energies

MDPI

Article

A LightGBM-Based Power Grid Frequency Prediction Method with Dynamic Significance–Correlation Feature Weighting

Jie Zhou [1], Xiangqian Tong [1,*], Shixian Bai [1] and Jing Zhou [2]

[1] School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China; 1190311027@stu.xaut.edu.cn (J.Z.); baishixian2008@163.com (S.B.)
[2] International of Faculty of Social Sciences and Liberal Arts, University College Sedaya, Kuala Lumpur 56000, Malaysia; zhoujing970@gmail.com
* Correspondence: lstong@mail.xaut.edu.cn

Abstract

Accurate grid frequency prediction is essential for maintaining the stability and reliability of power systems. However, the complex dynamic characteristics of grid frequency and the nonlinear correlations among massive time series data make it challenging for traditional time series prediction methods to balance efficiency and accuracy. In this paper, we propose a Dynamic Significance–Correlation Weighting (D-SCW) method, which generates dynamic weight coefficients that evolve over time. This is achieved by constructing a joint screening mechanism of feature time series correlation analysis and statistical significance test, combined with the LightGBM gradient-boosting decision tree (GBDT) framework; accordingly, high-precision prediction of grid frequency time series data is realized. To verify the effectiveness of the D-SCW method, this study conducts comparative experiments on two actual grid operation datasets (including typical scenarios with wind/photovoltaic (PV) installations, accounting for 5–35% of the grid); additionally, the Spearman's rank correlation coefficient method, mutual information (MI), Lasso regression, and the feature screening method of recursive feature elimination (RFE) are selected as the baseline control; root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are adopted as assessment indicators. The results show that the D-SCW-LightGBM framework reduces the root mean squared error (RMSE) by 5.2% to 10.4% and shortens the dynamic response delay by 52% compared with the benchmark method in high renewable penetration scenarios, confirming its effectiveness in both prediction accuracy and computational efficiency.

Keywords: grid frequency prediction; Dynamic Significance–Correlation Weighting; LightGBM; time series prediction

check for updates

Received: 26 May 2025
Revised: 18 June 2025
Accepted: 20 June 2025
Published: 24 June 2025

Citation: Zhou, J.; Tong, X.; Bai, S.; Zhou, J. A LightGBM-Based Power Grid Frequency Prediction Method with Dynamic Significance–Correlation Feature Weighting. *Energies* **2025**, *18*, 3308. https://doi.org/10.3390/en18133308

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Power system frequency is the core index for measuring a system's stable operation, and accurate frequency prediction plays a crucial role in preventing large-scale power outage accidents, enhancing the flexibility of grid regulation and safeguarding the reliability of a power supply [1]. In recent years, with the continuous optimization of the power system structure and the evolution of the operating environment [2], especially after the large-scale grid connection of new energy sources, the industry has put forward higher requirements for the accuracy and dynamic response of frequency prediction. However, the stochastic volatility of new energy sources such as wind and photovoltaic power, as

well as the strong nonlinearity of power electronic equipment, lead to significant non-stationary characteristics of grid frequency response on the time scale [3], which exacerbate the demand for real-time, accurate, and adaptive capabilities among prediction models. Therefore, effectively mining and utilizing informative features from frequency series data to capture nonlinear and correlated dynamic changes is essential for the development of advanced, data-driven regulation systems, which can provide real-time, adaptive control in power grids with high renewable energy integration, surpassing the limitations of traditional static or manually tuned regulation methods.

Feature selection methods in existing time series forecasting studies generally face multiple challenges, such as insufficient efficiency and limited adaptability. Common feature selection methods mainly include those based on a single statistical index (e.g., Pearson correlation coefficient, mutual information) [4,5], embedded methods (e.g., Lasso, Ridge regression), and encapsulated algorithms (e.g., recursive feature elimination, RFE) [6–9]. However, in real complex and dynamically nonlinear application environments, each of these methods has its obvious shortcomings. Specifically, although a single statistical index is easy to implement, it can only reflect simple linear or monotonic nonlinear relationships, and it is difficult to reflect the complex dynamic correlations in power systems. In addition, in response to the limitations of traditional methods, some scholars have also explored the integration of multiple feature selection strategies in recent years, such as combining Lasso with RFE, maximum correlation minimum redundancy (mRMR) with RFE, and other methods [10,11], with the intention of taking the respective advantages of different strategies into account. However, such fusion methods often suffer from increased algorithmic complexity, increased difficulty in parameter adjustment, and limited applicability to specific scenarios, etc.; the stability and effectiveness of their screening results in highly redundant features or non-stationary dynamic sequence environments still need to be further improved.

Current time series prediction methods are mainly divided into two major categories [12]: One is the traditional time series analysis methods based on probability statistics (e.g., ARMA, Prophet) [13–15]; such methods have high computational efficiency, but, due to their linear assumptions and parametric modeling mechanisms, it is difficult to effectively capture the nonlinear interactions and dynamic coupling relationships that exist among multidimensional features. The other category is based on data-driven time series analysis methods, among which intelligent time series prediction methods based on shallow machine learning (SML) and deep learning (DL) have become a hot research topic in recent years. SML methods usually use signal processing techniques to extract handmade features after data acquisition; they complete final regression prediction using different SML models, such as decision trees and support vector regression (SVR) [16–18]. On the other hand, DL methods can automatically extract abstract and effective features from the original signals through multi-layer neural network structures (e.g., LSTM, Transformer) [19,20] and have better prediction accuracy. Due to the "black-box" nature of the DL model, it lacks sufficient physical interpretation capability, and the training process of deep neural networks requires a large amount of data, time, and computational resources, which limits their wide application in real-world scenarios.

In contrast, models based on shallow machine learning are more concise in structure and consume fewer computational resources. However, such methods are prone to feature redundancy and overfitting problems when dealing with multidimensional manually extracted features, and are not sufficiently adaptive to time-varying system characteristics, which in turn affects the accuracy of prediction. Therefore, achieving a good balance between computational efficiency, prediction accuracy, and model generalization ability is still a key issue that needs to be solved in the field of time series prediction.

To address the above problems, this paper proposes a dynamic weight assignment mechanism based on multidimensional feature fusion—the Dynamic Significance–Correlation Weighting (D-SCW) method. The method generates adaptive weight coefficients that evolve over time by constructing a joint screening framework of feature statistical significance and time series correlation analysis, and deeply couples it with the LightGBM gradient-boosting decision tree (GBDT) framework to form a real-time response model oriented to the dynamic change in power grid frequency. This fusion strategy retains the efficient parallel computing capability of LightGBM, effectively mitigates the bias problem of a single indicator (e.g., relying on correlation coefficients only) in the traditional feature selection method through the two-dimensional feature screening criterion, and finally significantly improves the accuracy of the time series prediction of the frequency of the power system through the dynamic adaptive adjustment of the feature weights.

The main contributions of this paper include the following: (1) The deep combination of the Dynamic Significance–Correlation Weighting (D-SCW) method with LightGBM, which not only makes full use of the inherent advantages of the gradient-boosting decision tree in modeling nonlinear relationships, but also realizes the efficient processing of the multidimensional dynamic features generated by the D-SCW, and thus enhances overall computational efficiency; meanwhile, the dynamic weight allocation mechanism of D-SCW can effectively evaluate and utilize the feature information so that the model can flexibly adapt to the time-varying characteristics of the grid frequency, thus enhancing the accuracy and timeliness of the prediction results. (2) Global optimization of the key hyperparameters of the LightGBM model (e.g., the learning rate, the depth of the tree, the feature sampling ratio, and the number of leaf nodes) is carried out by utilizing grid search algorithms to further improve the prediction accuracy. (3) Validation of the effectiveness of the framework is performed, with two real novel power system frequency monitoring datasets; the experimental results show that, compared with the traditional feature selection method, the proposed method significantly improves the prediction performance index in different periods. (4) The framework can support real-time updates to the frequency prediction at the minute level, which provides a scalable solution for the scenarios of high percentage of renewable energy connected to the power grid.

This paper is structured as follows: the second part presents the theoretical basis and algorithmic steps of the D-SCW method, the principle of the LightGBM model, and the hyperparameter optimization method. The third part gives the data processing procedure and the prediction results of the model. The prediction performance of the proposed framework is evaluated by comparing it with other feature selection algorithms. Conclusions are drawn in the fourth part.

2. Methods

This section outlines the framework of the prediction method based on the fusion of Dynamic Significance–Correlation Weighting (D-SCW) and the LightGBM model. First, the phase space reconstruction technique and date–time information extraction method are used to expand the features of the original grid frequency time series dataset, and multidimensional features are generated to construct the time series prediction model; second, the correlation strength between the newly generated feature sequences is calculated based on the Spearman's rank correlation coefficient method, and the statistical significance level (p-value) of the features is evaluated to form the two-dimensional features. Then, the criteria are screened to construct a feature subset; again, a dynamic feature weight adjustment mechanism is introduced to dynamically respond to the time-varying fluctuation characteristics of the grid frequency. Finally, the LightGBM model is hyperparameter-tuned

and trained. Figure 1 demonstrates the technical route of the proposed D-SCW-LightGBM fusion framework in grid frequency prediction.

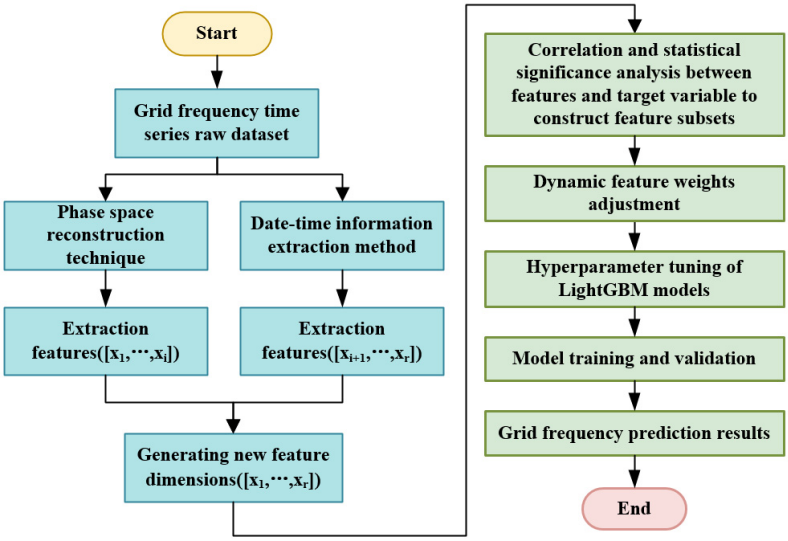


Figure 1. Schematic diagram of the D-SCW-LightGBM model.

2.1. Phase Space Reconstruction

The main goal of phase space reconstruction is to extract more valuable information from a set of time-varying observations by mining the time-dependent relationship between the observations. In the field of time series analysis, phase space reconstruction is an important technical tool, the core of which is to construct an m-dimensional vector representation by introducing different time delays for a given one-dimensional time series data, to capture the data's dynamic characteristics and time dependence accurately.

The technique involves two core parameters: the embedding dimension, m, and the delay time, τ. The embedding dimension, m, determines how many time points from the original time series should be selected to construct a new state vector. The delay time, τ, determines the interval between selecting neighboring time points. By adjusting these two parameters, the behavioral patterns of nonlinear dynamic systems can be analyzed and understood more effectively, to accurately capture the intrinsic features and potential patterns of the time series data and lay a solid information foundation for subsequent data analysis and prediction work. The embedding vector corresponding to the time index, t, can be constructed as follows:

$$X_t = [x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-(m-1)\tau)] \quad (1)$$

where $x(t)$ denotes the original time series data. Through this transformation, the original one-dimensional time series is converted into an m-dimensional state space trajectory that can reveal the structure and dynamic characteristics of the original time series, which helps to understand the intrinsic laws of the data better. At the same time, this also enables the data to fit their own temporal characteristics better, so that they can be better adapted to the temporal nature of the data, thus improving the accuracy and effectiveness of time series prediction.

