t) \tag{5}$$

(4) Repeat the above steps until the residual does not exceed two extrema, at which point no further decomposition is possible:

$$IMF_k(t) = \frac{1}{n}\sum_{i=1}^{n} EMD_1(r_{k-1}(t) + \varepsilon_{k-1} EMD_{k-1}(\omega^i(t))) \tag{6}$$

$$r_k(t) = r_{k-1}(t) - IMF_k(t) \tag{7}$$

(5) Complete the decomposition. At this point, the relationship between the original sequence and the modal components $R(t)$ and the final residual component is given by the following:

$$x(t) = \sum_{k=1}^{K} IMF_k(t) + R(t) \tag{8}$$

2.2. Sample Entropy

Sample entropy is an improvement over approximate entropy, used to more accurately assess the complexity of time series data [19]. It measures the randomness and unpredictability of the data by calculating the probability of similar patterns, thereby avoiding the biases and sample size dependence associated with approximate entropy. A larger sample entropy reflects greater complexity with more irregular changes in the data. Given a time series of length n, $[x(1), x(2), \ldots x(n)]$, the calculation steps for sample entropy are as follows:

(1) Construct a comparison sequence of length m:

$$x_m(i) = [x(i), x(i+1), \ldots x(i+m-1)], i = 1, 2, \ldots n-m+1 \tag{9}$$

(2) For each vector $x_m(i)$ and $x_m(j)$, calculate the distance d between the vectors:

$$d[x_m(i), x_m(j)] = \max[x_m(i+k) - x_m(j+k)], 0 \le k \le m-1 \tag{10}$$

(3) Calculate the similarity probability:

$$A^m(r) = \frac{1}{n-m}\sum_{i=1}^{n-m} \frac{v^m(i)}{n-m+1} \tag{11}$$

$$B^m(r) = \frac{1}{n-m}\sum_{i=1}^{n-m} \frac{\omega^{m+1}(i)}{n-m+1} \tag{12}$$

where r is the acceptance threshold for considering two sequences as similar; v^m is the number of counts where $d[x_m(i), x_m(j)] \le r$; ω^{m+1} is the number of counts

where $d[x_{m+1}(i), x_{m+1}(j)] \leq r$; $A^m(r)$ is the probability that the two sequences match in m dimensions; $B^m(r)$ is the probability that the two sequences match when the dimensions are increased to $m+1$.

(4) Finally, the calculation of sample entropy is defined as follows:

$$SE(m, r, n) = -\mathrm{In}\frac{B^m(r)}{A^m(r)} \tag{13}$$

2.3. VMD

VMD is a time-frequency decomposition method [20] that decomposes a signal into several modal components with specific center frequencies through a variational model. The objective of VMD is to minimize the bandwidth of each modal function, thereby adaptively decomposing the various modes. For a given sequence $x(t)$, the specific steps of VMD are as follows:

(1) Formulate the variational problem. VMD decomposes $x(t)$ into K modal components $u_k(t)$ by solving this variational problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + j\frac{1}{\pi} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{14}$$

where ∂_t represents the partial derivative with respect to t; $u_k(t)$ is the k-th modal component; ω_k is the center frequency of the k-th modal component; $\delta(t)$ is the Dirac delta function; $*$ denotes the convolution operation; j is the imaginary unit. The goal of this variational problem is to minimize the bandwidth of each modal function, that is, by adjusting the modal function and its corresponding center frequency, to minimize the spectral bandwidth of each modal function.

(2) To ensure that the sum of all the modal components equals the original sequence, a constraint is introduced:

$$s.t. \sum_{k=1}^{K} u_k(t) = x(t) \tag{15}$$

(3) Substitute the constraint into the objective function to construct the Lagrange function:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + j\frac{1}{\pi} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 \\ + < \lambda(t), f(t) - \sum_k u_k(t) > \end{aligned} \tag{16}$$

where $\lambda(t)$ is the Lagrange multiplier; α is the second-order penalty factor; $f(t)$ is the original input signal to be decomposed; $*$ is the convolution operation.

(4) Use the Alternating Direction Method of Multipliers (ADMM) to solve the above optimization problem:

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \dfrac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2}, \\ \hat{\omega}_k^{n+1} = \dfrac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}, \\ \hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \end{cases} \tag{17}$$

where τ is the tolerance for noise; and $\hat{u}_k^{n+1}(\omega), \hat{u}_i(\omega), \hat{f}(\omega), \hat{\lambda}(\omega)$ represents the corresponding Fourier transform of $u_k^{n+1}(t), u_i(t), f(t), \lambda(t)$. Repeat the above steps to iteratively optimize u_k, ω_k and λ, until the convergence condition is met, thus decomposing the original sequence $x(t)$ into K modal components.

2.4. LSTM

The LSTM model is a variant of the recurrent neural network (RNN) that can effectively capture long-term dependencies when processing time series data [21]. Its structure is shown in Figure 2. By introducing memory cells and gating mechanisms, LSTM effectively addresses the gradient vanishing and exploding problems commonly encountered by traditional RNNs when handling long sequences.

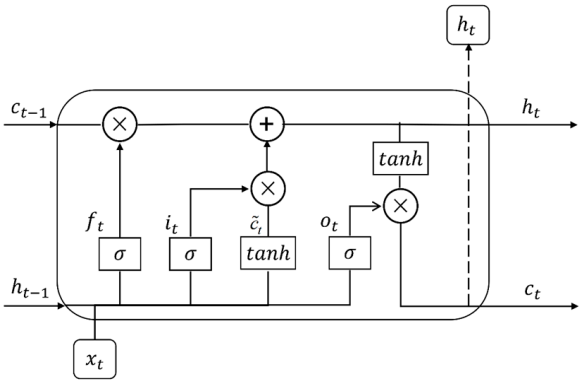

Figure 2. LSTM structure diagram.

LSTM consists of multiple LSTM cells, each of which mainly includes a forget gate, an input gate, and an output gate:

(1) Forget gate: the forget gate f determines which information from the previous time step's memory cell should be retained or discarded.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{18}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{19}$$

where $\sigma(x)$ is the sigmoid activation function; W_f is the weight matrix for the forget gate; b_f is the bias vector for the forget gate; h_{t-1} represents the hidden state from the previous time step, containing information from past inputs in the sequence; x_t is the input at the current time step t.

