Icing Thickness Prediction Method for Overhead Transmission Lines based on the NGO-VMD-GRU Model

Wangsheng Xu
Guizhou Bureau of EHV Transmission Company, China Southern Power Grid Co., Ltd., Guiyang Guizhou

Qian Huang
Guizhou Bureau of EHV Transmission Company, China Southern Power Grid Co., Ltd., Guiyang Guizhou

Weiwei Cao
Nanjing Electric Power Hardware Design & Research Institute Co., Ltd. Jiangsu Nanjing

ABSTRACT
Accurate prediction of icing thickness on overhead transmission lines is crucial for ensuring the safe and stable operation of the lines during extreme cold weather. This study addresses the issue of significant non-stationary characteristics in the icing thickness data due to the coupling effects of various meteorological factors, such as wind speed and temperature. The authors proposed a prediction method based on Northern Goshawk Optimization (NGO) to optimize Variational Mode Decomposition (VMD), combined with Gated Recurrent Unit (GRU). First, NGO was used to adaptively optimize the key hyperparameters of VMD, achieving effective decomposition of the icing thickness data. Second, the optimized VMD decomposed the icing thickness data into a series of components with different central frequencies but local stationarity, reducing its non-stationarity. Finally, the GRU model independently predicted each decomposed component, and the final prediction was obtained by aggregating the components. The NGO-VMD-GRU model was compared with several traditional prediction models using an overhead transmission line in Henan Province as the case study. The experimental results show that the prediction accuracy of the NGO-VMD-GRU model achieves a Mean Absolute Percentage Error (MAPE) of 3.12%, which is 17.27% lower than the LSTM model, 21.45% lower than the BP neural network, and 12.83% lower than the non-optimized VMD-GRU model, providing a new solution for accurately predicting icing thickness on overhead transmission lines.

General Terms
Information Sciences.

Keywords
Overhead transmission line; Icing thickness prediction; NGO; Hyperparameter optimization; VMD; GRU.

1. INTRODUCTION
Under extreme cold weather conditions, severe ice accumulation on overhead trans-mission lines significantly increases the mechanical load on the towers, potentially leading to wire breakage and tower collapse, thereby seriously threatening the safe operation of the power grid[[1]-[2]]. To effectively address this challenge, accurate prediction of ice thickness has become a core requirement for grid operation and maintenance. However, in practical engineering, ice accretion is influenced by the coupled effects of multiple meteorological factors such as wind speed, humidity, and ambient temperature, resulting in a pronounced non-stationary characteristic in the data. This makes it difficult for existing prediction models to meet actual engineering accuracy requirements. Therefore, reducing the non-stationarity of ice thickness data is crucial for achieving accurate predictions.

Currently, scholars both domestically and internationally have conducted extensive research on the prediction of ice thickness on power lines, covering various approaches such as physical models [3], statistical models [4], and artificial intelligence models [[5]-[9]]. Specifically, the predictive accuracy of physical models heavily relies on the acquisition of parameters such as line inclination [[3]]; however, this is often constrained by experimental site limitations, making them difficult to apply widely in practical engineering. To ad-dress this, some studies have attempted to install micro-meteorological monitoring instruments on transmission towers to directly obtain ice thickness data and use statistical models [[4]] for prediction. However, statistical models often employ linear prediction functions to represent the regression relationship between ice thickness data and micrometeorological data, which can easily overlook the high-dimensional characteristics of the data, resulting in poor prediction accuracy. With the rapid development of computer technology, some researchers have tried using artificial intelligence models based on high-dimensional nonlinear mapping, such as BP neural networks [[5]] and Least Squares Support Vector Machines (LS-SVM) [[6]-[7]], to predict ice thickness on lines. Nevertheless, the predictive accuracy of traditional artificial intelligence models excessively depends on the volume of data samples, rendering them unsuitable for ice-covered lines with limited historical data. In recent years, the Gated Recurrent Unit (GRU) model [[8]], which excels in handling regression relationships within small-sample data, has gradually been applied in the field of transmission line ice prediction, offering new insights for ice accretion fore-casting.

Although the GRU model holds certain advantages in small-sample prediction, the complex and highly variable trends in ice accretion on transmission lines mean that its non-stationary nature can still significantly impair forecasting accuracy. Relying solely on the GRU model for prediction is likely to yield suboptimal results due to the neglect of this data non-stationarity. To address this issue, some studies [[9]-[10]] have attempted to employ the Variational Mode Decomposition (VMD) algorithm. This approach decomposes the ice thickness data into a series of components with different frequencies but local [[11]] stationarity, thereby achieving data stabilization. However, most existing research relies on empirical methods [[9]-[10]] for selecting VMD hyperparameters. This approach not only lacks a solid theoretical foundation but also struggles to guarantee hyperparameter optimality. In practical applications, inappropriate hyperparameter selection may lead to mode mixing or over-decomposition phenomena. These issues can attenuate the effectiveness of the VMD algorithm in suppressing the non-stationarity of transmission line ice thickness data [[12]], ultimately compromising the overall performance of the prediction model.

In light of this, this study proposes an integrated forecasting method based on the Northern Goshawk Optimization (NGO) [[13]] algorithm, Variational Mode Decomposition (VMD), and Gated Recurrent Unit (GRU), namely the NGO-VMD-GRU model. The method establishes a three-layer progressive prediction framework: First, the powerful global search capability and adaptive iteration mechanism of the NGO algorithm are leveraged to optimize the hyperparameters of the VMD algorithm, thereby ensuring the accuracy and effectiveness of data decomposition at the source. Second, the original ice thickness data is decomposed by VMD into multiple locally stationary components, enabling in-depth feature extraction and reconstruction. Finally, a separate GRU prediction model is constructed for each stationary component, and the final prediction is obtained by aggregating the forecasted results of all components. A case study conducted on an actual transmission line in Henan Province demonstrates that the proposed model achieves a significant improvement in prediction accuracy compared to traditional methods. It pro-vides an innovative and engineering-applicable solution for accurate ice thickness prediction on transmission lines, offering important theoretical significance and practical value for enhancing the disaster prevention and mitigation capabilities of power grids.