Based on the basic theory of phase space reconstruction, to ensure the scientificity and accuracy of the subsequent research and analysis, it is necessary to clarify two core parameters: the delay time, τ, and the embedding dimension, m. The methods for determining these two parameters mainly include the following: the delay time τ determination technique is based on the average mutual information (AMI) method or the autocorrelation

function (ACF); the embedding dimension, m, is based on False Nearest Neighbor (FNN) or the Cao method. Among them, the ACF method is suitable for data with strong linear correlation, and its computational process is simple but may ignore nonlinear correlation features, while the AMI method takes nonlinear correlation into account and is suitable for complex nonlinear system analysis. The FNN method is more sensitive to the noise of the data, and the judgment threshold needs to be set manually in practice, which is more subjective, while the Cao method is an improvement of the FNN method, which does not require manual threshold setting and has significant advantages in the processing of noisy data.

For the power system's dynamic characteristics, this paper first adopts the AMI method to determine the delay time, τ, to capture the nonlinear structure inside the data. After determining the delay time, the Cao method determines the embedding dimension, m. This ensures that the system's dynamic characteristics are fully revealed and effectively avoids the computational complexity surge and "dimensional catastrophe" brought by high-dimensional reconstruction.

AMI is a method for determining the delay time, τ, by measuring the nonlinear dependence between two time points. The core idea is that, when the value of the delay time, τ, is taken (so that the value of mutual information between the state variable, $x(t)$, at the current moment and its state variable, $x(t-\tau)$, after a lag of τ time steps reaches a local minimum for the first time), it indicates that the information redundancy between these two time points reaches a theoretical minimum. At this time, the determined delay time, τ, can maximize the retention of independent dynamic information in the system dynamics evolution process, which not only avoids the problem of strong correlation of high-dimensional phase space coordinates caused by too short a delay time, but also overcomes the phenomenon of dynamics information breakage triggered by too long a delay time. The correlation formula is shown below:

$$I(\tau) = \sum_{x(t),x(t-\tau)} p(x(t), x(t-\tau)) \log \frac{p(x(t), x(t-\tau))}{p(x(t))p(x(t-\tau))} \tag{2}$$

where $I(\tau)$ denotes the mutual information of the original sequence, $x(t)$, and the delayed sequence, $x(t-\tau)$, and $p(x(t))$ and $p(x(t-\tau))$ are the marginal probability distributions of the original sequence, $x(t)$, and the delayed sequence, $x(t-\tau)$, respectively.

The Cao method is a widely used embedding dimension determination technique for nonlinear time series analysis, and its theoretical basis is to determine the optimal embedding dimension of the system by exploring the stability of the structural change in the distance structure between neighboring points in the reconstruction of sequences in high-dimensional phase space when the embedding dimension is gradually increased. Specifically, the Cao method uses two curves, $E1(m)$ and $E2(m)$, which change dynamically with the number of dimensions, as the key basis for determining the optimal embedding dimension. The $E1(m)$ curve is mainly used to reflect the change in the degree of freedom of the observation system, and if $E1(m)$ tends to saturate with the increase in the embedding dimension, m, the dimension m corresponding to its inflection point is the appropriate embedding dimension; the $E2(m)$ curve is used to assist in distinguishing between the deterministic and stochastic nature of the sequences, and the Cao method analyzes the change characteristics of the distance structure between neighboring points in the reconstructed space when the embedding dimension increases from m to $m+1$. The change in distance

structure is characterized. For each reconstructed vector, $X_t(m)$, in the m-dimensional space, its nearest neighbor vector, $X_{j(t)}(m)$, is found, and the distance ratio is defined as:

$$a(t,m) = \frac{\left\| X_t(m+1) - X_{j(t)}(m+1) \right\|}{\left\| X_t(m) - X_{j(t)}(m) \right\|} \tag{3}$$

For all moments averaged, the indicator $E1(m)$ is obtained:

$$E1(m) = \frac{1}{N - m\tau} \sum_{t=1}^{N-m\tau} a(t,m) \tag{4}$$

where N denotes the total length of the original time series, and N-$m\tau$ constructs the available number of phase space vectors. When the $E1(m)$ curve gradually converges to a specific stable value, it indicates that the system structure will not undergo a new change if it continues to expand to higher dimensions based on this embedding dimension. Then, the embedding dimension can fully characterize the system's dynamics at this point, and there is no need to increase the number of dimensions further to increase the complexity of the analysis.

A real-time data series may contain both deterministic signals and random noise components, which are significantly different in their dynamic properties. To effectively distinguish between deterministic signals and random noise, and thus analyze the system's dynamical behavior more precisely, the standardized distance metric $E2(m)$ is introduced. This metric provides a powerful quantitative basis for judging the deterministic or random nature of the time series through the standardized measure of the distance between neighboring points in the phase space under different embedding dimensions, which in turn assists in optimizing the selection process of the embedding dimensions.

$$E2(m) = \frac{1}{N - m\tau} \sum_{t=1}^{N-m\tau} b(t,m) \tag{5}$$

$$b(t,m) = \frac{|x(t - m\tau) - x(j(t) - m\tau)|}{\left\| X_t(m) - X_{j(t)}(m) \right\|} \tag{6}$$

Theoretically, for deterministic systems, the $E2(m)$ curve will converge to a constant and not equal to 1 as the embedding dimension m increases, while for random signals, the value of $E2(m)$ will gradually converge to 1. Therefore, based on the convergence characteristics of the $E1(m)$ metric mentioned above, and combined with the differentiated performance of the $E2(m)$ metric in deterministic systems and random signals, the optimal embedding dimension m is determined comprehensively. This approach balances spatial unfolding and signal property differentiation and is a commonly used and efficient criterion in phase space reconstruction.

2.2. Date–Time Information Extraction

In time series data analysis, the parsing and utilization of date–time information play a key role in improving prediction performance. Specifically, date–time information extraction is a refined disassembly of timestamps, and the structured parsing technique can isolate the time dimension features such as year, month, day, week, etc., from the raw data. These normalized time attributes not only carry the temporal positioning information of the data records but also contain the underlying logic that reveals the cyclical patterns. Among them, the daily dimension features can capture the daily fluctuation patterns, the

weekly dimension features can reflect the cyclical operation laws, and the yearly dimension features are significantly associated with the seasonal climate factors.

In this paper, temporal features can be systematically integrated into the construction of the forecasting model. Introducing these temporal elements not only enriches the dimensionality of the input space, enabling the model to capture different layers of cyclical features through multi-scale time windows, but also helps distinguish short-term stochastic fluctuations from long-term trend changes. By using the parsed temporal feature vectors as auxiliary inputs into the prediction model architecture, the model's ability to recognize composite periodic patterns in time series can be effectively enhanced. Applying this approach to grid frequency prediction scenarios helps the model more effectively capture the daily periodicity, intra-weekly variations, and seasonal variations in grid frequency fluctuations that are affected by external factors such as climate. At the same time, the model can comprehensively reflect the frequency dynamics characteristics at multiple time and space scales, thus improving the accuracy and generalization ability of the prediction.

2.3. Dynamic Significance–Correlation Weighting Method

The D-SCW method is a feature importance assessment method that takes into account the strength of the associations between features and target variables and statistical significance. The method firstly adopts the absolute value of Spearman's correlation coefficient to measure the correlation strength between features and target variables, and selects the features with strong correlation to ensure the stability of the prediction model; at the same time, it introduces the statistical significance test (p-value) as an auxiliary criterion, and incorporates the features with lower correlation but stronger statistical significance into the feature set, to avoid omitting the potentially effective information. In the weight calculation stage, the Spearman correlation coefficient of the screened features is normalized using the normalization method, and the weighting coefficient matrix corresponding to the feature set is finally generated. By dynamically adjusting the feature weight parameters, the model can more accurately portray the differences in the contribution of different features to the prediction target, thus optimizing the model's ability to fit complex time series relationships.

2.3.1. Significance–Correlation Weighting Method

The SCW method proposed in this paper first implements feature correlation analysis and a statistical significance test on the original feature set X and target variable Y. The absolute value of Spearman's rank correlation coefficient $|r|$ between each feature and the target variable is calculated to measure the strength of its monotonic association; meanwhile, the corresponding significance p-value is obtained for assessing the statistical significance of this correlation. In this paper, the initial screening strategy of features based on the correlation coefficient threshold is adopted to prioritize the features with significant monotonic associations with the target variables to form the initial screening feature set:

$$S_p = \{X_i | |r(X_i, Y)| \geq \theta\} \tag{7}$$

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{8}$$

where θ is the correlation screening threshold, the value of which needs to be determined according to the actual situation; r is the Spearman rank correlation coefficient, which is used to measure the strength of the monotonic relationship between the two variables; d_i is the absolute value of the difference in rank between the two variables in the i-th sample; and n is the total number of samples.

Based on completing the initial screening of features, this paper further combines the statistical significance test with the supplementary screening of features, aiming to mine those potential features with weak correlation but statistical significance. First, the significance *p*-value of all features in the initial screening feature set S_p is calculated, and its maximum value p_{max} is determined. Then, this p_{max} value is used as the supplementary screening threshold to screen the variables that satisfy the *p*-value of the features less than or equal to p_{max} from the remaining unselected features, and the final auxiliary screening feature set, S_s, is constructed with the following specific mathematical expressions:

$$S_s = \left\{ X_j \middle| X_j \notin S_p, p(X_j) \leq p_{\max} \right\} \tag{9}$$

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{10}$$

The corresponding *p*-value can be obtained by consulting the t-distribution table, thus measuring the feature's significance level.

Through the above joint screening strategy, not only can we prioritize the retention of features that are closely related to the target variables and statistically significant, but we can also make up for the potentially effective variables that may be missed when relying only on the correlation coefficient threshold screening, to enhance the characterization ability of feature engineering and the generalization performance of the prediction model.

2.3.2. Dynamic Feature Weighting Mechanism

The SCW method takes the absolute value of the normalized correlation coefficient as the feature weight. It applies this weight to the original features by element-by-element multiplication to realize the weighted adjustment of feature importance. Its basic computational expression is as follows:

$$X_w = X \cdot \omega^T \tag{11}$$

where X is the input feature data and ω denotes the feature weights obtained after normalization.

To better express the importance of the change in time series features, this paper adopts a dynamic correlation weighting method that combines the memory time series decay mechanism. The method forms a collection of data slices with local time series coverage by dividing the time series data into sliding windows of length W and realizing the continuous sliding of the windows with the time step, Δt. On this basis, the Spearman rank correlation coefficients are recalculated for the features and the target variables within each sliding window to capture the dynamic importance of the features in the local time period. By introducing the temporal decay factor, λ ($0 < \lambda \leq 1$), the importance of the current features is combined with the historical weights, achieving a smooth transition of weights across different windows in the temporal dimension. This enhances the temporal continuity and robustness of the weight sequence, preventing abrupt changes caused by noise. This approach is similar to the exponential smoothing method used in time series analysis [21]. The mathematical expression for dynamic weight updating is as follows:

$$\omega(t) = (1-\lambda)|r(X,Y)| + \lambda \cdot \omega(t-1) \tag{12}$$

where $\omega(t)$ is the feature weight of the current window and $|r(X,Y)|$ is the Spearman rank correlation coefficient of the features within the window. The decay factor, λ, is used to weigh the degree of influence of the current statistical significance against the historical weights. When $\lambda = 0$, the method only relies on the current window's significance analysis results for feature weighting; when $\lambda = 1$, it completely retains the previous moment's historical weights and ignores the current window's feature significance changes.