(2) Input gate: the input gate i controls how information is added to the memory cell, including two steps: selecting the information to be updated and generating the new information.
Selection of the information to be updated:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{20}$$

Generation of candidate information:

$$\widetilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{21}$$

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{22}$$

where tanh is the hyperbolic tangent function; $\widetilde{c}_t$ is the candidate cell state, representing the new information; W_c is the weight matrix for the cell state; b_c is the bias vector for the cell state.

(3) Cell state update: The cell state c_t is updated through the forget gate and input gate. First, the previous time step's cell state c_{t-1} is multiplied by the output of the forget gate, and then the candidate information selected by the input gate is added.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c}_t \tag{23}$$

(4) Output gate: The output gate o determines which part of the cell state c_t will be output, and the hidden state h_t is also generated from this gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{24}$$

$$h_t = o_t \cdot \tanh(o_t) \tag{25}$$

where W_o is the weight matrix for the output gate; b_o is the bias vector for the output gate.

2.5. SHAP Explainable Method

SHAP is a game theory-based explanation method that provides global and local interpretations by calculating feature contributions to model predictions [22]. Applicable to any type of machine learning model, it serves as a model-agnostic explanation tool.

SHAP calculations are based on Shapley values, which are a method in game theory for fairly distributing gains. In machine learning, features are viewed as "participants", and the prediction outcome is the "gain". Shapley values measure the marginal contribution of each feature to the prediction outcome, determining its importance. The formula for calculating SHAP values is as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{26}$$

where ϕ_i represents the SHAP value for the i-th feature; S is the subset of the features that does not include the i-th feature; N is the set of all the features; and $f(S)$ is the output of the model trained using only the feature subset S. Each term in the formula represents the change in the model output before and after including the i-th feature. By taking a weighted average over all the possible subsets of the features, the SHAP values can fairly allocate each feature's contribution to the final prediction.

The relationship between the SHAP values and the model prediction results can be expressed as follows:

$$f(x) = \phi_0 + \sum_i \phi_i \tag{27}$$

where $f(x)$ is the prediction function of the trained machine learning model; x is the input feature vector; ϕ_0 is the baseline output of the model, which is typically the mean of the target variable across all the samples, which is a fixed value; and $\sum_i \phi_i$ represents the model's basic prediction level when no features are considered. This represents the contribution of each feature to the model's prediction, indicating how each feature shifts the prediction result from the baseline output ϕ_0.

The SHAP explanation method has the advantage of enabling each stage of the model's predictions to be linked to specific feature contributions, thereby enhancing the model's interpretability. Users can understand how each feature influences the prediction results through SHAP values, thereby gaining deeper insights into the model's decision-making process.

3. Case Study

3.1. Data Source

In recent years, extremely high temperatures during the summer have frequently occurred in Sichuan, placing immense pressure on the power system and leading to a tense supply–demand situation. Therefore, this study analyzed the 15 min interval total load data (MWh) from the main electricity meter of a residential community in Nanchong, Sichuan, in 2023, to investigate air conditioning load patterns under high summer temperatures in this region.

With the assistance of smart meters, it becomes possible to understand the load consumption patterns of individual household appliances [23]. However, due to the

incomplete deployment of smart meters in China, obtaining user-level air conditioning load data remains challenging. As a result, an indirect method was required to estimate the air conditioning load. The electric load was categorized into the baseline load and temperature-sensitive load, with the air conditioning load significantly influenced by temperature. Based on historical data, the loads in March and April in Sichuan typically excluded the air conditioning load, making them representative of the baseline load. By subtracting this baseline from the loads in July and August, the summer air conditioning load curve was derived, resulting in a total of 5952 data points, as shown in Figure 3.

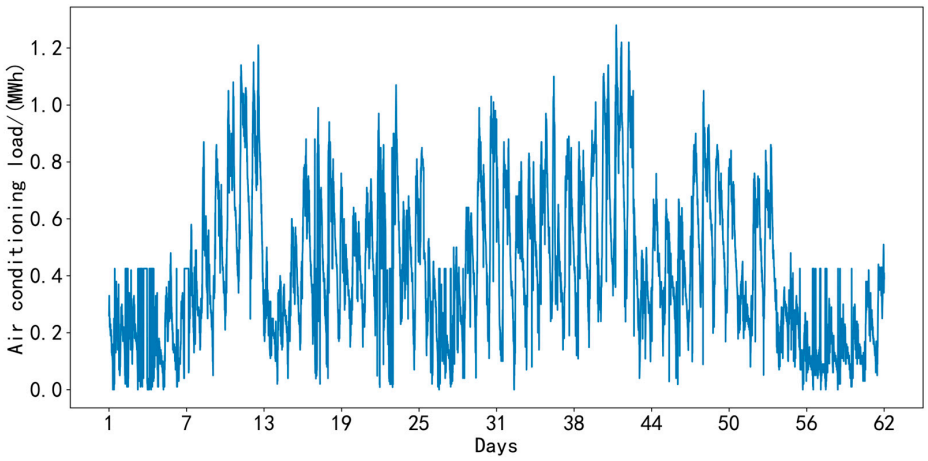

Figure 3. Air conditioning load curve.

As shown in Figure 3, the air conditioning load curve exhibited rapid changes and significant high-frequency fluctuations, demonstrating obvious randomness and volatility. These characteristics complicated the task of load forecasting.

3.2. Evaluation Metrics

Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2) are commonly used evaluation metrics in load forecasting. MAE measures the average deviation between the predicted and actual values, with smaller values indicating higher model accuracy; RMSE emphasizes the impact of large errors, with smaller values reflecting better control over significant deviations; MAPE presents errors in percentage form, with smaller values indicating higher prediction accuracy; R^2 represents the model's ability to explain variations in the data, with values closer to 1 indicating better fit. Together, these four metrics provided a comprehensive assessment of the model's forecasting performance. The calculation formulas are as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{28}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{29}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \tag{30}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{31}$$

where N is the number of samples in the test set; y_i is the actual value of the i-th sample in the test set; $\hat{y}_i$ is the predicted value of the i-th sample in the test set; and $\overline{y}$ is the average value of the samples in the test set.