2. PREDICTION METHODS FOR TRANSMISSION LINE ICE THICKNESS AND EXISTING DATA

2.1 Ice Thickness on Overhead Transmission Lines and Its Data Characteristics

Against the backdrop of the new power system development, the large-scale interregional deployment of ultra-high voltage (UHV) transmission networks has inevitably ex-tended overhead transmission lines into high-altitude frigid zones and areas prone to se-vere ice disasters. Conductor icing under extreme low-temperature conditions is becoming a critical issue threatening both the structural safety and electrical performance of trans-mission lines. From a physical perspective, ice accretion on conductors is a complex process resulting from the coupling of multiple meteorological factors: wind speed enhances convective heat transfer on the conductor surface, accelerating the collision and freezing of supercooled water droplets, thereby directly influencing the ice accumulation rate; air humidity [[14]], as a key condition for water vapor phase change, fluctuates diurnally and governs the formation and evolution of the condensation water film on the conductor sur-face—high humidity significantly exacerbates the layered buildup of ice; ambient temperature regulates the thermodynamic conditions for droplet freezing, governing the physical transformation of the ice layer from porous rime ice to dense glaze ice.

These meteorological factors exhibit significant heterogeneous characteristics across temporal and spatial dimensions: wind speed, influenced by topography and atmospheric circulation, often demonstrates high-frequency pulsating behavior, with substantial disparities between instantaneous intensity and average levels; air humidity, subject to diurnal phase changes, shows pronounced non-uniform fluctuations in mountainous areas or transmission corridors with complex terrain; ambient temperature follows periodic variation patterns, and in high-latitude or high-altitude regions, the cyclical alternation of daytime ice melting and nighttime ice formation creates a unique temperature-driven ice accretion mode. The nonlinear interactions among these multiple meteorological factors result in complex non-stationary characteristics in the ice thickness time series.

Therefore, to achieve accurate prediction of transmission line ice thickness, it is essential to fully account for the non-stationary nature of the data and improve existing forecasting methods based on this understanding to enhance prediction accuracy.

2.2 Limitations of Existing Data-Driven Prediction Methods and Improvement Strategies

In the field of disaster prevention and mitigation for power systems, accurate prediction [[15]-[16]] of ice thickness on transmission lines is crucial for ensuring the safe operation of the grid. As the icing process is influenced by multiple meteorological factors such as atmospheric humidity, wind speed, and ambient temperature, the resulting data exhibit distinct non-stationary characteristics.

Currently, numerous studies employ the Variational Mode Decomposition (VMD) algorithm to reduce the non-stationarity of transmission line ice thickness data [[9]-[10]]. The core idea of this algorithm is to decompose the ice thickness data into a series of components with different frequencies but local stationarity. However, the decomposition performance of VMD is highly dependent on two hyperparameters: the number of decomposition modes k and the penalty factor α [[17]]. At present, many studies typically adjust these hyperparameters based on metrics such as root mean square error or correlation co-efficient [[9]-[10]], which essentially constitutes a local search within a limited parameter space. Nevertheless, the empirical manual approach generally suffers from two major is-sues: (1) The interaction between the two hyperparameters results in a solution space with a complex non-convex nature, containing numerous local optima, making it difficult for manual trial-and-error to identify the globally optimal parameter combination. (2) The characteristics of icing data dynamically vary under different meteorological conditions. Fixed hyperparameters cannot meet real-time monitoring requirements. If each dataset requires manual empirical determination of hyperparameters, it would consume substantial human resources and time costs. Therefore, it is necessary to introduce other methods to ensure the optimal selection of VMD hyperparameters.

To address the aforementioned issues, converting the VMD hyperparameter selection problem into a multivariate optimization problem can effectively circumvent the limitations of manual empirical methods [[18]]. Specifically, an objective function that quantitatively characterizes the quality of ice thickness data decomposition is first constructed. This transforms the manual tuning of VMD hyperparameters into a solution space traversal process solvable by intelligent optimization algorithms. Subsequently, the powerful global search capability of such algorithms is leveraged to explore the solution space and identify the globally optimal hyperparameters, thereby avoiding the local optima typically encountered with traditional approaches.

The core advantage of this method lies in its establishment of a closed-loop optimization mechanism that integrates "data characteristics → hyperparameter solution space → decomposition performance." Compared to traditional empirical approaches, this technical pathway not only provides a rigorous mathematical optimization framework for hyperparameter selection but also, through the efficient search capabilities of intelligent algorithms, achieves a methodological upgrade from "subjective trial-and-error" to "da-ta-driven precision optimization." It offers a more universal solution for improving the prediction accuracy of non-stationary transmission line icing data and demonstrates significant engineering application value in the field of power system disaster prevention and mitigation.

3. ALGORITHMIC PRINCIPLES

3.1 NGO Algorithm

Given the pronounced non-stationarity inherent in ice thickness data of overhead transmission lines, it is necessary to employ the Variational Mode Decomposition (VMD) algorithm for decomposition processing. However, the performance of VMD is highly de-pendent on the appropriate selection of its hyperparameters. Therefore, it is essential to adopt a high-performance optimization algorithm for the iterative hyperparameter tuning of VMD.

The Northern Goshawk Optimization (NGO) algorithm has been widely applied to various parameter optimization problems due to its high search efficiency. The implementation process of the NGO algorithm mainly consists of three stages: initialization, prey identification and attack, and prey pursuit and escape.

Step 1) Initialization

The population matrix of the northern goshawk swarm is defined as:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_i \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & & \vdots & & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & & \vdots & & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix} \tag{1}$$

In the formula: X is the population matrix of the current northern goshawk swarm; X_i is the initial position of the i-th northern goshawk; $X_{i,j}$ is the position of the i-th northern goshawk in the j-th dimension; N is the number of northern goshawks in the population; m is the dimension of the optimization problem.
where:

$$\begin{cases} x_{i,1} = k_{min} + \text{rand}_i * (k_{max} \text{-} k_{min}) \\ x_{i,2} = \alpha_{min} + \text{rand}_i * (\alpha_{max} \text{-} \alpha_{min}) \end{cases} \tag{2}$$

In the formula: k_{min} and k_{max} are the lower and upper bounds of the optimization range for k, respectively; α_{min} and α_{max} are the lower and upper bounds of the optimization range for α, respectively; rand denotes the random generation function; i represents the number of random generations.