To ultimately achieve comparability and stability of feature weighting, here, we conduct min–max normalization on the dynamic weight sequence and linearly map it to the [0, 1] interval. Thus, we eliminate the scale difference of weight magnitude under different time windows and provide standardized inputs for subsequent feature selection and model training.

$$\widetilde{\omega}(t) = \frac{\omega(t) - \min(\omega)}{\max(\omega) - \min(\omega)} \tag{13}$$

In summary, the dynamic significance weighting method proposed in this paper can take into account the dynamic changes in the relationship between features and target variables over time and the smoothing effect of historical information, giving weighted features higher stability and effectiveness in the actual model construction.

2.4. LightGBM Model

The Light Gradient-Boosting Machine (LightGBM) is a decision-tree-based gradient-boosting framework. As an implementation of the gradient boosting decision tree (GBDT), LightGBM shows remarkable results in time series forecasting, especially in predicting grid load. Compared with traditional forecasting methods, LightGBM is not only faster to train but also more accurate in forecasting.

In machine learning, a dataset is usually denoted as $\{X, y\}$, which contains l samples; $x_k = (x_k{}^1, x_k{}^2, \ldots, x_k{}^r)$ represents the k-th input vector containing r features; y_k denotes the corresponding label or target value.

At the beginning of the gradient-boosting algorithm, the model needs to be initialized. At this stage (i.e., round 0 iteration), an initial prediction model needs to be set, usually with a simple model, such as a constant model. This initial model will serve as the starting point for the entire gradient-boosting algorithm, and subsequent iterations will continuously add new learners to this starting point to gradually improve the model's predictive performance. This initialization aims to find an optimal constant value that can minimize the total loss of the sample data. The equation to obtain the initial constant value is shown below:

$$F_0(x) = \text{argmin}_\gamma \sum_{k=1}^{l} L(y_k, \gamma) \tag{14}$$

where $F_0(x)$ is the predicted value at the initial state in the gradient-boosting algorithm, γ is a constant, and $L(y_k, \gamma)$ is the loss function.

In each iteration of the gradient-boosting algorithm, the residuals for each sample point need to be calculated. It is based on how well the current model fits the training data to guide the next step of adjusting the model to minimize the error. The model is updated by calculating the residuals along the direction of the gradient. The expression for calculating the residuals is shown below:

$$r_{kp} = -\left[\frac{\partial L(y_k, F(x_k))}{\partial F(x_k)}\right]_{F(x)=F_{p-1}(x)} \tag{15}$$

where r_{kp} denotes the residual value of the p-th iteration for the k-th sample point, and $F_{p-1}(x_k)$ is the model's predicted value for the sample point x_k after the p-1st round of iterations.

After each iteration, a base learner needs to be fitted to the residuals of the current round. The basic learner (e.g., decision tree) is trained by constructing a new training set $\{(x_k, r_{kp})\}_{k=1}{}^l$ consisting of all training samples, x_k, and their corresponding residuals, r_{kp}. The goal of this basic learner is to predict the residual values of each sample as accurately as possible, thus gradually improving the model's predictive performance. The predictions of

the basic learner are adjusted to better fit the target, i.e., the pseudo-residuals, by calculating the output multiplier, γ_p, of the basic learner in each iteration to minimize the overall loss.

$$\gamma_p = \text{argmin}_\gamma \sum_{k=1}^{L} L(y_k, F_{p-1}(x_k) + \gamma h_p(x_k)) \tag{16}$$

where γ_p denotes the output multiplier of the base learner in the p-th iteration, and $L()$ is the loss function to measure the difference between the true label, y_k, and the model prediction, $F_{p-1}(x_k) + \gamma h_p(x_k)$.

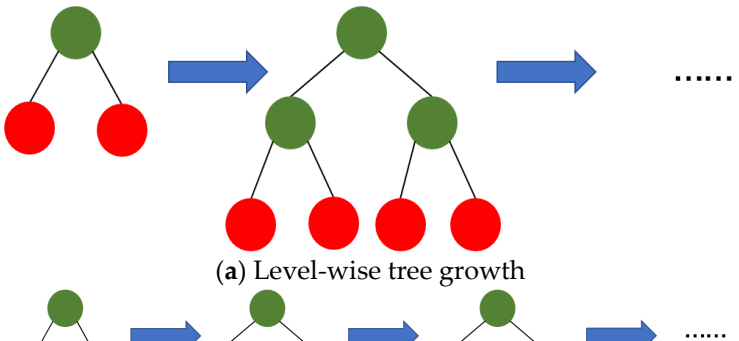
An optimal value of γ_p is found through the above equation such that the overall loss L is minimized when the prediction of the p-th basic learner $h_p(x)$ is multiplied by this γ_p and added to the model prediction $F_{p-1}(x)$ of the previous p-1 rounds. This γ_p value determines the relative importance of each model in the final prediction. After obtaining the weights, γ_p, for the current model, the model is updated as follows:

$$F_p(x) = F_{p-1}(x) + \gamma_p h_p(x) \tag{17}$$

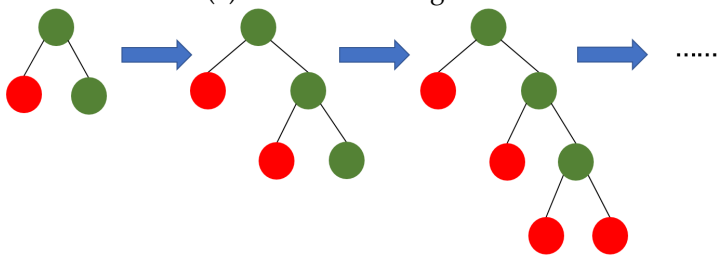
where $F_p(x)$ denotes the model's predicted value for the input sample after the p-th iteration; $h_p(x)$ denotes the predicted value of the p-th base learner for the input sample.

This process is repeated iteratively until a specified number of iterations is reached or a convergence condition is satisfied. By performing this iterative process, the GBDT algorithm can effectively reduce the prediction error and improve the accuracy of the prediction.

LightGBM aims to improve the training efficiency and model performance of GBDT. Its core lies in combining the Gradient-based One-Side Sampling (GOSS) strategy, which prioritizes the selection of the most representative high-gradient instances from the samples to improve the information utilization rate during the training process; at the same time, the continuous features are discretized by the histogram-based algorithm, which effectively reduces the computational and memory overheads. In addition, LightGBM adopts a leaf-based growth strategy for tree structure optimization, enabling the model better to capture the complex distribution features of the data, thus further improving the prediction accuracy. Figure 2 demonstrates the two strategies of hierarchical growth and leaf growth, where green nodes represent decision nodes and red nodes denote leaf nodes.



(**a**) Level-wise tree growth



(**b**) Leaf-wise tree growth

Figure 2. Comparison of level-wise and leaf-wise growth strategies.

These properties make LightGBM the model of choice for many machine learning tasks. This paper uses LightGBM to model the relationship between temporal features

extracted during the feature engineering phase and the target frequency. Specifically, the model is trained with historical grid frequency data and its corresponding temporal features as inputs and is used to predict future grid frequency values.

2.5. Hyperparameter Tuning of LightGBM Models

In machine learning applications, the reasonable determination of hyperparameters is crucial in improving the model's prediction performance. LightGBM, an efficient gradient-boosting decision tree algorithm, contains multiple adjustable hyperparameters. To further improve the model's performance, optimizing some of the key hyperparameters is often necessary. In the LightGBM hyperparameter tuning process in this paper, four key parameters, 'num_leaves', 'learning_rate', 'max_depth', and 'feature_fraction', are optimized, while the remaining parameters are set by default. These core parameters have a significant impact on the expressiveness, convergence speed, and generalization performance of the model, and are the main factors determining the performance of LightGBM. In contrast, parameters such as 'bagging_fraction', 'lambda_l1', and 'lambda_l2' have relatively small impacts on the model performance, and the default values of LightGBM are reasonable enough in most scenarios, so they are not taken as the focus of this tuning. By focusing on the optimization of core parameters, we improve the efficiency of parameter tuning and effectively reduce the computational overhead caused by too many parameter combinations.

Given the limited number of parameters and their ranges in this experiment, grid search was chosen to exhaustively explore all combinations within reasonable computational resources, ensuring identification of the global optimum. Compared to Bayesian optimization, grid search is more intuitive and easier to implement and analyze. While Bayesian optimization offers higher efficiency in large-scale, high-dimensional parameter spaces, for the scope of this study, grid search achieves a good balance between efficiency and effectiveness, while enhancing the interpretability and reproducibility of the results.

3. Results and Discussions

To validate the methodology's effectiveness proposed in this paper, a systematic evaluation of the method is carried out based on real grid frequency monitoring data. The dataset used was selected from the publicly available monitoring data of the UK National Grid System Operator, covering the grid frequency for September 2023, with a sampling interval of 1 s, and containing a total of 2,592,030 observations. The system is rated at 50 Hz with a permissible frequency fluctuation interval of ±0.5 Hz [22]. Natural gas has the highest share, 33%, in this power system. Wind energy is the next highest with 24.9%, and solar energy has a share of 5.8% [23].

3.1. Performance Metrics

To comprehensively evaluate the prediction performance of the proposed method, root mean squared error (RMSE), coefficient of determination (R^2), mean absolute percentage error (MAPE), and mean absolute error (MAE) were selected as the evaluation indices. These metrics are widely used in machine learning to measure the accuracy of regression models. The specific calculation formula is expressed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{19}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{20}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (21)$$

where N denotes the total number of samples, y_i denotes the actual value of the i-th sample, $\hat{y}^i$ denotes the predicted value of the i-th sample, and $\overline{y}$ denotes the mean of the actual values of all samples.

3.2. Data Analysis

To fully explore the effective information in the data, this paper extends the multi-dimensional features of the original grid frequency time series based on the phase space reconstruction theory and the date–time feature extraction method to construct a multidimensional feature matrix. This process can characterize the potential information in the frequency series more comprehensively, providing a solid data foundation for constructing a high-precision prediction model.

3.2.1. Determination of Phase Space Parameters

In this paper, the average mutual information method is used to plot the curve of mutual information between the original grid frequency sequence and its delayed sequence with delay time (see Figure 3). The horizontal axis is the delay time, τ, and the vertical axis is the mutual information value, reflecting the correlation degree between the original sequence and the delayed τ sequence. When determining the optimal delay time, the first local minimum of the mutual information curve is usually selected. At this time, the data redundancy can be effectively reduced, and the dynamic information of the sequence is maintained. If there is no apparent local minimum in the curve, the position where the mutual information value tends to be stable is selected as the optimal delay time.

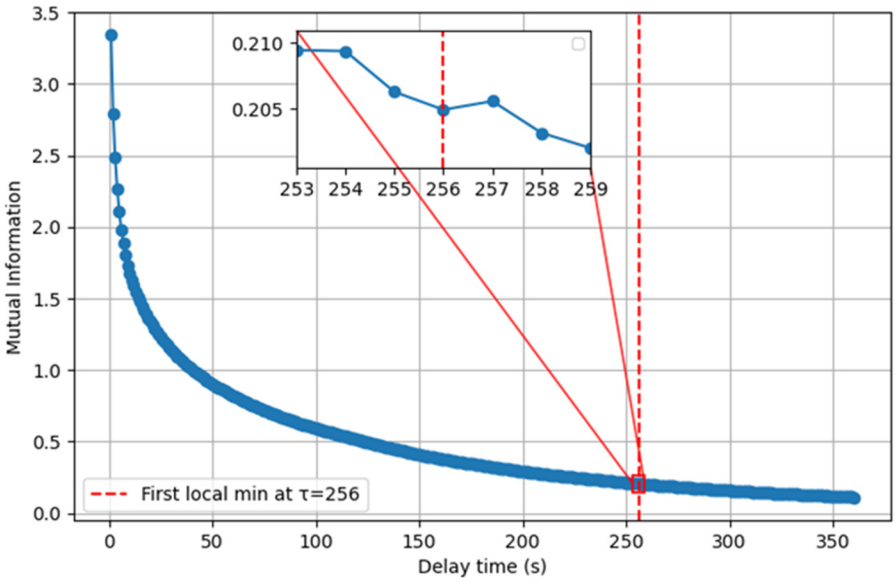


Figure 3. Mutual information curve of grid frequency and its delayed sequence.