Due to the presence of zero values in the test set samples of the air conditioning load series, this can lead to very large values for MAPE, which do not accurately reflect the model's prediction accuracy. Therefore, this paper used the Modified Mean Absolute Percentage Error (MMAPE) [24] to replace MAPE, addressing the issue of a zero denominator in MAPE. The calculation formula for MMAPE is as follows:

$$\text{MMAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{\overline{y}} \times 100\% \tag{32}$$

3.3. Air Conditioning Load Data Decomposition

3.3.1. CEEMDAN Decomposition and Sample Entropy Integration

Based on the theory in Section 2, this study employed CEEMDAN to conduct a preliminary decomposition of the original air conditioning load series, with the results shown in Figure 4. Figure 4 illustrates that the high-frequency subsequences exhibited complex variations with unclear trends, while the low-frequency subsequences showed simpler variations with evident trends.

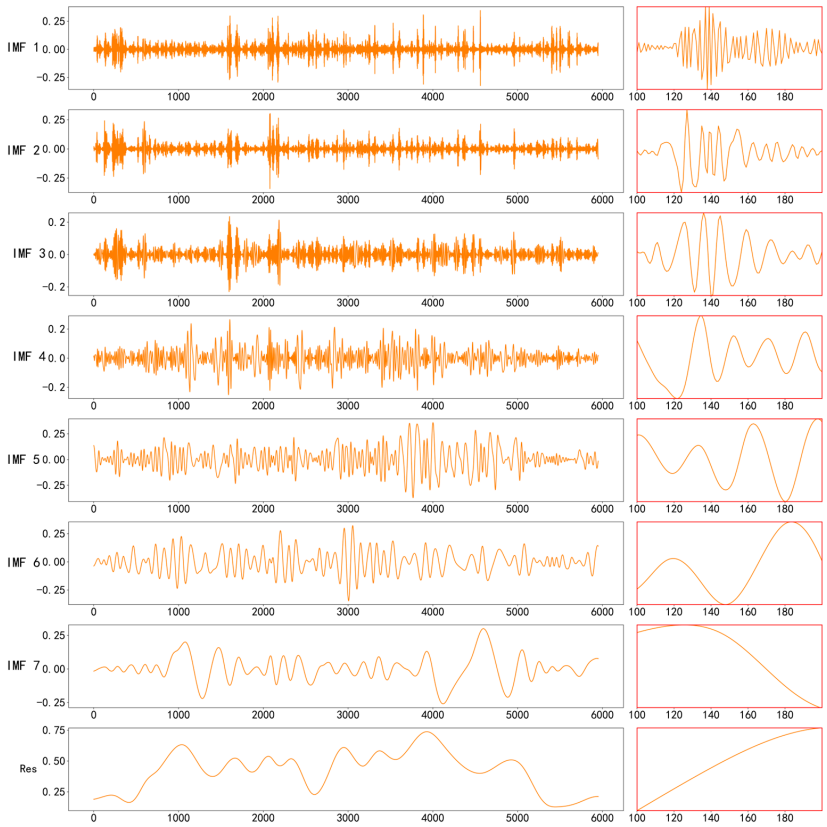

Figure 4. CEEMDAN decomposition results.

To enhance prediction efficiency, the sample entropy was calculated for the subsequences obtained from the CEEMDAN decomposition to assess their complexity. Sample entropy is influenced by two key parameters: m, which represents the length of subsequences used to construct similarity patterns, and r, the threshold for defining similarity between subsequences. Specifically, m controls the length of the patterns being compared; larger values capture more detailed information from the time series but may lead to fewer matches, especially in shorter sequences. The parameter r determines the tolerance level for similarity; smaller values enforce stricter similarity criteria, which may result in

lower entropy values, while larger values provide a more relaxed criterion, yielding higher entropy values.

In this study, subsequences with similar complexities, as indicated by sample entropy, were grouped into high-frequency, medium-frequency, and low-frequency categories. Figure 5 displays the sample entropy results under different m and r settings. Notably, IMF1 and IMF2 consistently showed higher sample entropy values across the different parameters, indicating that they are more complex and difficult to predict. In contrast, IMF7 and IMF8 exhibited lower sample entropy values, suggesting more regular and predictable patterns. Based on these observations, IMF1 and IMF2 were combined into Co-IMF1; IMF3 to IMF6 into Co-IMF2; IMF7 and IMF8 into Co-IMF3, representing high-frequency, medium-frequency, and low-frequency subsequences, respectively.

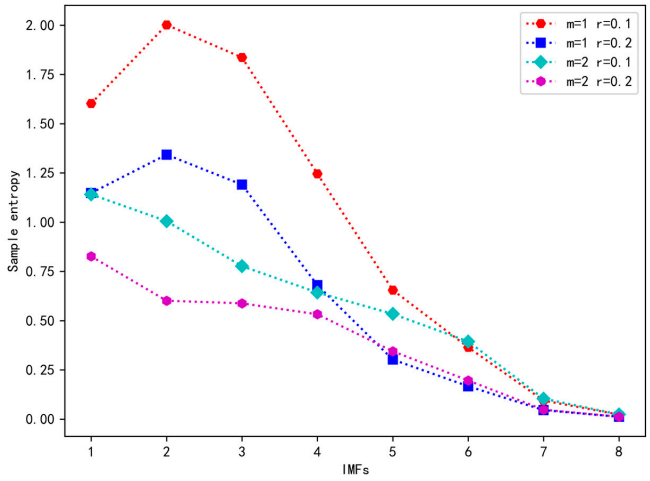

Figure 5. Sample entropy results.

3.3.2. VMD Secondary Decomposition

Figure 6 shows that the high-frequency subsequence Co-IMF1, after the preliminary decomposition using CEEMDAN and integration based on sample entropy, still exhibited high complexity. To reduce its complexity and enhance predictability, this study performed a second decomposition of Co-IMF1 using VMD. The results of the second VMD second are presented in Figure 7.