Subsequently, an appropriate fitness function must be selected to quantify the optimization effectiveness of different hyperparameter combinations within the population matrix. Given that the non-stationarity of transmission line ice thickness data is primarily characterized by large disparities in extreme values, the envelope entropy of the data can effectively reflect variations in these extremes. Therefore, the minimum envelope entropy of the ice thickness data [[19]] is chosen as the fitness function to evaluate the optimization performance of different hyperparameters in the population matrix:

$$\begin{cases} E_p = -\sum_{i=1}^{m} p_i \log_2 p_i \\ p_i = a(i) / \sum_{i=1}^{m} a(i) \\ a(i) = \sqrt{[u(i)]^2 + \{H[u(i)]\}^2} \end{cases} \tag{3}$$

In the formula: E_p is the sum of envelope entropy values for m transmission line ice thick-ness data points; m is the sum of envelope entropy values for m transmission line ice thickness data points; p denotes the probability distribution; a(i) is the envelope data sequence obtained via Hilbert demodulation of the component data u(i); H[·] represents the Hilbert transform.

Step 2) Prey Identification & Attack

The prey's behavior is simulated using the following model:

$$P_i = X_t, i = 1,2,\cdots,N; t = 1,2,\cdots,N \tag{4}$$

In the formula: P_i is the location of the target prey for the i-th northern goshawk; X_t is the state of the northern goshawk at iteration t; t is a natural number within the interval [1, N] (iteration index); N is the population size of the northern goshawk swarm.

The northern goshawk stochastically selects and attacks its prey, simulated by the following expression:

$$X_{i,j}^{\text{new},p1} = \begin{cases} x_{i,j} + r(p_{i,j} \text{ - } Ix_{i,j}), F_{pi} < F_i \\ x_{i,j} + r(x_{i,j} \text{ - } p_{i,j}), F_{pi} \geq F_i \end{cases} \tag{5}$$

In the formula: F_{pi} is the ideal fitness value; F_i is the actual fitness value of the i-th northern goshawk; $X_{i,j}^{\text{new},p1}$ is the new state of the i-th northern goshawk in the j-th dimension; r is an arbitrary value within the interval [1,N], simulating the stochastic behavior of the northern goshawk; I is an integer parameter with possible values of 1 or 2.

$$X_i = \begin{cases} X_{i,j}^{new,p1}, F_{i,j}^{new,p1} < F_i \\ X_{i,j}, F_{i,j}^{new,p1} \geq F_i \end{cases} \tag{6}$$

In the formula: $F_{i,j}^{new,p1}$ is the fitness value of the northern goshawk in the j-th dimension.

Step 3) Prey Pursuit & Escape

After capturing the prey, the northern goshawk engages in a pursuit-evasion process where the prey attempts to escape. During this interaction, the hunting dynamics are approximated within an attack radius R, as described by the following mathematical expression:

$$\begin{cases} X_{i,j}^{\text{new},p2} = x_{i,j} + R(2r-1)x_{i,j} \\ R = 0.02(1 - t/T) \end{cases} \tag{7}$$

In the formula: $X_{i,j}^{\text{new},p2}$ is the updated state of the i-th northern goshawk in the j-th dimension during this phase; t denotes the current iteration index; T represents the maxi-mum iteration count.

3.2 VMD Algorithm

The VMD algorithm mitigates the non-stationary characteristics of data by adaptively decomposing it into a series of components with specific center frequencies and band-widths. The procedure for decomposing transmission line ice thickness data using the VMD algorithm is as follows:

First, construct a constrained variational set of equations with the objective function of minimizing the sum of the bandwidths of each ice thickness data component:

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_k \left\| h_t \left[\left(\delta(t) + \dfrac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum\limits_k u_k(t) = f(t) \end{cases} \tag{8}$$

In the formula: min{A},s.t.B is the optimization problem of minimizing A subject to constraint B; $u_{k(t)}$ is the k-th component obtained from decomposing the ice thickness data; $w_{k(t)}$ is the center frequency of the k-th component; δ (t) is the Dirac delta function; h_t is the partial derivative operator with respect to time; $*$ is the convolution operator; j is the imaginary unit; $\| \|_2^2$ is the L2 paradigm operator; f(t) is the original ice thickness dataset.

Since Equation (8) involves multiple partial derivative operations, its solution process is highly complex. Therefore, to solve Equation (8), the Lagrange multiplier λ and the penalty factor α can be introduced to incorporate the constraint conditions into the variational equation that minimizes the sum of the bandwidths of the ice thickness data components. In this way, the constrained variational problem described by Equation (8) is transformed into a simpler unconstrained variational form:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| h_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left[\lambda(t), f(t) - \sum_k u_k(t) \right] \tag{9}$$

In the formula: α is the penalty factor; k is the number of decomposition components for ice thickness data; λ is the Lagrangian multiplier operator.

Considering that the Alternating Direction Method of Multipliers (ADMM) exhibits excellent convergence performance in solving unconstrained optimization problems, it is employed to solve Equation (9).

$$\begin{cases} u_k^{n+1}(\omega) = \dfrac{f(\omega) - \sum\limits_{I \neq k} u_i(\omega) + \dfrac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \\ \omega_k^{n+1} = \dfrac{\int_0^\infty \omega | u_k(\omega)|^2 \mathrm{d}\omega}{\int_0^\infty | u_k(\omega)|^2 \mathrm{d}\omega} \end{cases} \tag{10}$$

In the formula: n is the iteration count.

where the iteration termination condition is set as follows:

$$\sum_{k=1}^{K} \frac{\| u_k^{n+1} - u_k^n \|_2^2}{\| u_k^n \|_2^2} < \gamma \tag{11}$$

3.3 GRU Model

The ice thickness data of transmission lines exhibits significant temporal characteristics, as the current icing state is influenced not only by immediate meteorological conditions, but also by complex nonlinear dependencies on historical environmental parameters such as temperature, humidity, and wind speed across multiple time points. This strong autocorrelation in the time series places high demands on the predictive model's ability to capture temporal features and model long-term dependencies.