From Figure 3, it can be seen that with the increase in delay time, τ, the mutual information value between the original sequence and the delayed sequence decreases rapidly, which indicates that the information redundancy between the data points is gradually reduced and the correlation is weakened. When τ is small, the mutual information value is high, reflecting a high degree of correlation between the sequences, which easily leads to excessive redundant information in phase space reconstruction, and is not conducive to the presentation of the system's dynamic features. As τ increases, the mutual information value gradually decreases, and the information independence increases, which helps to

enrich the phase space structure. The red dashed line marked at $\tau = 256$ in the figure is the first local minimum of the mutual information curve, so it is selected as the optimal delay time and verified by the local zoom-in diagram. The suitable delay time reduces the redundant information in the reconstruction and preserves the system's dynamics. Based on the phase space reconstruction theory and mutual information analysis, $\tau = 256$ is finally determined to be the optimal delay time for the subsequent dynamics analysis and predictive modeling.

After determining the delay time, this paper uses the Cao method to determine the optimal embedding dimension of the grid frequency time series. Figure 4 demonstrates the curves of the two key metrics, $E1$ and $E2$, in the Cao method with the embedding dimension.

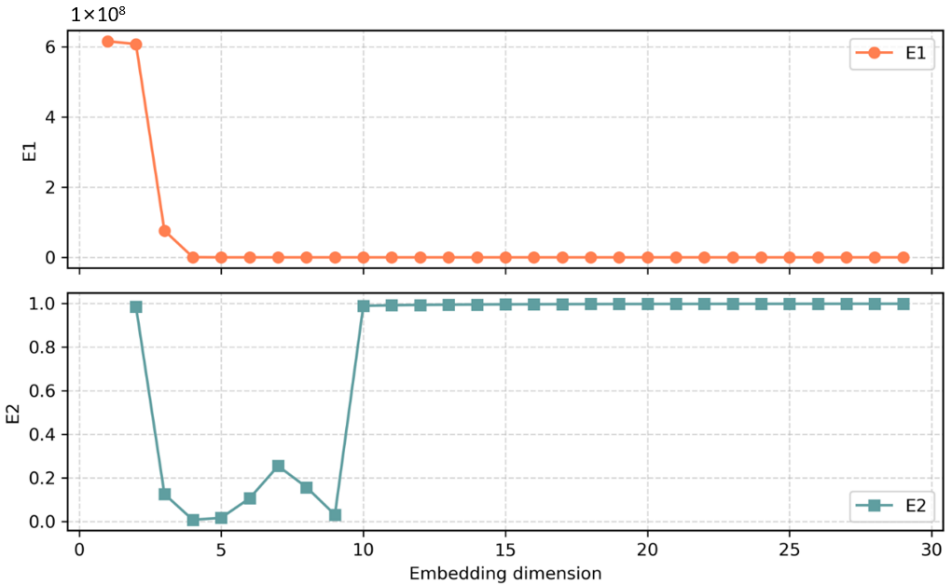


Figure 4. $E1$ and $E2$ vs. embedding dimension.

The upper half of Figure 4 shows that $E1$ increases with the embedding dimension m and stabilizes after $m = 4$, indicating that the system's degrees of freedom have been sufficiently inscribed, and the effect of continuing to increase the embedding dimension is limited. The second half shows the curve of $E2$ with m. $E2$ approaches 1 when $m \geq 10$, further verifying the dynamics of the sequence.

Combining the optimal delay time determined by the mutual information method and the analytical results of Cao's method, this paper finally determines the embedding dimension of the grid frequency time series as $m = 4$.

3.2.2. Extraction of Date and Time Information

In the preprocessing stage of grid time series data, reasonable extraction and encoding of temporal features are crucial to improve the prediction model's performance. Based on the one-month grid frequency data collected above, this paper extracts the following temporal features: week of month, day of month, day of week, is_weekend, hour, minute, and second. Among them, hour, minute, second, day of week, and day of month belong to discrete variables with strong periodicity, and sine and cosine coding are used to reveal their cyclic patterns better and avoid model misjudgement; is_weekend is a binary feature, which is directly coded with 0/1 to distinguish between weekdays and weekends; week of month is_weekend is a binary feature, which is directly coded with 0/1 to distinguish weekdays and weekends; week of month is coded with integers from 1 to 5, which is convenient for the model to capture the cyclical changes within the month. Through diversified temporal features and appropriate coding methods, the temporal patterns in

the grid frequency series can be more comprehensively explored, effectively improving the accuracy of model analysis and prediction.

In this paper, the correlation between the features extracted based on phase space reconstruction and date–time information and the grid frequency is further analyzed and demonstrated by a scatter plot in Figure 5. Here, the horizontal axis represents the absolute value of Spearman's correlation coefficient ($|r|$) between each feature and the grid frequency, and the vertical coordinate is its corresponding significance level, which is quantified by $-\log10$ (p value) to reveal the degree of correlation and statistical significance of each feature with the target variable.

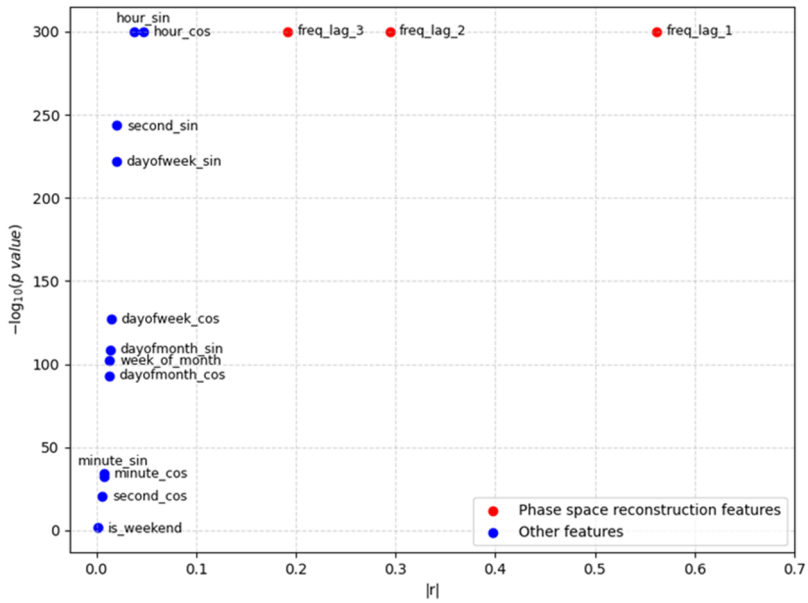


Figure 5. Feature correlation and significance scatter plot.

In Figure 5, the red scatters represent the features extracted based on phase space reconstruction, while the blue scatters indicate the features obtained from date–time information. In time series forecasting, Spearman correlation coefficient ($|r|$) is commonly used for feature screening. However, relying only on the correlation coefficient may ignore some statistically significant features despite the low correlation coefficient. Therefore, this paper also introduces the significance test (p-value) as an auxiliary criterion in feature screening.

In this paper, we take the lowest correlation coefficient in the red scatter as the threshold value of $-\log10$ (p value) and filter the blue scatter to keep only those date and time features whose $-\log10$ (p value) is not less than this threshold value. This method finally filters "hour" as a valid time feature. This double-criteria feature screening avoids information omission, improves the features' predictive ability and statistical reliability, and helps improve the model's performance in practical applications.

To summarize, the final feature set selected in this paper contains the following: phase space reconstruction features with a delay time of 256 s and an embedding dimension of 4, as well as the "hour" feature in the date–time information.

3.3. Parameter Settings for DSCW Method

Based on the subset of features screened in the previous section, this paper employs a grid search optimization algorithm to traverse the parameter space. It performs model training and validation evaluation for all combinations of sliding window sizes, step ratios, and attenuation factor λ parameters. The parameter combination that minimizes

the root mean square error (RMSE) is finally selected to determine the optimal values of each parameter.

Figures 6 and 7 show the trend and distribution of the model RMSE under different combinations of sliding window length and step ratio, respectively. From the figures, it can be intuitively observed that different parameter combinations significantly affect the model performance.

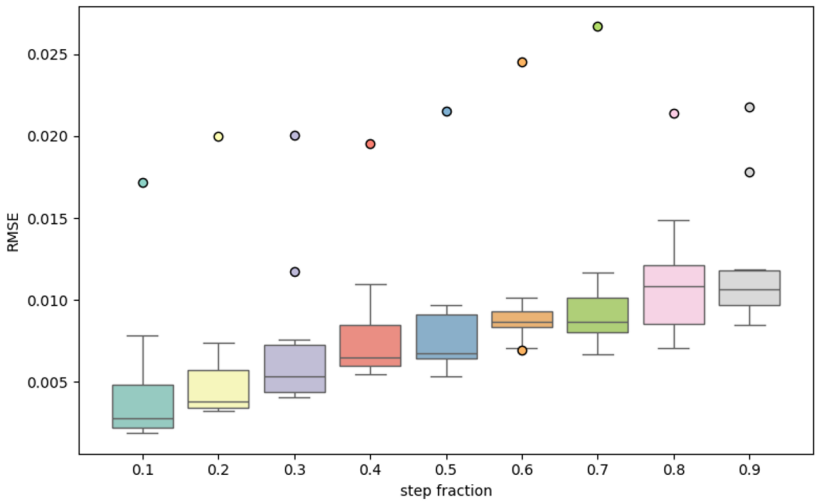


Figure 6. Model prediction performance under different step ratios.

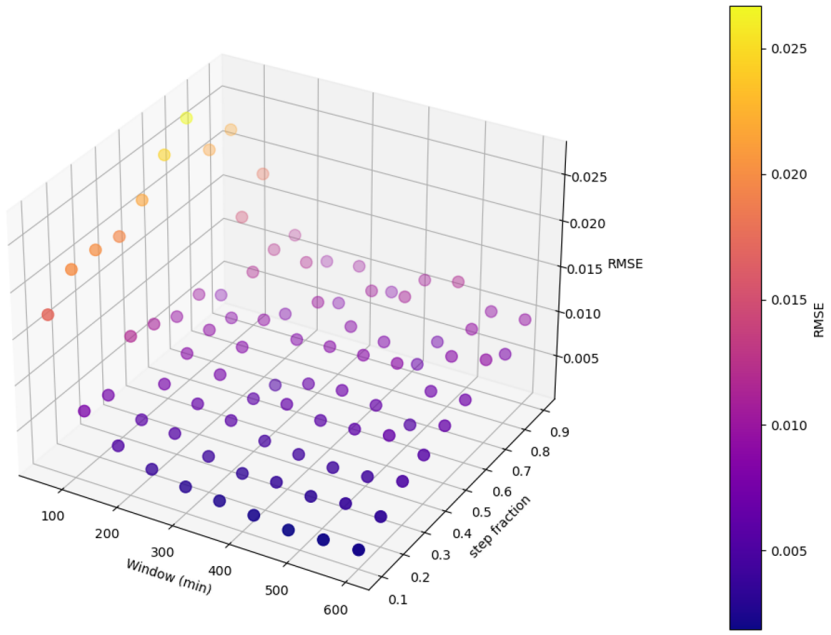


Figure 7. Effect of window and step ratio on prediction performance.