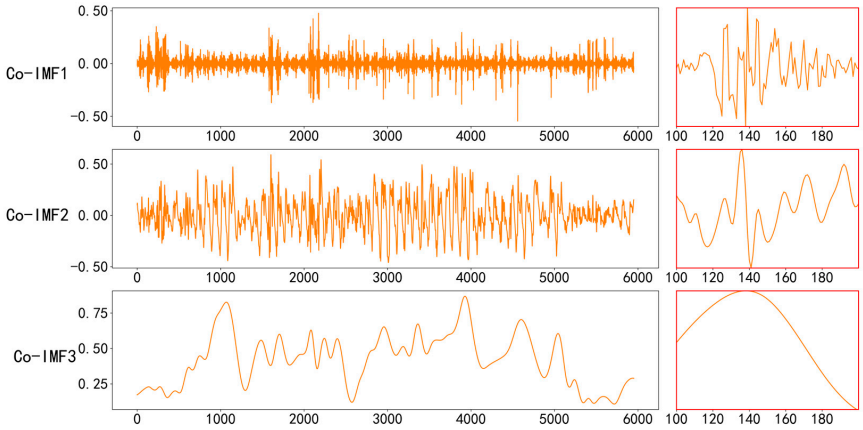

Figure 6. Sample entropy integration results.

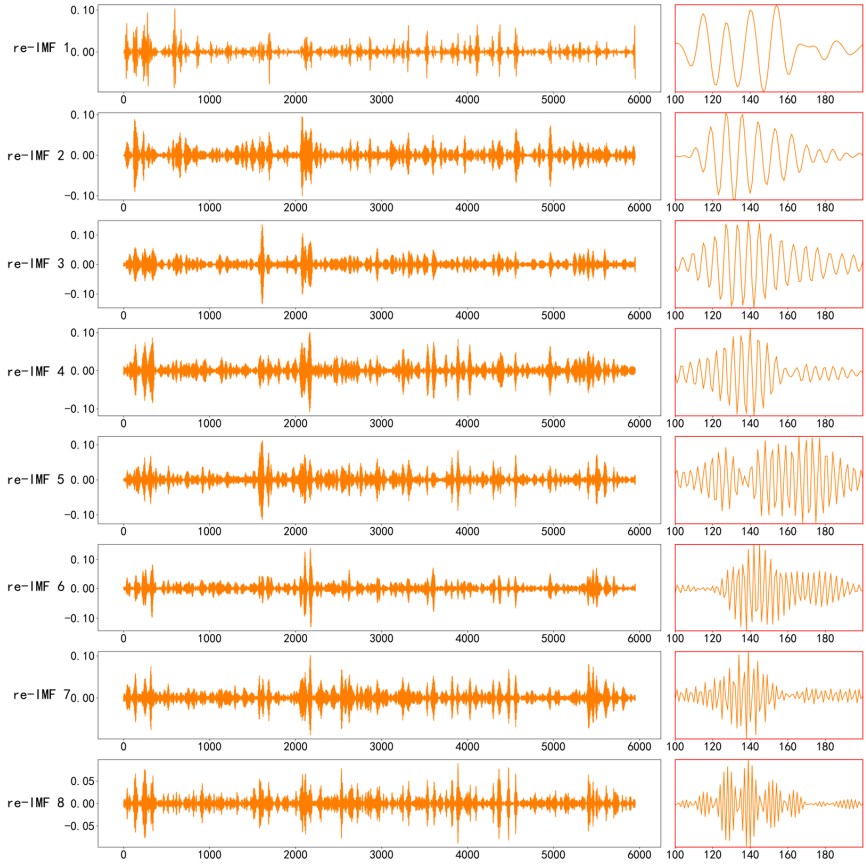

Figure 7. VMD secondary decomposition results of Co-IMF1.

3.4. Air Conditioning Load Prediction Model

This study used LSTM to establish the prediction model, with the hyperparameters optimized using the Optuna library. Optuna is an efficient and flexible hyperparameter optimization tool that can automatically search for the best parameter combinations to enhance model performance. The dataset was divided into training, validation, and test sets in a 7:1:2 ratio. The final hyperparameters determined through Optuna's global optimization and adaptive sampling methods are shown in Table 1, ensuring the optimal performance of the LSTM model in prediction tasks.

Table 1. Hyperparameters of LSTM.

Description	Value
Input Dimension	8
Hidden Layer Size	64
Input Sequence Length	3
Output Sequence Length	1
Number of LSTM Layers	2
Dropout Probability	0.2
Batch Size	32
Learning Rate	0.001
Number of Epochs	100
Optimizer	Adam
Loss Function	MSE Loss

Additionally, since LSTM is sensitive to the scale of input data, the data underwent min–max normalization before being input into the model. This normalization eliminated the dimensional differences between the features, improving the training efficiency and

the stability of the model. After the prediction was completed, to facilitate evaluation and visualization, the predicted values were reverse normalized to restore them to the original data scale, ensuring the accuracy and interpretability of the results.

To accurately capture the multidimensional factors affecting air conditioning load variation, this study incorporated eight features into the LSTM model, including historical load (MWh), temperature (°C), humidity (%), precipitation (mm), wind speed (m/s), wind direction (°), hour (h), and quarter (15 min), as shown in Table 2. This comprehensive selection of the features enabled a thorough consideration of the environmental and temporal influences on air conditioning load, significantly enhancing the accuracy and adaptability of the predictions.

Table 2. Input feature list.

Feature Name
Historical load (MWh)
Temperature (°C)
Humidity (%)
Precipitation (mm)
Wind speed (m/s)
Wind direction (°)
Hour (h)
Quarter (15 min)

3.5. Prediction Results

To validate the effectiveness of the proposed prediction framework, the following models were selected for comparison:

(1) LSTM: no decomposition was performed; predictions were made directly using the original features;
(2) EMD-LSTM: The original air conditioning load sequence was decomposed using EMD, and LSTM was used to predict each component separately. The final prediction result was obtained by summing these predictions;
(3) CEEMDAN-LSTM: The original air conditioning load sequence was decomposed using CEEMDAN, and LSTM was used to predict each component separately. The final prediction result was obtained by summing these predictions;
(4) VMD-LSTM: The original air conditioning load sequence was decomposed using VMD, and LSTM was used to predict each component separately. The final prediction result was obtained by summing these predictions;
(5) CEEMDAN-SE-LSTM: The original air conditioning load sequence was decomposed using CEEMDAN, and then aggregated based on sample entropy to obtain high-frequency, mid-frequency, and low-frequency subsequences. LSTM was then used to predict each of these subsequences separately;

The prediction results are shown in Table 3, and Figure 8 displays the prediction results for 24 randomly selected time steps.