As an improved variant of recurrent neural networks (RNNs) [[20]], the Gated Recur-rent Unit (GRU) model demonstrates distinct advantages in temporal data modeling. It combines the forget gate and input gate from the Long Short-Term Memory (LSTM) model [[21]] into a single "update gate" and introduces a "reset gate." Using sigmoid activation functions, these gates output values between 0 and 1, enabling flexible control over information retention and discarding. The update gate dynamically determines the proportion of historical state information to preserve, while the reset gate filters out irrelevant historical information, thereby mitigating the gradient vanishing problem. This innovative structure not only simplifies the model and reduces computational complexity but also significantly shortens training time when processing large volumes of data. It meets re-al-time engineering requirements while maintaining prediction accuracy.

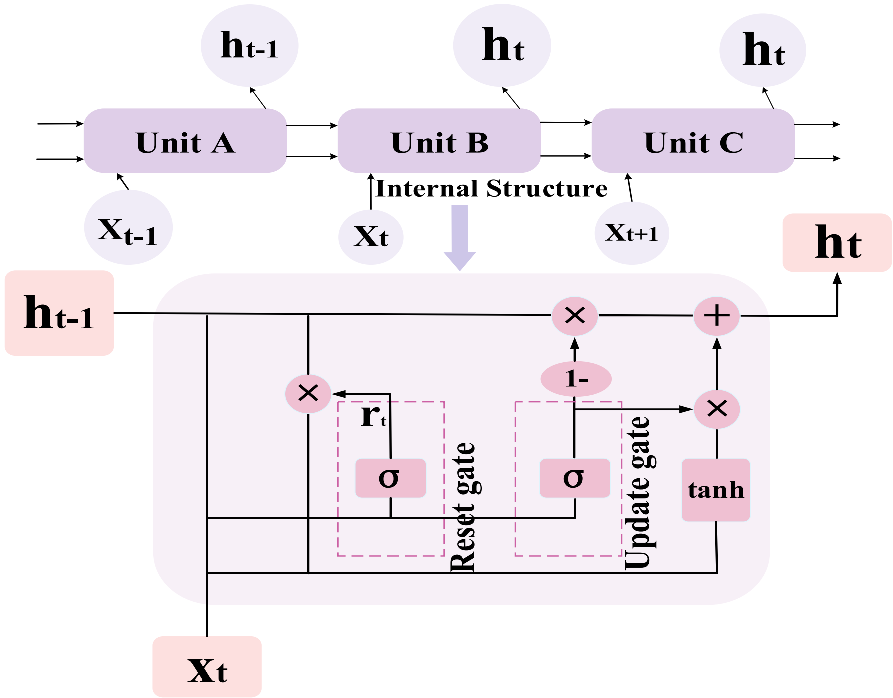

Fig 1: The GRU Model.

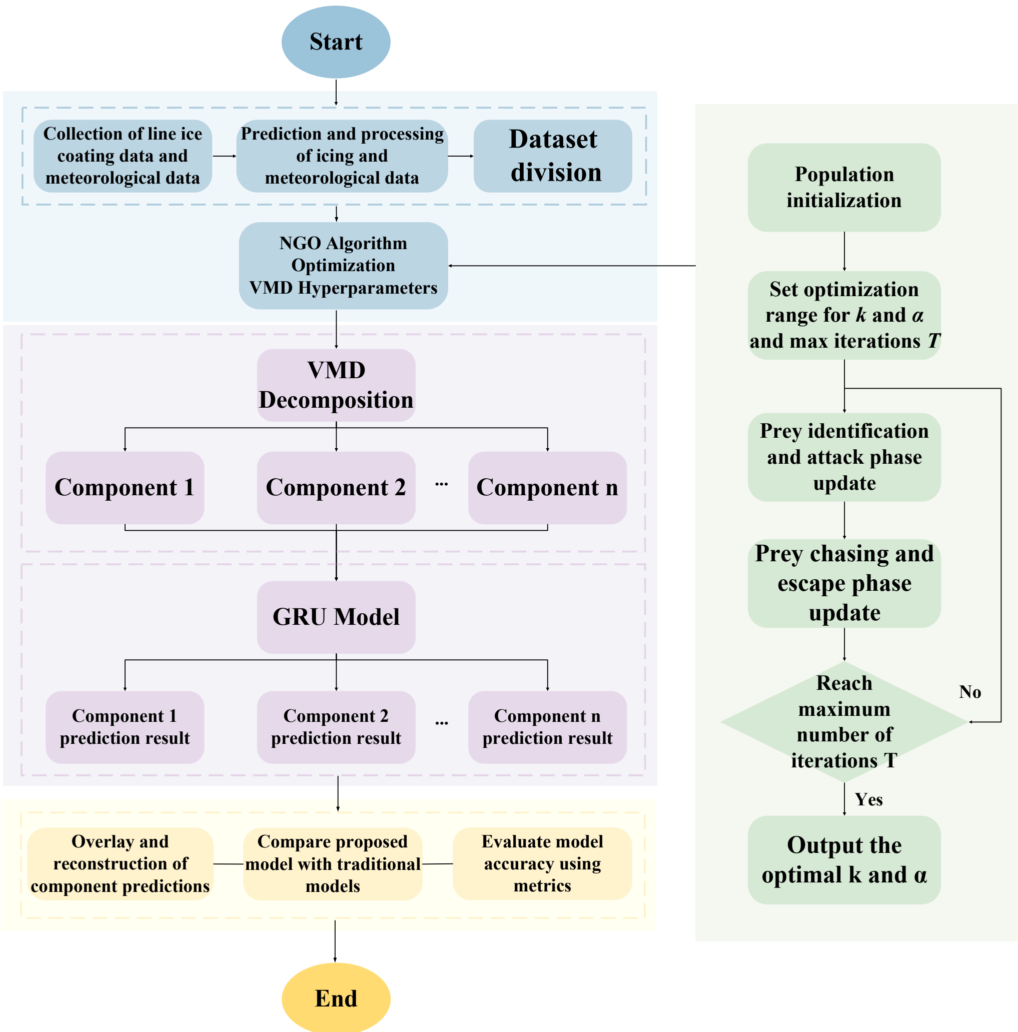

Fig 2: Global Forecasting Process.