According to the results in Figures 6 and 7, the effect of the step ratio on the model RMSE is more significant. The box plots in Figure 6 show that the RMSE shows an overall upward trend with the increase in the step size ratio, indicating that a smaller step size ratio helps to improve the model prediction accuracy. When the step size ratio is larger, the prediction error and volatility of the model increase significantly, and the outliers of some distributions tend to increase, indicating that a step size ratio that is too large is not conducive to model stability. Combined with the three-dimensional scatter plot analysis in Figure 7, the combination of different sliding window lengths and step size ratios has an obvious relationship with the effect of RMSE. Overall, the smaller step ratio and

longer sliding window length exhibit lower RMSE values, indicating that the simultaneous optimization of these two hyperparameters can effectively improve the model performance.

Based on the results in Figures 6 and 7, the step size ratio is finally determined to be 0.1. Subsequently, this paper further filters the optimal sliding window length and λ value by comparing the RMSE performances of the model with different sliding window lengths and values of the decay factor λ based on a step size ratio of 0.1. Figure 8 demonstrates the effect of λ values on the RMSE of the model under different sliding windows.

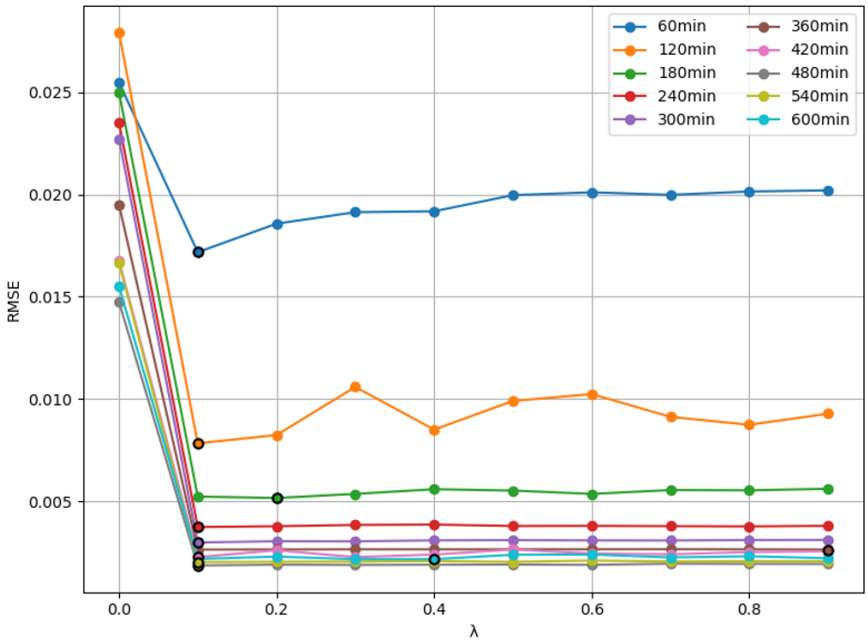


Figure 8. Effect of λ on prediction performance at different window lengths.

As can be seen from the figure, the RMSE of the model decreases significantly when λ is gradually increased to 0.1. Subsequently, it tends to stabilize under most of the sliding window lengths. This phenomenon suggests that the appropriate fusion of historical feature weights helps improve the model's generalization ability. Further analysis shows that, in the case of shorter sliding windows (e.g., 60 min, 120 min), the RMSE value of the model is more sensitive to changes in λ. In the case of longer sliding windows, the value of RMSE is at a lower level and relatively less affected by the change in λ. Some of the data points in the figure are labeled with circles with black outlines, representing the best λ parameter points under different sliding window lengths. Overall, a reasonable setting of the λ value can effectively reduce the prediction error, and there is a certain degree of difference in the value of λ to optimize the model performance under the conditions of different lengths of the sliding window.

Combined with the above analysis results, it can be seen that the sliding window size, step ratio, and attenuation factor, λ, all significantly affect the prediction performance of the model. When the sliding window is set to 480 min, the step ratio is 0.1, and the attenuation factor λ is 0.1, the model achieves the lowest RMSE and the best prediction performance on the validation set. Therefore, in this paper, the sliding window of 480 min, step ratio of 0.1, and λ value of 0.1 are finally selected as the optimal parameters of the model to improve the accuracy and stability of time series prediction.

3.4. LightGBM Hyperparameter Selection

Figure 9 illustrates the effect of the four core hyperparameters on the model's RMSE on the validation set, and Table 1 lists the optimized LightGBM model hyperparameter configurations.

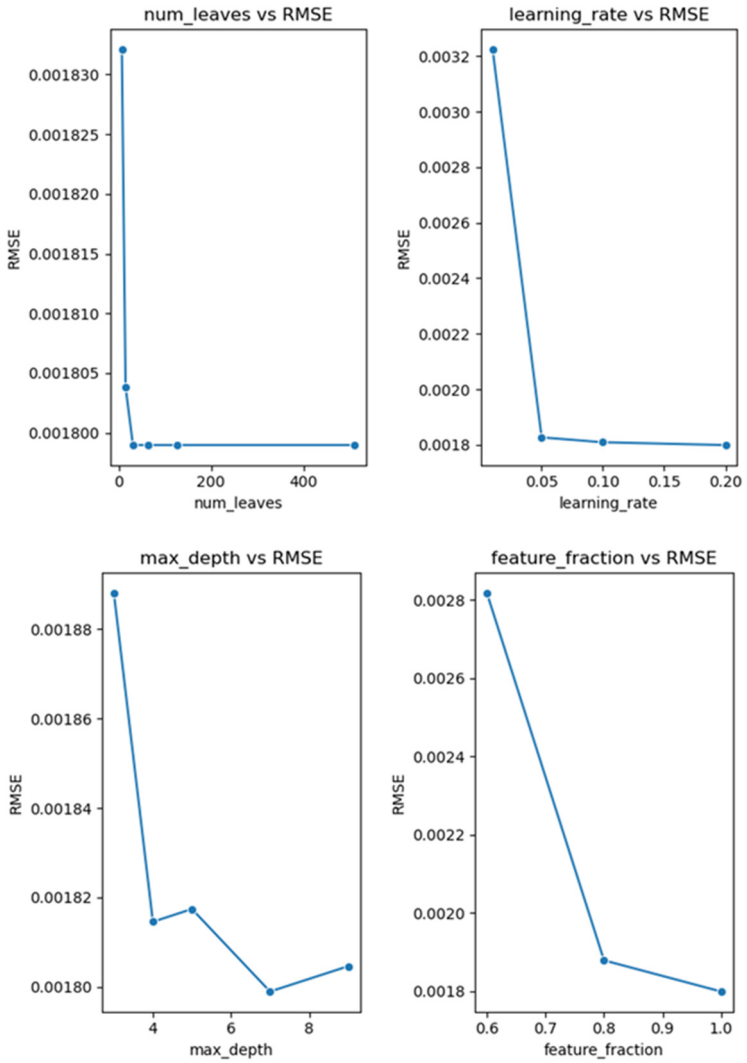


Figure 9. Impact of LightGBM hyperparameters on model performance.

Table 1. Hyperparameter settings for the LightGBM model.

Hyperparameter	Value	Hyperparameter	Value
Num_leaves	31	Max_depth	7.0
Learning_rate	0.2	Feature_fraction	1.0

From Figure 9, it can be seen that with the gradual increase in num_leaves, learning_rate, and feature_fraction, the RMSE of the model shows an obvious decreasing trend. This indicates that appropriately increasing the number of leaves, improving the learning rate, and increasing the feature sampling ratio help improve the LightGBM model's performance. Max_depth and RMSE show some fluctuation between them. This indicates that the model depth has a nonlinear effect on the prediction accuracy, and it is not that the larger the model depth, the higher the prediction accuracy, but there is an optimal depth range. Overall, selecting larger num_leaves, moderate learning_rate, reasonable max_depth, and higher feature_fraction helps reduce the model error and improve the generalization ability.

3.5. Prediction Results of the DSCW-LightGBM Method

In this paper, a systematic comparative analysis is conducted with the proposed DSCW method as the core, and its performance is compared horizontally with four benchmark

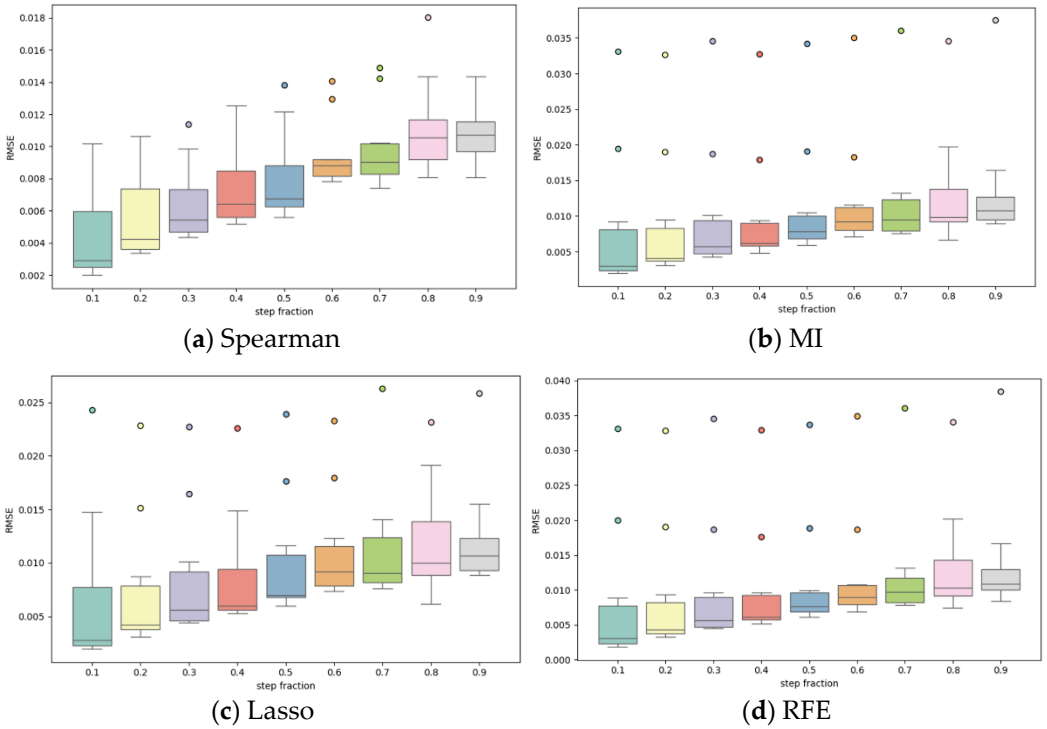
feature selection methods, including the following: a traditional feature screening method based on correlation coefficient analysis, a feature importance assessment method based on mutual information, a regularized feature selection method based on Lasso regression, and a method based on recursive feature elimination (RFE). Table 2 lists the number of selected features under each method, which provides data support for the subsequent performance evaluation.

Table 2. Feature numbers for different feature selection methods.

Method	Number of Features	Method	Number of Features
SCW	5	Lasso	8
Spearman	3	RFE	15
MI	14	/	/

According to the statistical results in Table 2, the SCW method retains only five core features, showing its advantages in feature redundancy control and effective information extraction. In contrast, the Spearman and Lasso methods screened 3 and 8 features, respectively, the mutual information (MI) method retained more features with a total of 14 features due to its focus on nonlinear dependency mining, and the number of features was significantly higher than that of the previous methods, while the RFE retained all 15 features. This table visualizes the difference in the number of features between different feature selection methods.