Table 3. Comparison of prediction results.

Model	MAE/(MWh)	RMSE/(MWh)	MMAPE (%)	R^2
LSTM	0.0498	0.0713	18.4275	0.8660
EMD-LSTM	0.0313	0.0487	11.5679	0.9375
CEEMDAN-LSTM	0.0292	0.0464	10.8118	0.9431
VMD-LSTM	0.0213	0.0308	7.8690	0.9750
CEEMDAN-SE-LSTM	0.0331	0.0520	12.2601	0.9286
Proposed	0.0132	0.0184	4.8984	0.9910

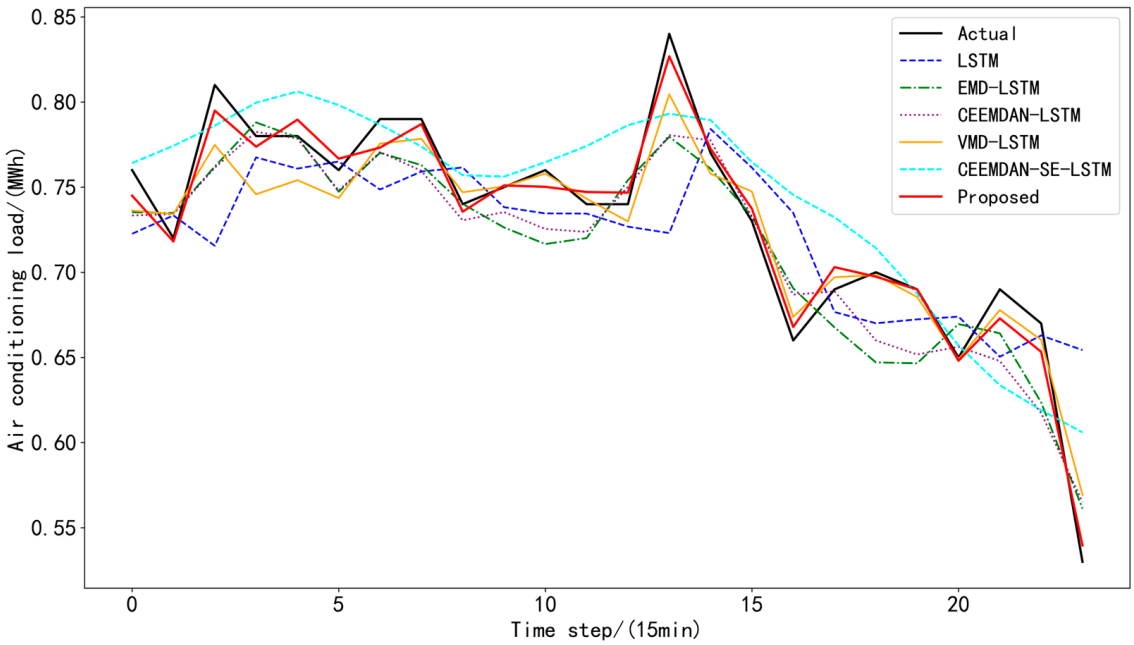

Figure 8. Comparison of prediction results.

From Table 3 and Figure 8, it can be seen that the CEEMDAN-SE-VMD-LSTM method outperformed the LSTM model in terms of prediction accuracy, with improvements of 73.4940%, 74.1935%, 73.4180%, and 14.4342% in the MAE, RMSE, MMAPE, and R^2 metrics, respectively. Compared to VMD-LSTM, these metrics improved by 38.0282%, 40.2597%, 37.7507%, and 1.6410%. This indicates that the multi-level decomposition and aggregation approach effectively utilized the characteristics of the different frequency components, reducing the complexity of the air conditioning load series and significantly decreasing prediction errors. In contrast, the LSTM model without decomposition had the lowest prediction accuracy due to the interference of frequency components and noise in the original data.

Among the combined decomposition prediction methods, CEEMDAN-SE-LSTM had the lowest prediction accuracy. Although sample entropy aggregation simplified the input structure of the model, the complexity of the high-frequency subsequence increased, leading to greater prediction difficulty. On the other hand, VMD-LSTM demonstrated better prediction accuracy than CEEMDAN-LSTM and EMD-LSTM, indicating that VMD was more effective in signal decomposition and better at extracting frequency components that aid in prediction.

Additionally, to provide a more comprehensive visualization of the prediction results, Figure 9 illustrates the predictions for the entire test dataset, along with a zoomed-in view of a selected portion. From the figure, it can be observed that the proposed method (in red) aligned more closely with the actual load curve (in black) across most time steps, particularly excelling in capturing detailed fluctuations. The zoomed-in view highlights that even in regions with more pronounced load fluctuations, the proposed method effectively captured the high-frequency variations in the load curve, demonstrating its superiority in forecasting short-term load dynamics.

This study adopted the strategy of using CEEMDAN for initial decomposition, followed by VMD for secondary decomposition of the high-frequency subsequence. This combination leveraged the strengths of both the decomposition methods, avoiding the potential complexity and performance decline that may arise from directly applying two rounds of VMD. The combination of CEEMDAN and VMD effectively enhanced the prediction accuracy of the model.