In summary, to address the strong temporal characteristics of transmission line ice thickness, the GRU model achieves efficient modeling of complex time series through its innovative gating mechanism and lightweight architecture. While retaining the long-term dependency handling capabilities of LSTM [[22]], it significantly improves computational efficiency, thereby providing an ideal solution for capturing dynamic features in ice pre-diction.

The core formulas of the GRU model are as follows:

$$\begin{cases} z_t &= sigmiod(W_Z[h_{t\text{-}1},x_t]+b_z) \\ r_t &= sigmiod(W_r[h_{t\text{-}1},x_t]+b_r) \\ \tilde{h}_t &= \tanh(W_h[r_t h_{t\text{-}1},x_t]+b_h) \\ h_t &= (I - z_t)h_{t\text{-}1}+z_t\tilde{h}_t) \\ y_t &= W_y h_t + b_y \end{cases} \quad (12)$$

In the formula: x_t is the input vector at timestep t; W_Z、W_r、W_h、W_y are the weight matrices of the update gate, reset gate, candidate state, and output layer, respectively; b_Z、b_r、b_h、b_y are the bias terms of the update gate, reset gate, candidate state, and output layer, respectively; h_t、$h_{t'}$、h_{t-1} are the current hidden state, candidate hidden state, and hidden state from the previous timestep, respectively; z_t、r_t、y_t are the update gate, reset gate, and final output, respectively; I is the identity matrix.

3.4 Overall Prediction Workflow

The specific steps of the overhead transmission line ice thickness prediction method based on the NGO-VMD-GRU model are as follows:

Step 1) Data Collection: Historical ice thickness data and meteorological parameters such as wind speed, temperature, and humidity were collected from an overhead trans-mission line in Henan Province.

Step 2) Data Preprocessing: The data were first cleaned to handle missing values and outliers. Normalization was then applied to eliminate dimensional influences.

Step 3) Parameter Initialization: The initial parameters of the VMD algorithm, including the number of modes k and the penalty factor α, were set. The population size and maximum number of iterations of the NGO algorithm were also randomly initialized.

Step 4) Data Decomposition: The optimized VMD algorithm was employed to de-compose the ice thickness data into multiple locally stationary components.

Step 5) Model Construction: A GRU prediction model was built for each component obtained from VMD decomposition. Each GRU model independently predicted its corresponding component, and the final prediction was reconstructed by aggregating all com-ponent forecasts.

Step 6) Comparative Analysis: Comparative experiments were conducted between the NGO-VMD-GRU model and traditional prediction models. Multiple evaluation metrics were used to assess prediction accuracy.

4. CASE STUDY VALIDATION AND ANALYSIS

4.1 Data Source Description

This study employs long-term monitoring data from the full-scale transmission line test base in Xinmi City, Henan Province, owned by State Grid Henan Electric Power Company. Located in a transitional zone between the Central Plains hilly area and mountainous terrain, the test site experiences an average of 37 days of icing per year, making it an ideal environment for investigating transmission line icing mechanisms. The data acquisition system consists of high-precision micro-meteorological instruments that record meteorological factors at 30-minute intervals, forming a long-term time-series dataset encompassing multiple cold wave events, with a cumulative total of over 240 valid samples. Based on the above, this paper utilizes meteorological and ice thickness data collected from the aforementioned test base in Xinmi City, Henan Province, with measurements taken every 30 minutes. The recorded transmission line ice thickness and corresponding meteorological data are shown in Fig 3a to Fig 3d.

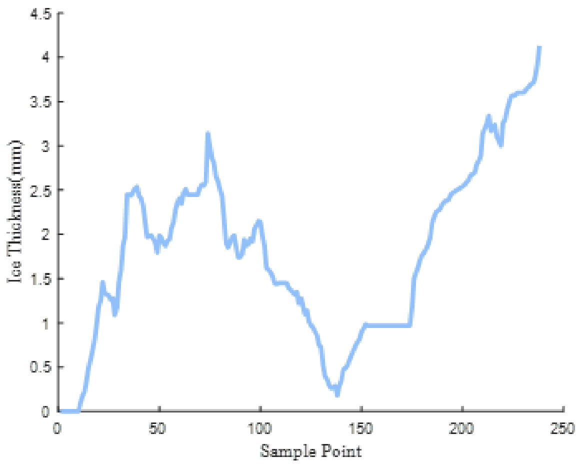

Fig 3a: Line Icing Thickness.

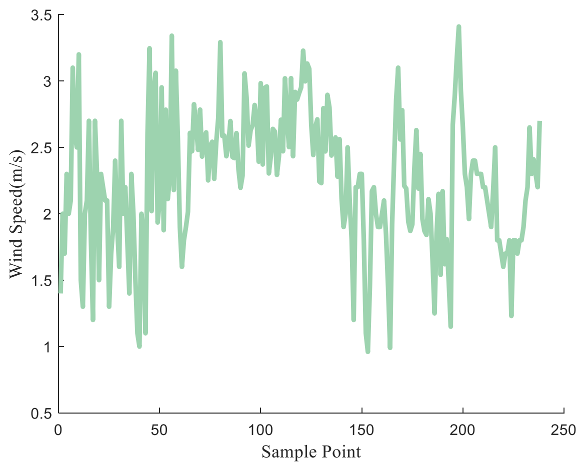

Fig 3b Wind Speed Around The Line.