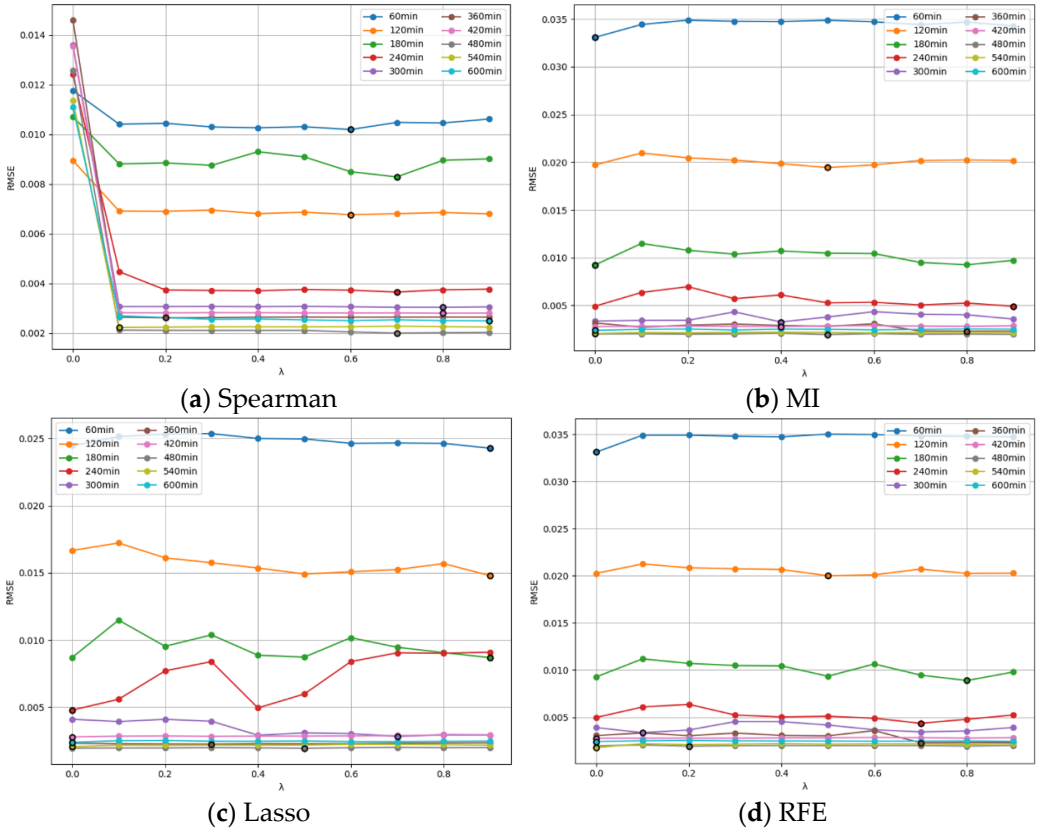
In this paper, the control variable method investigates the effects of the parameter configurations of sliding window size, step ratio coefficient, and attenuation factor, λ, on the model performance under different feature selection strategies. Figures 10 and 11 show the distribution patterns of the values of the above three key parameters and their effects on the model performance in each feature selection method, respectively.



(**a**) Spearman (**b**) MI

(**c**) Lasso (**d**) RFE

Figure 10. Model prediction performance under different step ratios with various feature selection methods.



(**a**) Spearman (**b**) MI

(**c**) Lasso (**d**) RFE

Figure 11. Effect of λ parameter on model performance with different sliding window lengths under various feature selection methods.

According to the experimental results in Figures 10 and 11, Table 3 gives the optimal values of each parameter under different feature selection strategies. In addition, the hyperparameter optimization results based on the LightGBM model under the corresponding parameter combinations are also listed in the table.

Table 3. Parameter settings for different feature selection methods.

Method	W (min)	Δ*t*	λ	Num_Leaves	Max_Depth	learning_Rate	Feature_Fraction
Spearman	480	0.1	0.7	31	9.0	0.2	1.0
MI	480	0.1	0.5	31	5.0	0.2	1.0
Lasso	480	0.1	0.5	15	4.0	0.2	1.0
RFE	480	0.1	0.2	31	9.0	0.2	1.0

According to the experimental parameter settings in Tables 1 and 3, it can be seen that the sliding window length parameter and step size scaling factor are 480 min and 0.1 for different feature selection methods; also, the feature sampling rate and learning rate of the LightGBM algorithm are set to 1.0 and 0.2, respectively. In terms of the maximum depth parameter selection, both the Spearman and RFE methods use a larger maximum depth (9), while the Lasso method cooperates with a lower maximum depth (4) and fewer number of leaves (15), which enhances the adaptability to high-dimensional features. In contrast, the SCW method proposed in this paper employs a maximum depth value of 7 at a more eclectic level.

Based on the parameter configuration information under different feature selection strategies given in Tables 1 and 3, the LightGBM model prediction results are analyzed.

Table 4 lists the main performance metrics of the LightGBM model on the test set and the average computation time of the sliding window weights under different feature selection methods. Figure 12 then visually compares the performance of different feature selection methods on multidimensional evaluation metrics through radar charts.

Table 4. Regression performance and efficiency of different feature selection methods.

Method	RMSE↓	R^2↑	MAE↓	MAPE/%↓	Computation Time (ms)
SCW	1.799×10^{-3}	0.9924	1.170×10^{-3}	2.339×10^{-3}	16.57
Spearman	1.984×10^{-3}	0.9907	1.244×10^{-3}	2.488×10^{-3}	11.13
MI	2.002×10^{-3}	0.9905	1.261×10^{-3}	2.522×10^{-3}	32.82
Lasso	1.940×10^{-3}	0.9911	1.228×10^{-3}	2.456×10^{-3}	22.38
RFE	2.008×10^{-3}	0.9905	1.265×10^{-3}	2.530×10^{-3}	34.70

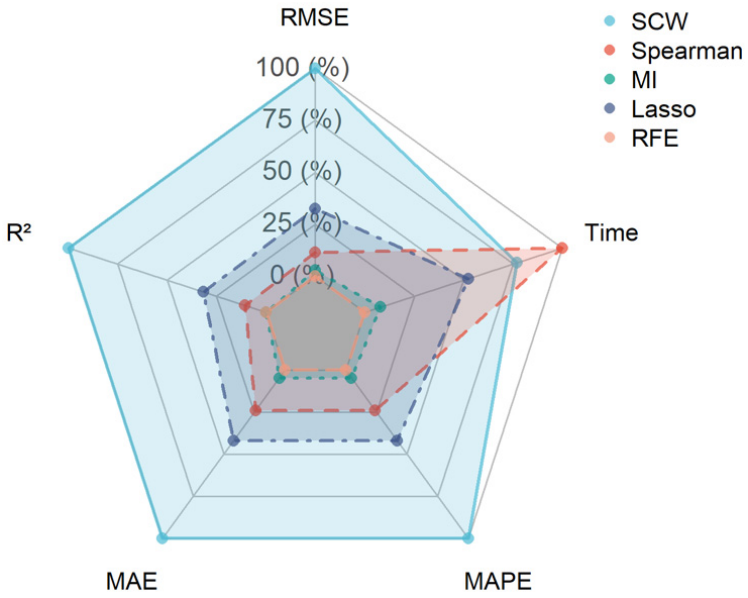


Figure 12. Regression performance of different feature selection methods.

According to the quantitative evaluation results in Table 4, the SCW method proposed in this paper shows a significant advantage in the regression task, with a low RMSE value of 1.799×10^{-3}, an R^2 as high as 0.9924, and better indicators such as MAE and MAPE than the other methods. In this table, the downward arrows indicate that a smaller value means better regression performance for that metric, while the upward arrow for R^2 means that a larger value is preferable. While the Spearman and Lasso methods show good fitting capabilities, their overall regression performance is slightly inferior to SCW. The MI and RFE methods have increased model complexity due to higher feature dimensions, and the weight calculation time consumed is 32.82 ms and 34.70 ms, respectively. Combined with the radar charts in Figure 12, the SCW method has a balanced and outstanding performance in the core indexes such as R^2 and RMSE; the Spearman method has an advantage in calculation time consumed, but the performance of the other indexes is average; the MI and RFE have minimal differences in error metrics such as MAE and MAPE, and their regression accuracy is comparable. Overall, the SCW method not only improves model accuracy but also enhances computational efficiency by effectively controlling the feature dimensions, which verifies its feasibility and value in engineering applications.

Figure 13 further compares the time series fitting curves of the LightGBM model optimized based on five feature selection methods: SCW, Spearman, MI, Lasso, and RFE. The original sequences here are within the test sample interval, allowing us to analyze the dynamic consistency between the prediction results of each method and the actual values.

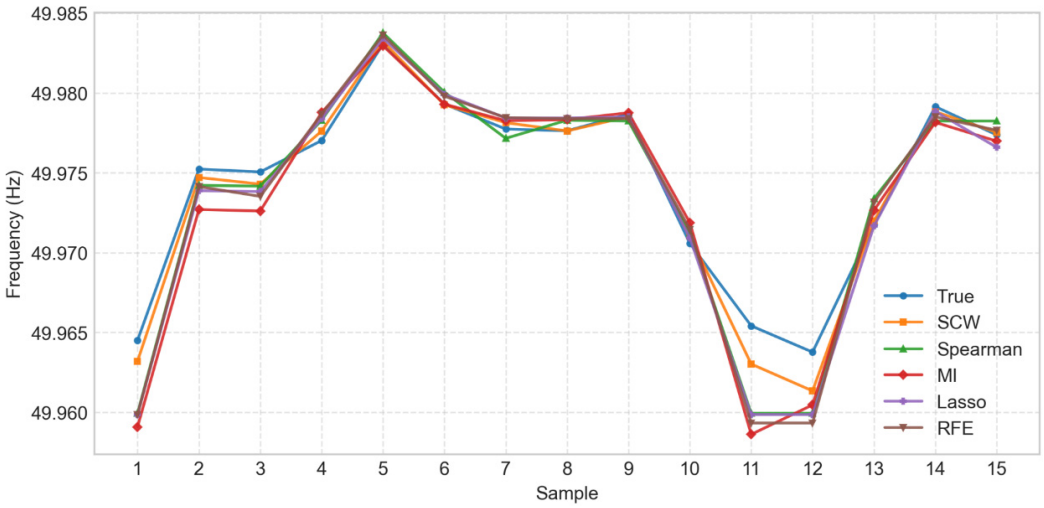


Figure 13. Grid frequency prediction comparison using different feature selection methods.

From the figure, it can be seen that feature selection methods such as SCW, Spearman, MI, Lasso, and RFE can more accurately fit the trend of the target series in terms of the overall trend and fluctuation interval. The SCW method is closer to the true value (True) in most of the sample points, which shows a strong dynamic adaptation ability.

In order to further evaluate the impact of different feature weight allocation strategies on the LightGBM model in grid frequency timing prediction. In this paper, three feature input methods are designed: unweighted input (LightGBM), static weight assignment (SCW-LightGBM), and dynamic weight assignment (D-SCW-LightGBM). All three methods are compared and experimented upon with the same dataset and training process. Table 5 presents the results of the three methods on the key regression metrics (RMSE, MAE, MAPE). Figure 14 then presents the fitting effect of the predicted sequences of each method to the real sequences on the validation set.

Table 5. LightGBM performance with different feature weighting methods.

Model	RMSE↓	MAE↓	MAPE/%↓
D-SCW-LightGBM	1.799×10^{-3}	1.170×10^{-3}	2.339×10^{-3}
SCW-LightGBM	1.810×10^{-3}	1.177×10^{-3}	2.354×10^{-3}
LightGBM	2.088×10^{-3}	1.273×10^{-3}	2.546×10^{-3}

As can be seen from Table 5, all three methods exhibit low error levels in the regression task, with D-SCW-LightGBM slightly outperforming the other two methods in terms of RMSE, MAE, and MAPE. The downward arrows in the table indicate that lower values correspond to better model performance for these metrics. This indicates that dynamically adjusting the feature weights can improve the model's prediction accuracy. Specifically, D-SCW-LightGBM utilizes the dynamic weighting mechanism to dynamically adjust the feature weights according to the correlation between the features within the sequence and the target, which makes the model more flexible in responding to changes in the sequence information, and thus has the best prediction performance, with an RMSE of

1.799×10^{-3}, which outperforms both the static-weighted SCW-LightGBM (1.810×10^{-3}) and the unweighted LightGBM model (2.088×10^{-3}).