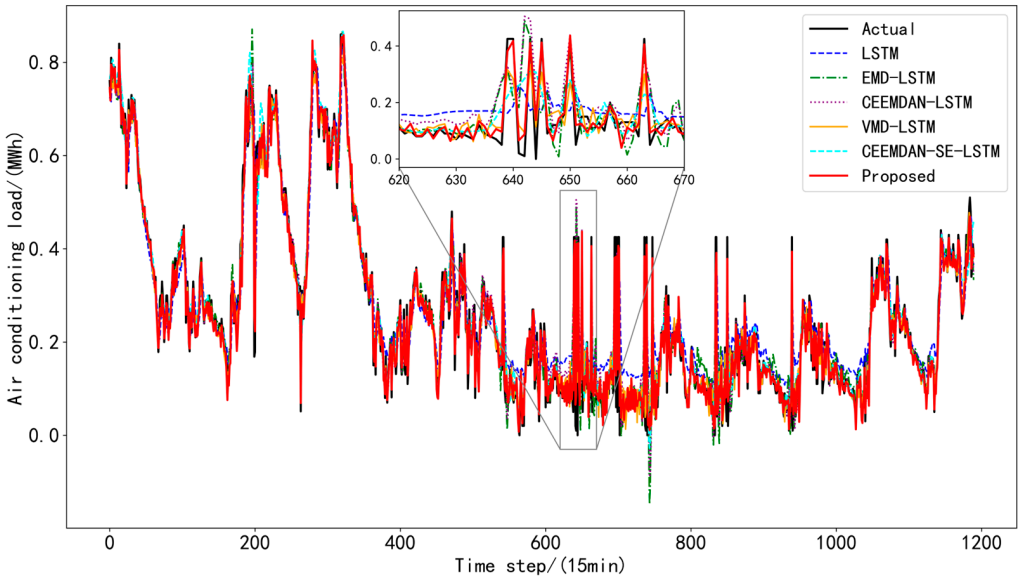

Figure 9. Prediction results on the test set.

3.6. Effect of Sample Entropy Parameter Selection

In this section, we discuss the impact of sample entropy parameter selection on the prediction accuracy.

As shown in Figure 5, the sample entropy results varied significantly with the different parameter selections. The embedding dimension m and tolerance threshold r had a direct impact on the calculated entropy values across the IMFs. This variability indicates that parameter selection is crucial for accurately assessing the complexity of different frequency components, influencing the subsequent prediction accuracy. Appropriate selection of the parameters m and r enabled a more reliable classification of the high-, medium-, and low-frequency components based on their entropy characteristics, thereby improving the prediction accuracy.

With all the other settings remaining constant, the prediction results with only the parameters m and r adjusted are shown in Table 4.

Table 4. Effect of parameter m and r selection on prediction results.

Parameter	MAE/(MWh)	RMSE/(MWh)	MMAPE (%)	R^2
m = 1, r = 0.1	0.0306	0.0412	11.3380	0.9551
m = 1, r = 0.2	0.0149	0.0201	5.5168	0.9894
m = 2, r = 0.1	0.0132	0.0184	4.8984	0.9910
m = 2, r = 0.2	0.0232	0.0323	8.5882	0.9726

As shown in Table 4, when m = 2 and r = 0.1, the model achieved the best prediction performance, with MAE, RMSE, and MMAPE reaching 0.0132, 0.0184, and 4.8984, respectively, and R^2 reaching 0.9910. This indicates that, under this parameter setting, the classification of the frequency components and complexity were more reasonable, thereby improving the prediction accuracy of the model. The next best combination was m = 1 and r = 0.2, where MAE, RMSE, and MMAPE reached 0.0149, 0.0201, and 5.5168, respectively, and R^2 reached 0.9894, which is slightly lower than the case of m = 2 and r = 0.1. In the cases of m = 1, r = 0.1, and m = 2, r = 0.2, the errors were relatively larger, indicating that these two combinations did not balance complexity and stability well in the sample entropy calculation, resulting in reduced prediction accuracy.

From the perspective of m and r, this result can be analyzed as follows: the embedding dimension m represents the length of the subsequences compared in the calculation of sample entropy. A smaller m (e.g., m = 1) can capture finer details and more complex

content but increases sensitivity to noise. In contrast, a larger m (e.g., m = 2) captures more representative patterns, resulting in smoother entropy values.

The tolerance threshold r controls the similarity criterion. A smaller r (e.g., r = 0.1) applies a stricter criterion, capturing finer variations but also amplifying noise, which may affect stability. A larger r (e.g., r = 0.2) provides a more relaxed criterion, effectively filtering high-frequency noise and helping to identify overall trends in the data, although it may miss some detailed features.

Overall, the combination of m = 2 and r = 0.1 performed best, likely because this setting achieved a good balance between capturing the overall structure and fine details of the frequency components. The larger embedding dimension (m = 2) helped to filter out some high-frequency noise, while the smaller tolerance threshold (r = 0.1) preserved the subtle variations in the frequency components, enabling a more effective classification of the high-, medium-, and low-frequency components and ultimately improving the model's prediction accuracy.

3.7. Predictive Results Interpretability

3.7.1. Global Interpretability

Global interpretability aims to provide a comprehensive understanding of the overall impact of various features on the predictive output of a model, rather than just the influence of individual predictions. Since the subsequences obtained from the decomposition algorithms did not fully represent the true air conditioning load levels, this study employed the LSTM model for direct prediction and utilized the SHAP method for model interpretation.

Figure 10 illustrates the global interpretation of features. Each row represents a feature, with the horizontal axis showing the SHAP values, indicating the contribution of that feature to the model output. The colors indicate the magnitude of the feature values, with blue representing lower values and red representing higher values. The distribution range of points reflects the extent of the feature's impact on the prediction results; a wider range indicates a greater influence. The red points on the positive axis signify that higher feature values had a positive impact on the prediction, meaning that, as the feature value increased, the prediction result also increased; conversely, the red points on the negative axis indicate that higher feature values had a negative impact, implying that as the feature value increased, the prediction result decreased.

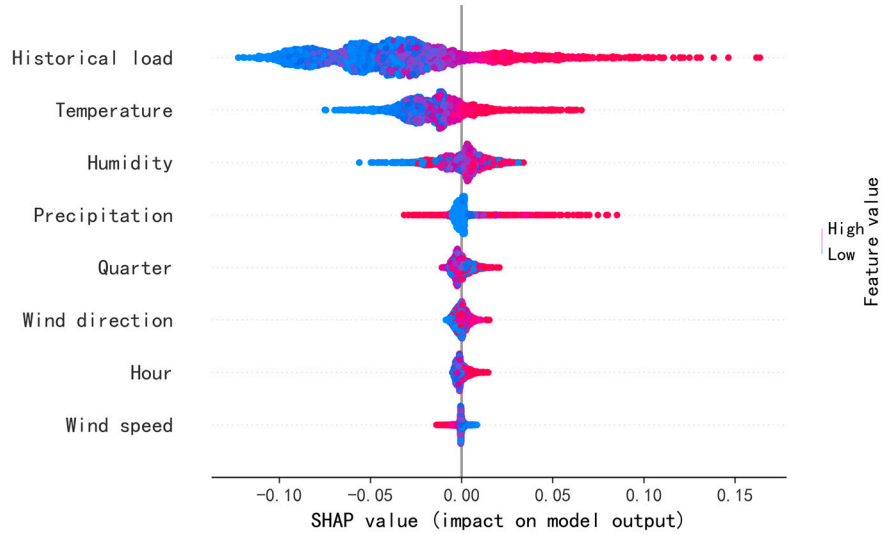

Figure 10. Global interpretability of features.