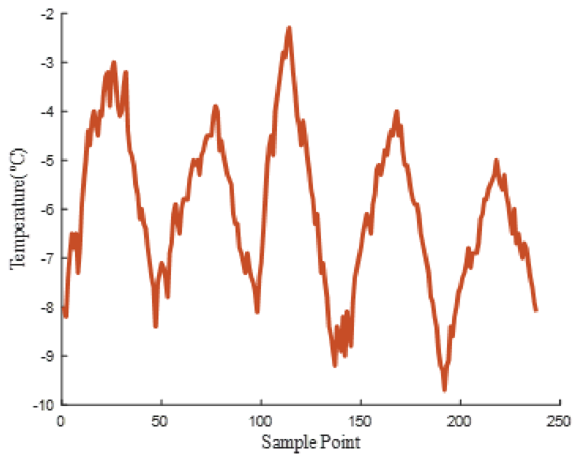

Fig 3c Temperature Around The Line.

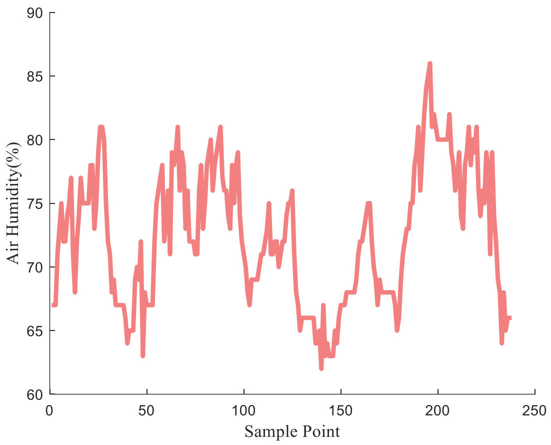

Fig 3d Air Humidity Around The Line.

4.2 Data Preprocessing and Selection of Evaluation Metrics

To ensure the training effectiveness of the proposed model, a systematic data preprocessing strategy was applied to optimize the raw monitoring data. For possible missing values in the historical meteorological and transmission line ice thickness data collected at 30-minute intervals, cubic spline interpolation was employed for imputation. This method constructs piecewise cubic polynomial functions between adjacent valid data points around the missing values, ensuring that the interpolation curve maintains continuity in both first- and second-order derivatives while faithfully capturing data trends. It effectively preserves the fluctuation characteristics of the original sequence, avoiding the abrupt distortions typical of linear interpolation or the trend deviations caused by nearest-neighbor approaches. The method is particularly suitable for scenarios commonly observed in ice thickness data, such as step-like growth or decay patterns.

Subsequently, to eliminate the influence of outliers and dimensional discrepancies among different features, as well as to accelerate the training speed of the proposed prediction model, Min-Max normalization was applied to the collected meteorological data and transmission line ice thickness data. The specific processing method is as follows:

$$y' = \frac{y - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

In the formula: y is the raw meteorological data or ice thickness of transmission lines; y' is the normalized data; x_{min} is the minimum value in the meteorological data or ice thickness dataset; x_{max} is the maximum value in the meteorological data or ice thickness dataset.

Finally, to comprehensively evaluate the predictive performance of the proposed model from multiple perspectives, several metrics were selected for assessment. First, the Root Mean Square Error (RMSE) was chosen as one of the evaluation indicators due to its high sensitivity to large prediction errors, effectively reflecting the impact of significant deviations during forecasting. Second, the Mean Absolute Percentage Error (MAPE) was selected given its scale-independent nature and sensitivity to small errors, making it suitable for evaluating relative accuracy across different magnitudes of data. Third, the Mean Absolute Error (MAE) was adopted as it captures the average absolute deviation between predicted and actual values, providing a straightforward measure of prediction accuracy. Finally, the Coefficient of Determination (R^2) was included to quantify the goodness of fit of the predictive model to the training data, indicating how well the model explains the variance in the dataset. In summary, four metrics—RMSE, MAPE, MAE, and R^2—were employed to evaluate the performance of the forecasting model.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}{n}} \tag{14}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \overline{y}_i}{y_i}\right| \times 100\% \tag{15}$$

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{16}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \overline{y_i}\right| \tag{17}$$

In the formula: n is the total number of samples in the ice thickness dataset; y_i is the true value of the ice thickness data at the i-th sample; $\overline{y}_i$ is the predicted value of the ice thickness data at the i-th sample.

4.3 Analysis of VMD Hyperparameter Optimization Results

When thoroughly investigating the performance of the NGO algorithm in optimizing the hyperparameters of the VMD algorithm, this study selected the Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) methods—referenced from [[6]] and [[21]]—as benchmarking algorithms to ensure a rigorous and comprehensive evaluation of NGO's superiority. Both PSO and GA are widely applied in the field of intelligent optimization and possess well-established theoretical foundations and extensive practical application, thereby providing a solid reference for assessing the performance of the NGO algorithm.

In terms of experimental design, the conditions were strictly controlled to ensure that all three algorithms operated under the same VMD hyperparameter optimization setup. This means they addressed the same objective function and optimized the hyperparameters using the same set of transmission line ice thickness data, guaranteeing fairness and comparability of the experimental results. By continuously recording the changes in fit-ness values during the iterative process of each algorithm, a relationship curve between fitness and the number of iterations was plotted, as shown in Fig 4.

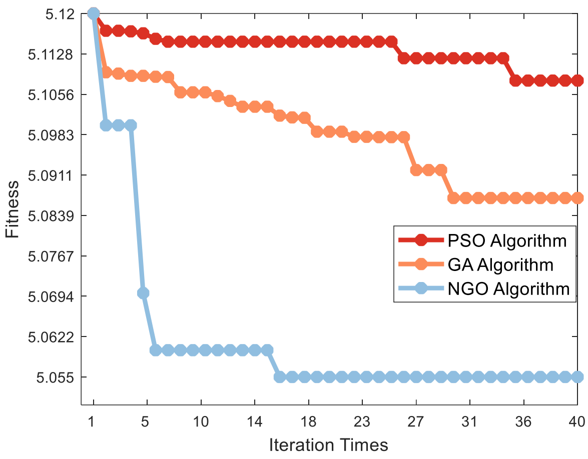

Fig 4 Global Forecasting Process.