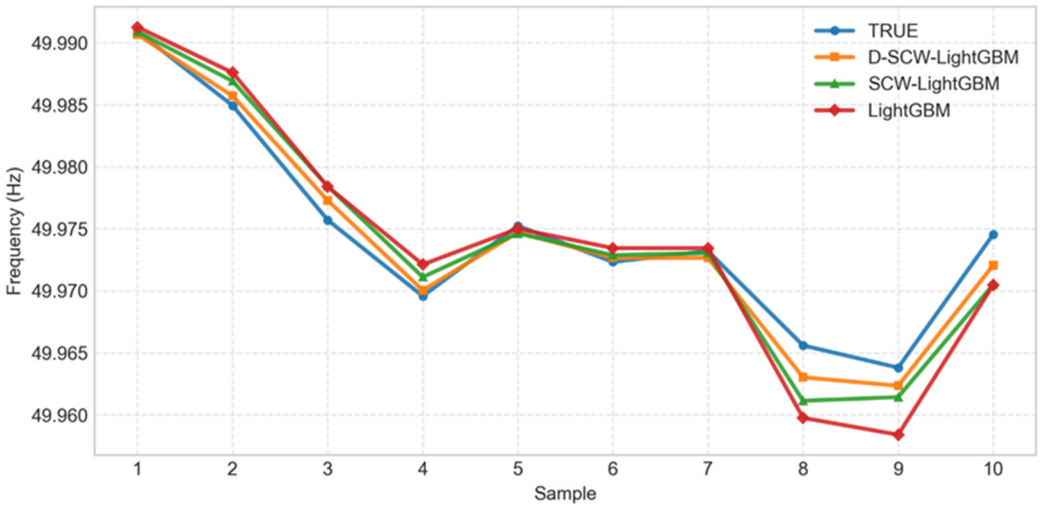


Figure 14. Grid frequency prediction comparison using different weighting methods.

Based on the SHAP value analysis presented in Table 6, the D-SCW method consistently demonstrates an advantage in enhancing the importance of key lagged frequency features compared to both the static SCW weighting strategy and the unweighted LightGBM baseline model. Notably, the SHAP values of core lagged frequency features (such as Freq_lag_1) are significantly higher than those of other features, which aligns well with the theoretical assumption in power systems that "recent historical states primarily govern frequency dynamic responses." This empirical evidence substantiates the model's ability to accurately capture and represent the underlying physical mechanisms.

Table 6. Feature importance comparison by mean absolute SHAP values under different weighting methods.

Feature	D-SCW-LightGBM	SCW-LightGBM	LightGBM
Freq_lag_1	1.797×10^{-2}	1.791×10^{-2}	1.794×10^{-2}
Freq_lag_2	4.08×10^{-4}	2.5×10^{-5}	3.5×10^{-5}
Freq_lag_3	1.2×10^{-4}	6.3×10^{-5}	3.6×10^{-5}
Hour_cos	2.11×10^{-4}	2.08×10^{-4}	1.86×10^{-4}
Hour_sin	7.1×10^{-5}	2.8×10^{-5}	2.7×10^{-5}

Figure 14 visualizes the effect of the sequence fitting of the three feature weighting methods through the time series curves. It can be seen that the model as a whole can fit the fluctuation trend of the target variable better, regardless of dynamic weighting, static weighting, or no weighting configuration. However, the LightGBM model with no weighting configuration of the input features deviates more significantly from the actual value at some inflection points and where the fluctuations are large. In contrast, SCW-LightGBM improves the stability of the fit by static weighting, while D-SCW-LightGBM enhances the responsiveness to sudden changes by the dynamic weighting mechanism, and its fitted curves are closer to the actual values at the turning points of the trend. In summary, the feature weighting method, especially the introduction of the dynamic adaptation mechanism, significantly improves the accuracy and adaptability of the LightGBM model in grid frequency time series prediction, especially when dealing with non-stationary time series data.

To further validate the effectiveness of the dynamic significance–correlation-based weighting method proposed in this paper in grid frequency prediction, the real-time time series dataset of grid frequency in February 2025 is used as a test sample to validate the LightGBM model. Table 7 presents the proposed method's regression performance and computational efficiency compared with four feature selection strategies: Spearman, MI, Lasso, and RFE. By quantitatively assessing each method's prediction accuracy and the average computation time for sliding window weights, the actual effectiveness of each method in the prediction task can be objectively reflected. Figure 15 visually illustrates the differences in various metrics among the methods using a multiYaxis bar chart, providing a clear and robust basis for comprehensively evaluating the proposed method.

Table 7. Regression performance and efficiency of different feature selection methods.

Method	RMSE↓	R^2↑	MAE↓	MAPE/%↓	Computation Time (ms)
SCW	9.003×10^{-4}	0.9948	7.368×10^{-4}	1.473×10^{-3}	17.88
Spearman	9.347×10^{-4}	0.9942	7.450×10^{-4}	1.490×10^{-3}	11.22
MI	9.155×10^{-4}	0.9945	7.372×10^{-4}	1.474×10^{-3}	31.77
Lasso	9.496×10^{-4}	0.9940	7.627×10^{-4}	1.525×10^{-3}	33.11
RFE	9.233×10^{-4}	0.9944	7.442×10^{-4}	1.488×10^{-3}	34.60

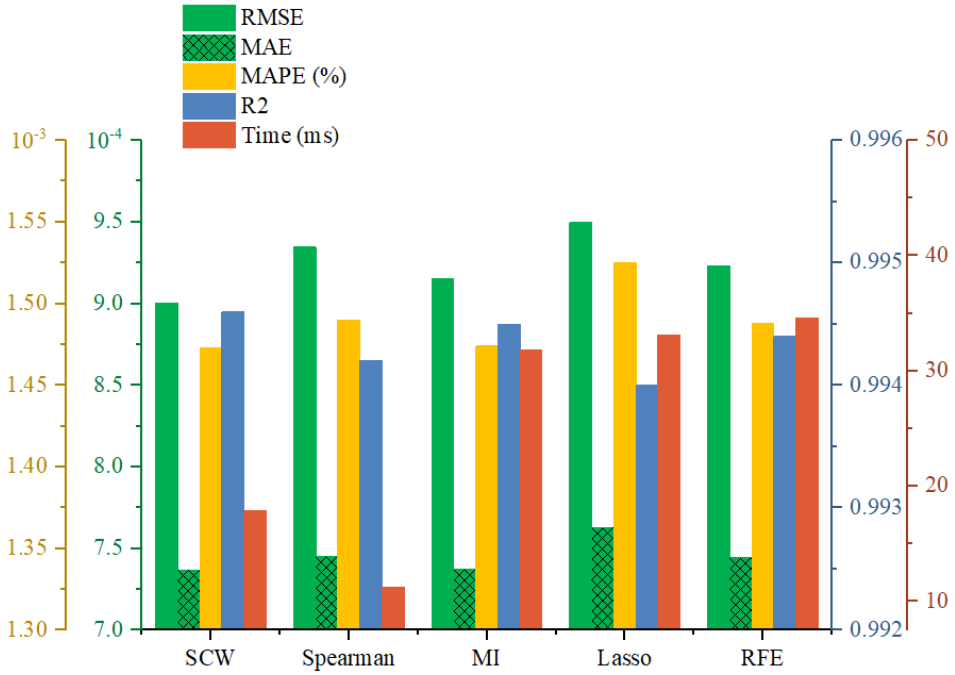


Figure 15. Performance comparison of different feature selection methods.

According to the results in Table 7, the differences between the feature selection methods on the RMSE and MAE metrics are small, and the overall error is low. Among them, the SCW method performs optimally on these two error indicators. The performance of the methods on the MAPE indicator is relatively close, and the R^2 values of all methods are higher than 0.994, indicating that the model has good interpretability. However, there are obvious differences in the efficiency of feature weight computation: the Spearman method has the shortest weight computation time of 11.22 ms, while the RFE and Lasso methods have relatively longer weight computation times of more than 30 ms. Figure 15 further verifies the above conclusions through the visualization comparison. It can be seen that the SCW method outperforms the other methods in terms of prediction accuracy

(RMSE, MAE) and the degree of model fit (R^2), and although the weight calculation time is slightly higher than that of the Spearman method, it is much lower than that of the algorithms such as MI and RFE. In addition, the Spearman method makes its MAPE slightly higher than that of SCW due to the screening of features based on correlation only, further validating the importance of statistical significance testing. In summary, the SCW method balances prediction accuracy and computational efficiency in the grid frequency prediction task by integrating correlation analysis and a statistical significance test.

3.6. Robustness Analysis

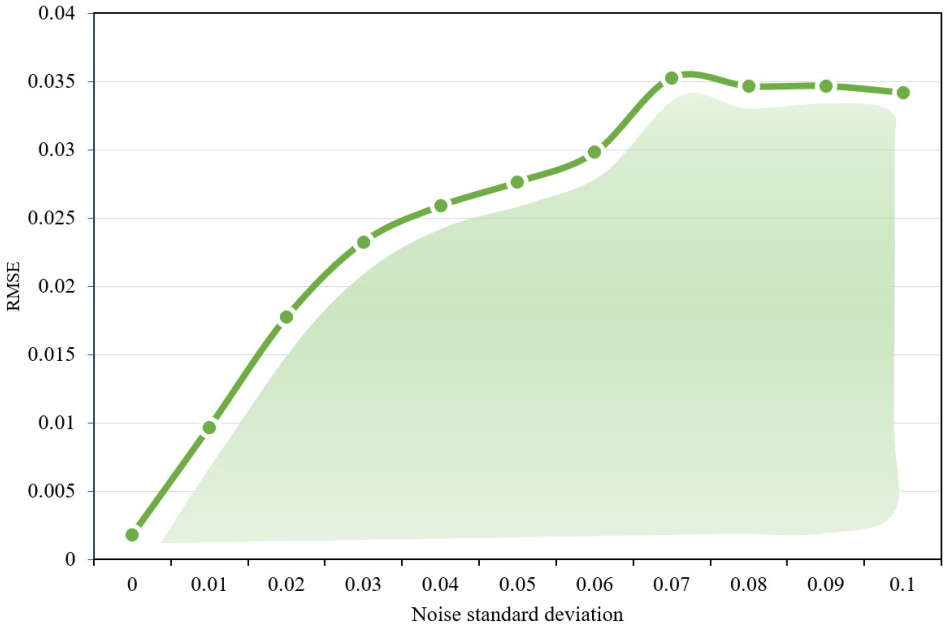
To systematically evaluate the proposed method's resistance to input perturbations, this study conducted noise sensitivity tests on the two aforementioned datasets. By incrementally adding Gaussian noise with different standard deviations to the validation set, the RMSE of the model's predictions was recorded under each scenario. The noise–error relationship curves are shown in Figure 16, and key statistical indicators are summarized in Table 8.

Table 8. RMSE values of the model under different noise levels across different datasets.