From Figure 10, it is evident that the primary feature influencing air conditioning load predictions was historical load. The SHAP values indicate that when the historical load was high, the model's predicted values also significantly increased, demonstrating

the clear positive impact of historical load on predictions; conversely, when the historical load was low, the predicted values were correspondingly lower. Additionally, temperature had a notable positive influence on the model output, with higher temperatures leading to increased predicted load values, while lower temperatures resulted in decreased predictions. Humidity ranked as the third most important feature, as high humidity typically leads to an increase in the model's predicted values, while low humidity tends to reduce them. The impact of precipitation was minimal, with the SHAP values close to zero, likely due to the limited rainfall during July and August, which restricted its effect on the model output. Other features such as wind speed, wind direction, hour, and quarter showed a minor influence on the air conditioning load predictions, with the SHAP values clustered closely together, indicating that these features did not significantly contribute to the model's predictive results.

3.7.2. Local Interpretability

Local interpretability aims to explain the results of individual prediction samples by showing how each feature contributes to the prediction for a specific sample. In our air conditioning load forecasting model, SHAP was used to calculate the contribution of features such as temperature, historical load, and humidity for each time step, allowing us to understand which features had the most significant impact on that specific prediction. This approach enhanced the model's transparency and helped us interpret its behavior under various conditions.

Figure 11 shows the local interpretability of a randomly selected prediction sample. Each row represents a feature along with its specific value for this sample (normalized result). At the bottom, the model's expected prediction value is displayed, representing the average predicted output in the absence of any feature inputs. At the top is the model's actual predicted value for this sample, derived from the contributions of each feature interacting together. From Figure 11, it can be seen that the historical load feature contributed +0.06 to the predicted value, and the temperature feature contributed +0.02, while the humidity feature contributed −0.01. This indicates that the historical load had a positive impact on the prediction result for this sample, while the humidity feature had a negative effect on the prediction.

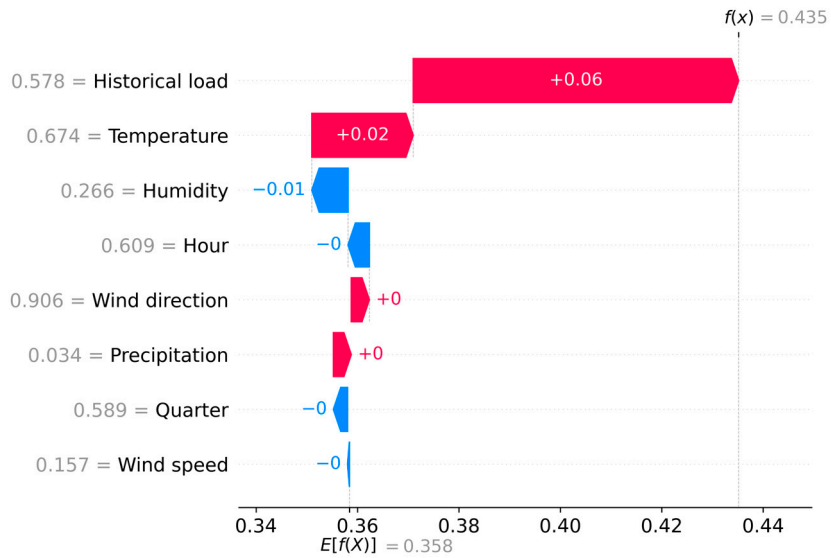

Figure 11. Local interpretability of features.

4. Conclusions

This study proposes an interpretable air conditioning load forecasting framework based on CEEMDAN-SE-VMD-LSTM with SHAP. First, the load series was decomposed using CEEMDAN, while subsequences were classified into high, medium, and low fre-

quencies based on sample entropy. The high-frequency component underwent further decomposition with VMD, and each subsequence was predicted using LSTM to ultimately reconstruct the load prediction results. The SHAP method was then employed to explain the prediction results and reveal the impact of the key features.

The proposed multi-level decomposition strategy, integrating CEEMDAN, sample entropy (SE), and VMD, effectively reduced the data complexity, mitigated the high-frequency noise, and improved the predictive accuracy. Experimental results show that the CEEMDAN-SE-VMD-LSTM model achieves notable gains over LSTM, with reductions of 73.49%, 74.19%, and 73.42% in MAE, RMSE, and MMAPE, and a 14.43% increase in R^2. Compared to VMD-LSTM, these metrics improved by 38.03%, 40.26%, 37.75%, and 1.64%, respectively, underscoring the effectiveness of the decomposition strategy. By leveraging the strengths of both CEEMDAN and VMD, this two-step approach captured essential frequency components, mitigated high-frequency noise interference, and enhanced prediction accuracy.

The SHAP method further strengthened this framework by providing both global and local interpretability of the prediction results. By quantifying the impact of features such as historical load, temperature, and humidity, SHAP offered dual-layer interpretability. Globally, SHAP revealed the overall importance and influence of key features across the entire dataset, offering insights into how various factors generally affect model predictions. Locally, it allows us to understand and trust the specific factors driving individual predictions, enhancing transparency for particular instances. This comprehensive interpretability is crucial for decision-making in energy management, as it enables stakeholders to grasp both broader feature trends and the specific drivers behind predictions.

Although the proposed method performed well in this case study, the dataset was limited to summer air conditioning load data from a single residential area. In real-world applications, air conditioning loads are more diverse, with significant regional and seasonal variations. Future work will therefore expand the dataset to cover multiple seasons and geographic regions, enabling validation of the method's adaptability across varied climatic conditions. Additionally, considering substantial load variations, future research could apply optimization algorithms to fine-tune the parameters within both the decomposition and forecasting stages, enhancing the model's adaptability and ensuring robust accuracy for broader applications.