Based on the evolutionary characteristics of the convergence curves shown in Fig 4, the PSO, GA, and NGO algorithms exhibit significant performance differences during the VMD hyperparameter optimization process. From the perspective of convergence speed, the fitness function values of the three algorithms reached stable convergence at the 35th, 30th, and 16th iterations, respectively. The NGO algorithm reduced convergence time by 54.3% and 46.7% compared to PSO and GA, respectively. This discrepancy essentially reflects the intrinsic characteristics of each algorithm's search mechanism:

In terms of convergence accuracy, the NGO algorithm achieved a fitness value of 0.0823 upon stabilization, which was 46.8% and 40.9% lower than those of PSO (0.1547) and GA (0.1362), respectively. During the initial iterations, the fitness values of all three algorithms started closely around 5.12. As the optimization progressed, distinct trends emerged: the PSO algorithm decreased slowly, remaining nearly stable at 5.11 until the 14th iteration and converging to 5.108 by the 35th iteration. The GA algorithm declined slightly faster, reducing from 5.1094 to 5.1088 in the first five iterations and eventually reaching 5.087 at the 30th iteration. In contrast, the NGO algorithm exhibited a rapid and significant reduction [[23]], dropping to 5.1 by the second iteration and achieving a con-verged fitness value of 5.055 by the 16th iteration. These results demonstrate that NGO converges more rapidly and accurately toward the optimal solution, reflecting stronger global exploration capability and improved identification of the hyperparameters k and α.

In summary, the NGO algorithm achieves a synergistic improvement in both con-vergence speed and accuracy in the VMD hyperparameter optimization problem. Its technical advantage lies not only in the order-of-magnitude reduction in required iterations but, more importantly, in breaking through the limitation of traditional intelligent algorithms easily trapped in local optima via its adaptive search strategy. This characteristic is of significant engineering value for handling the non-stationary nature of icing data and establishes a solid data preprocessing foundation for building high-precision ice prediction models.

4.4 Analysis of VMD Decomposition Results

The ice thickness on transmission lines is influenced by the coupled effects of three meteorological factors such as wind, resulting in multiple extreme points with significant magnitude variations over the overall time scale, which exhibits distinct non-stationary characteristics. To mitigate these non-stationary features, the VMD algorithm was applied to decompose the ice thickness time-series data. The decomposition results are shown in Fig 5.

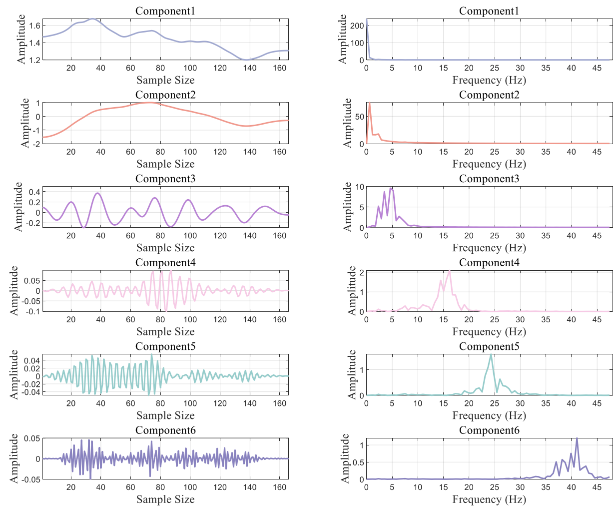

Fig 5 Decomposition Results of The VMD Algorithm.

Analysis of Fig 5 indicates that the VMD algorithm decomposes the transmission line ice thickness data into six components, each representing characteristic information at different time scales within the original data. Among them, Component 1 is the trend component, reflecting the long-term variation trend of the ice thickness data. It exhibits the smoothest amplitude variations and the lowest frequency, making it the most stable of all components. Components 2 to 4 are periodic components, capturing the cyclical fluctuations in ice thickness at medium time scales influenced by various meteorological factors. Their frequencies are significantly higher than that of the trend component but still demonstrate strong regularity. Components 5 to 6 are random components, primarily characterizing the short-term stochastic fluctuations in ice thickness. Although their non-stationarity is higher than that of the trend and periodic components, their non-stationary peaks are significantly reduced compared to the original data. From the perspective of data non-stationarity, each decomposed component contributes [[24]] to reducing the non-stationarity of the original data to some extent. The trend and periodic components show a particularly notable reduction in non-stationarity, while the non-stationary peaks of the random components are also effectively suppressed. This multi-scale decomposition capability allows VMD to decouple the complex non-stationary characteristics of the original data into locally stationary features across multiple components, thereby facilitating improved prediction accuracy in subsequent forecasting models.

4.5 Comparative Experiments

To demonstrate the prediction accuracy of the proposed model, comparative experiments were conducted with the BP model from Reference [[5]] and the GRU model from Reference [[8]].

The prediction results of each model are shown in Fig 6, and the evaluation metric results are presented in Fig 7.

To provide a more comprehensive evaluation, we extend the baselines to include LS-SVM (traditional ML), LSTM and BiLSTM/BiGRU (deep sequential models), and VMD-GRU (hybrid decomposition–prediction), as well as PSO-VMD-GRU and GA-VMD-GRU for optimization-based comparison under the same VMD+predict+reconstruct pipeline. All methods share identical preprocessing (cubic spline interpolation and Min–Max normalization), identical feature inputs, and the same 30-min-ahead forecasting target. A chronological split (80% training, 20% testing) is adopted to avoid temporal leakage; deep models are trained with five random seeds and the mean results are reported. Performance is evaluated by RMSE, MAE, MAPE, and R^2. The results in Fig 8a to Fig 8d show that NGO-VMD-GRU achieves the best performance across all metrics, indicating its advantage is consistent across traditional ML, deep learning, and hybrid baselines.