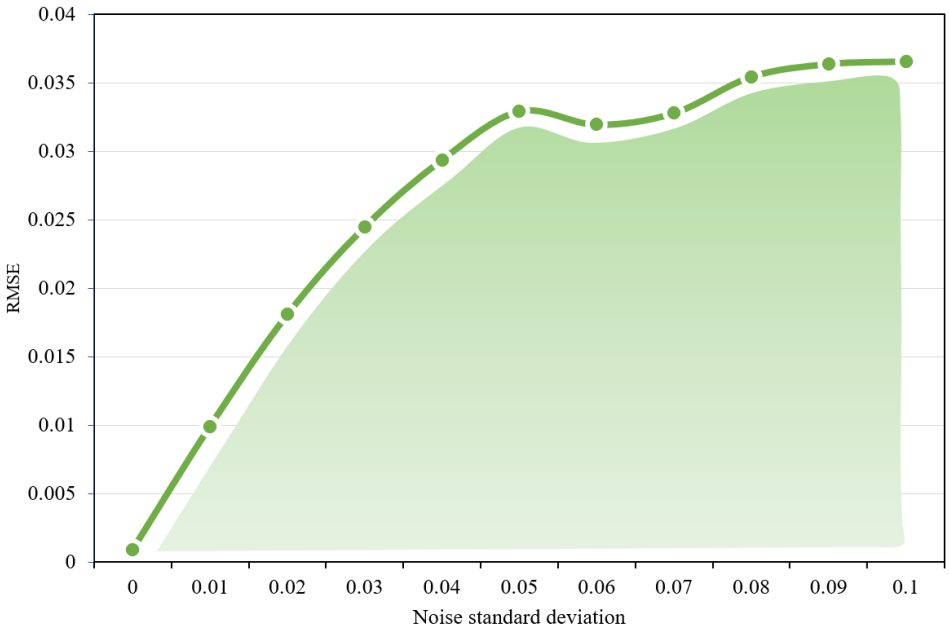
Noise Standard Deviation	Datasets	
	2025-2	**2023-9**
0	9.00×10^{-4}	1.80×10^{-3}
0.01	9.91×10^{-3}	9.68×10^{-3}
0.02	1.81×10^{-2}	1.77×10^{-2}
0.03	2.45×10^{-2}	2.33×10^{-2}
0.04	2.94×10^{-2}	2.59×10^{-2}
0.05	3.29×10^{-2}	2.76×10^{-2}
0.06	3.19×10^{-2}	2.98×10^{-2}
0.07	3.28×10^{-2}	3.53×10^{-2}
0.08	3.55×10^{-2}	3.47×10^{-2}
0.09	3.64×10^{-2}	3.47×10^{-2}
0.1	3.66×10^{-2}	3.42×10^{-2}

As shown in Figure 16, as the input noise standard deviation gradually increases from 0 to 0.1, the RMSE values for both datasets exhibit a steady upward trend. This indicates that the prediction error increases as expected with higher levels of noise. However, throughout the tested noise range, the RMSE remains at a relatively low level, and no abrupt changes in performance due to noise accumulation are observed. Table 8 further quantitatively compares the RMSE at each noise level, revealing similar RMSE growth patterns across both datasets, with the increases being gradual and controllable. These findings demonstrate that the model maintains stable responsiveness to noise disturbances.

The experimental results above demonstrate that the proposed model achieves satisfactory prediction accuracy under typical noise scenarios. Specifically, when the noise standard deviation reaches 0.1 (corresponding to a 10% fluctuation in the features), the increase in RMSE is still minimal. This indicates that the model exhibits strong robustness on both datasets, confirming its applicability and reliability in real-world scenarios where inevitable noise or slight disturbances may be present in the input data.



(**a**) September 2023 dataset.



(**b**) February 2025 dataset.

Figure 16. Model robustness analysis curves under noise perturbation on different datasets.

4. Conclusions

This paper proposes a prediction framework based on the fusion of D-SCW and LightGBM, which mainly focuses on the prediction study of power system frequency time series data. By constructing a dual feature screening mechanism of relevance and significance, the method not only realizes the effective expansion and precise selection of feature dimensions, but also significantly enhances the computational efficiency and prediction performance of the prediction model in power system data scenarios, which improves the accuracy and real-time performance of the grid frequency timing prediction. Based on the research presented, the following conclusions are drawn:

(1) The D-SCW-LightGBM fusion framework is an efficient method for power system timing prediction modeling. The method is based on LightGBM's efficient histogram algorithm and GOSS technology, which improves the processing efficiency of large-scale

feature engineering data; meanwhile, the SCW method optimizes the feature space through the dual screening mechanism of correlation and statistical significance, and enables the model to better adapt to time-varying fluctuations in grid frequency through a dynamic feature weight adjustment strategy. The fusion model retains the advantages of LightGBM in handling high-dimensional data and improves the accuracy of feature selection.

(2) The D-SCW method combines the Spearman rank correlation coefficient and statistical significance test to realize the dual assessment of feature space and introduce the dynamic weighting mechanism for feature adjustment. Different from the single correlation method, the D-SCW method comprehensively examines the correlation and statistical significance in feature screening, especially when the correlation analysis is insufficient, and the significance test enhances the scientificity of feature screening. The experimental results show that this two-dimensional evaluation mechanism enhances the interpretability of feature selection and improves the model's ability to capture the time-varying characteristics of grid frequency, thus optimizing the prediction performance.

(3) This paper systematically compares the SCW method with commonly used feature selection methods such as Spearman, MI, Lasso, and RFE. The experimental evaluation based on real grid frequency monitoring data shows that the SCW method significantly improves prediction accuracy while guaranteeing computational efficiency, and demonstrates superiority in terms of error control and time-consumption compared to computationally complex methods such as RFE and Lasso. This indicates that the feature selection strategy based on the correlation–significance joint testing can achieve a balance between prediction performance and computational efficiency in real-time scenarios such as grid frequency dynamic prediction, providing a new modeling idea.

In summary, the D-SCW-LightGBM fusion framework proposed in this paper demonstrates high prediction accuracy and model stability in grid frequency prediction. The framework fully exploits LightGBM's parallel computing capability in handling large-scale data and effectively utilizes the two-dimensional optimization of feature space by the SCW method, which significantly improves the predictive capability and interpretability of the model. This technology effectively supports improving the operation security and intelligence level of the power system and has good application prospects.

Author Contributions: Conceptualization, J.Z. (Jie Zhou); methodology, J.Z. (Jie Zhou) and S.B.; software, J.Z. (Jie Zhou); validation, J.Z. (Jing Zhou); formal analysis, J.Z. (Jie Zhou) and J.Z. (Jing Zhou); investigation, J.Z. (Jie Zhou); resources, X.T.; data curation, S.B.; writing—original draft preparation, J.Z. (Jie Zhou); visualization, J.Z. (Jie Zhou); supervision, X.T.; project administration, X.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cao, Z.N.; Wang, J.Z.; Xia, Y.R. Combined electricity load-forecasting system based on weighted fuzzy time series and deep neural networks. *Eng. Appl. Artif. Intell.* **2024**, *132*, 108375. [CrossRef]
2. Saleem, M.I.; Saha, S. Assessment of frequency stability and required inertial support for power grids with high penetration of renewable energy sources. *Electr. Power Syst. Res.* **2024**, *229*, 110184. [CrossRef]
3. Qin, B.Y.; Wang, M.J.; Zhang, G.; Zhang, Z. Impact of renewable energy penetration rate on power system frequency stability. *Energy Rep.* **2022**, *8*, 997–1003. [CrossRef]
4. Gong, H.H.; Li, Y.Y.; Zhang, J.N.; Zhang, B.S.; Wang, X.L. A new filter feature selection algorithm for classification task by ensembling pearson correlation coefficient and mutual information. *Eng. Appl. Artif. Intell.* **2024**, *131*, 107865. [CrossRef]

5. Gu, X.Y.; Guo, J.C.; Xiao, L.J.; Li, C.Y. Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy. *Appl. Intell.* **2021**, *52*, 1436–1447. [CrossRef]
6. Miswan, N.H.; Chan, C.S.; Ng, C.G. Hospital readmission prediction based on improved feature selection using grey relational analysis and LASSO. Grey Syst. *Theory Appl.* **2021**, *11*, 796–812. [CrossRef]
7. Fan, J.Q.; Li, R.Z. Comment: Feature Screening and Variable Selection via Iterative Ridge Regression. *Technometrics* **2020**, *62*, 434–437. [CrossRef]
8. Lin, X.Q.; Ren, C.; Li, Y.; Liang, Y.J.; Yue, W.T.; Liang, J.Y. Extraction of eucalyptus plantation forest based on Relief F-RFE feature selection. *Sci. Surv. Mapp.* **2023**, *48*, 107–115.
9. Xiang, S.Y.; Xu, Z.H.; Zhang, Y.W.; Zhang, Q.; Zhou, X.; Yu, H.; Li, B.; Li, Y.F. Construction and Application of ReliefF-RFE Feature Selection Algorithm for Hyperspectral Image Classification. *Spectrosc. Spectr. Anal.* **2022**, *42*, 3283–3290. [CrossRef]
10. Ai, C. A Method for Cancer Genomics Feature Selection Based on LASSO-RFE. *Iran. J. Sci. Technol. Trans. A Sci.* **2022**, *46*, 731–738. [CrossRef]
11. Tang, T.T.; Chen, T.; Gui, G. A Novel mRMR-RFE-RF Method for Enhancing Medium- and Long-Term Hydrological Forecasting: A Case Study of the Danjiangkou Basin. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 14919–14934. [CrossRef]
12. Han, X.S.; Shi, Y.; Tong, R.J.; Wang, S.T.; Zhang, Y. Research on short-term load forecasting of power system based on IWOA-KELM. *Energy Rep.* **2023**, *9*, 238–246. [CrossRef]
13. Li, Y.T.; Wu, K.; Liu, J. Self-paced ARIMA for robust time series prediction. *Knowl.-Based Syst.* **2023**, *269*, 110489. [CrossRef]
14. Almazrouee, A.I.; Almeshal, A.M.; Almutairi, A.S.; Alenezi, M.R.; Alhajeri, S.N. Long-Term Forecasting of Electrical Loads in Kuwait Using Prophet and Holt–Winters Models. *Appl. Sci.* **2020**, *10*, 5627. [CrossRef]
15. Shakeel, A.; Chong, D.T.; Wang, J.S. Load forecasting of district heating system based on improved FB-Prophet model. *Energy* **2023**, *278*, 127637. [CrossRef]
16. Ji, L.; Li, S.L. A dynamic financial risk prediction system for enterprises based on gradient boosting decision tree algorithm. *Syst. Soft Comput.* **2025**, *7*, 200189. [CrossRef]
17. Samantaray, S.; Sahoo, A.; Baliarsingh, F. Groundwater level prediction using an improved SVR model integrated with hybrid particle swarm optimization and firefly algorithm. *Clean. Water* **2024**, *1*, 100003. [CrossRef]
18. Wang, X.M.; Liang, X.J.; Zhang, C.B.; Yang, C.H.; Gui, W.H.; Liu, Y.Q. Towards real-time adaptive prediction of rotary kiln processes: An enhanced framework combining parallel temporal convolution and long short-term memory networks. *Eng. Appl. Artif. Intell.* **2025**, *144*, 109993. [CrossRef]
19. Dey, M.; Wickramarachchi, D.; Rana, S.P.; Simmons, C.V.; Dudley, S. Power Grid Frequency Forecasting from μPMU Data using Hybrid Vector-Output LSTM network. In Proceedings of the 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Grenoble, France, 23–26 October 2023; pp. 1–5.
20. Li, J.; Deng, D.Y.; Zhao, J.B.; Cai, D.S.; Hu, W.H.; Zhang, M.; Huang, Q. A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2443–2452. [CrossRef]
21. Yan, H.F.; Yu, X.Y.; Li, D.W.; Xiang, Y.; Chen, J.; Lin, Z.Y.; Shen, J.W. Research on Commercial Sector Electricity Load Model Based on Exponential Smoothing Method. *Artif. Intell. Secur.* **2022**, *13338*, 189–205. [CrossRef]
22. Cao, Y.J.; Wu, Q.W.; Zhang, H.X.; Li, C.G.; Zhang, X. Chance-Constrained Optimal Configuration of BESS Considering Uncertain Power Fluctuation and Frequency Deviation Under Contingency. *IEEE Trans. Sustain. Energy* **2022**, *13*, 2291–2303. [CrossRef]
23. NESO National Energy System Operator. Available online: https://data.nationalgrideso.com/system/system-frequency-data?from=40resources (accessed on 24 April 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.