Author Contributions: X.Y.: Writing—Original draft preparation, Data curation; L.Z.: Visualization; H.Z.: Project administration, Supervision; W.Z.: Investigation, Supervision; C.L.: Formal analysis, Supervision; G.W.: Validation, Supervision; J.Z.: Methodology, Writing—Original draft preparation; X.S.: Conceptualization, Funding acquisition, Writing—Reviewing and Editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Project of State Grid Sichuan Electric Power Company: 521996240005.

Data Availability Statement: The authors do not have permission to share the data.

Conflicts of Interest: Authors Xinting Yang, Chuan Long, and Gang Wu were employed by the State Grid Sichuan Economic Research Institute. Authors Ling Zhang, Hong Zhao, and Wenhua Zhang were employed by the State Grid Sichuan Electric Power Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Duan, H.; Ming, X.; Zhang, X.B.; Sterner, T.; Wang, S. China's Adaptive Response to Climate Change through Air-Conditioning. *iScience* **2023**, *26*, 106178. [CrossRef] [PubMed]
2. He, N.; Qian, C.; Liu, L.; Cheng, F. Air Conditioning Load Prediction Based on Hybrid Data Decomposition and Non-Parametric Fusion Model. *J. Build. Eng.* **2023**, *80*, 108095. [CrossRef]
3. Ejenakevwe, K.A.; Li, S. Investigation of Smart Thermostat Fault Detection and Diagnosis Potential for Air-Conditioning Systems Using a Modelica/EnergyPlus Co-Simulation Approach. *Energy Build.* **2024**, *309*, 114053. [CrossRef]

4. Wang, M.; Hu, E.; Chen, L. Simulation of a Radiation-Enhanced Thermal Diode Tank (RTDT) Assisted Refrigeration and Air-Conditioning (RAC) System Using TRNSYS. *J. Build. Eng.* **2024**, *82*, 108168. [CrossRef]
5. Mi, J.; Fan, L.; Duan, X.; Qiu, Y. Short-Term Power Load Forecasting Method Based on Improved Exponential Smoothing Grey Model. *Math. Probl. Eng.* **2018**, *2018*, 3894723. [CrossRef]
6. Dahl, M.; Brun, A.; Andresen, G.B. Using Ensemble Weather Predictions in District Heating Operation and Load Forecasting. *Appl. Energy* **2017**, *193*, 455–465. [CrossRef]
7. Dhaval, B.; Deshpande, A. Short-Term Load Forecasting with Using Multiple Linear Regression. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 3911–3917. [CrossRef]
8. Yang, X.; Zhou, G.; Ren, Z.; Qiao, Y.; Yi, J. High-Precision Air Conditioning Load Forecasting Model Based on Improved Sparrow Search Algorithm. *J. Build. Eng.* **2024**, *92*, 109809. [CrossRef]
9. Gao, Z.; Yu, J.; Zhao, A.; Hu, Q.; Yang, S. A Hybrid Method of Cooling Load Forecasting for Large Commercial Buildings Based on Extreme Learning Machine. *Energy* **2022**, *238*, 122073. [CrossRef]
10. Zhou, M.; Wang, L.; Hu, F.; Zhu, Z.; Zhang, Q.; Kong, W.; Zhou, G.; Wu, C.; Cui, E. ISSA-LSTM: A New Data-Driven Method of Heat Load Forecasting for Building Air Conditioning. *Energy Build.* **2024**, *321*, 114698. [CrossRef]
11. Wang, H.T. Typical Building Thermal and Thermal Load Forecasting Based on Wavelet Neural Network. *Procedia Comput. Sci.* **2020**, *166*, 529–533. [CrossRef]
12. Wang, L.; Lee, E.W.M.; Yuen, R.K.K. Novel Dynamic Forecasting Model for Building Cooling Loads Combining an Artificial Neural Network and an Ensemble Approach. *Appl. Energy* **2018**, *228*, 1740–1753. [CrossRef]
13. Xu, C.; Chen, H.; Xun, W.; Zhou, Z.; Liu, T.; Zeng, Y.; Ahmad, T. Modal Decomposition Based Ensemble Learning for Ground Source Heat Pump Systems Load Forecasting. *Energy Build.* **2019**, *194*, 62–74. [CrossRef]
14. Zhang, L.; Alahmad, M.; Wen, J. Comparison of Time-Frequency Analysis Techniques Applied in Building Energy Data Noise Cancellation for Building Load Forecasting: A Real-Building Case Study. *Energy Build.* **2021**, *231*, 110592. [CrossRef]
15. Huang, X.; Han, Y.; Yan, J.; Zhou, X. Hybrid Forecasting Model of Building Cooling Load Based on EMD-LSTM-Markov Algorithm. *Energy Build.* **2024**, *321*, 114670. [CrossRef]
16. Karijadi, I.; Chou, S.-Y. A Hybrid RF-LSTM Based on CEEMDAN for Improving the Accuracy of Building Energy Consumption Prediction. *Energy Build.* **2022**, *259*, 111908. [CrossRef]
17. He, F.; Zhou, J.; Feng, Z.K.; Liu, G.; Yang, Y. A Hybrid Short-Term Load Forecasting Model Based on Variational Mode Decomposition and Long Short-Term Memory Networks Considering Relevant Factors with Bayesian Optimization Algorithm. *Appl. Energy* **2019**, *237*, 103–116. [CrossRef]
18. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 4144–4147.
19. Zhou, F.; Huang, Z.; Zhang, C. Carbon Price Forecasting Based on CEEMDAN and LSTM. *Appl. Energy* **2022**, *311*, 118601. [CrossRef]
20. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]
21. Hochreiter, S. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
22. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30, Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774.
23. Stanescu, D.; Enache, F.; Popescu, F. Smart Non-Intrusive Appliance Load-Monitoring System Based on Phase Diagram Analysis. *Smart Cities* **2024**, *7*, 1936–1949. [CrossRef]
24. Zhao, W.; Wei, Y.M.; Su, Z. One Day Ahead Wind Speed Forecasting: A Resampling-Based Approach. *Appl. Energy* **2016**, *178*, 886–901. [CrossRef]