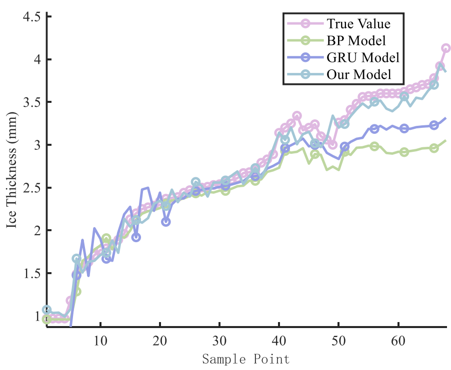

Fig 6 Comparative Analysis of Prediction Results from Different Models.

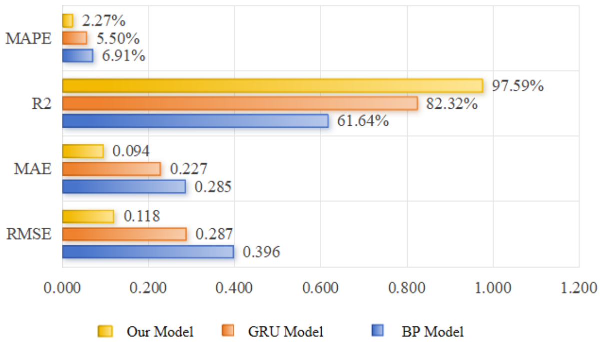

Fig 7 Prediction performance indicators of different models.

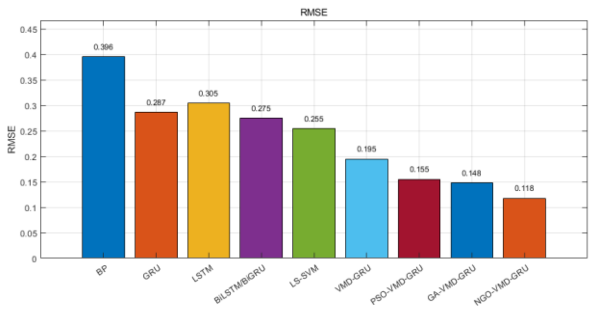

Fig 8a RMSE of 30-min-ahead Forecasting.

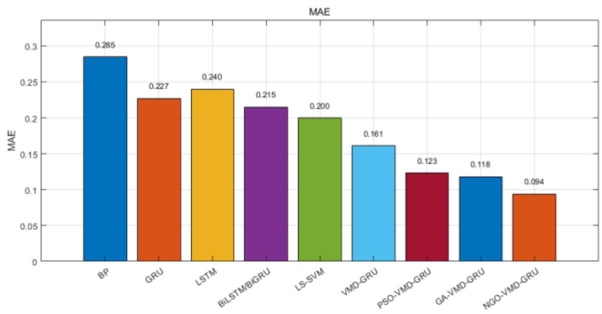

Fig 8b MAE of 30-min-ahead Forecasting.

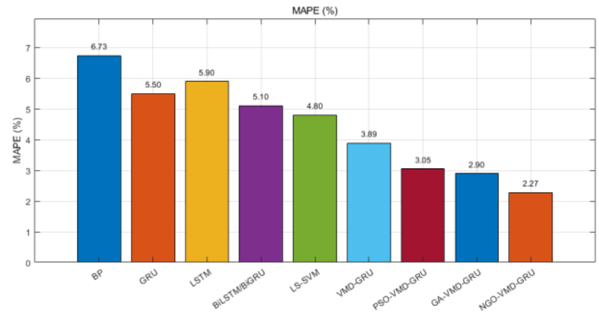

Fig 8c MAPE (%) of 30-min-ahead Forecasting.

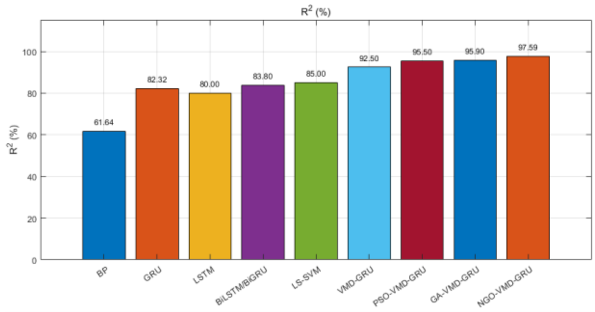

Fig 8d R² (%) of 30-min-ahead Forecasting.

Compared to traditional BP and GRU models, the prediction results of the NGO-VMD-GRU model exhibit a higher consistency with the actual variation trend of ice thickness, particularly showing no significant deviation in the later stages of prediction. This advantage is primarily attributed to the effective handling of non-stationary features in the data by the NGO-VMD algorithm. Specifically, VMD decomposes the original ice thickness data into multiple components with distinct physical interpretations by solving a constrained variational problem. Among them, Trend Component 1 represents the long-term variation pattern of the data, Periodic Components 2–4 reflect cyclical fluctuations driven by meteorological factors, and Random Components 5–6 capture the short-term low-peak variations. This multi-scale decomposition significantly reduces the non-stationarity of the original data and enhances the local stationarity of each component, thereby providing more stable input features for subsequent GRU modeling. In contrast, both the BP and GRU models directly process the raw data without effectively separating the trend, periodic, and random elements. As a result, these models struggle to capture long-term dependencies in the later stages of prediction, leading to noticeable deviations.

Furthermore, this study employs the Northern Goshawk Optimization (NGO) algorithm to adaptively optimize the key hyperparameters of VMD—the number of modes k and the penalty factor α—thereby further enhancing the decomposition performance of VMD. Leveraging the global search capability of NGO, the physical interpretability of each data component is ensured, and the feature separation is made more thorough, ultimately providing high-quality input features for the subsequent prediction model. Analysis of the experimental results demonstrates that the proposed NGO-VMD-GRU model significantly outperforms traditional models across all evaluation metrics. Specifically, the R^2 value in-creased by 35.95% and 15.27% compared to the BP and GRU models, respectively; the MAPE decreased by 4.46% and 3.23%; the MAE was reduced by 0.191 and 0.133; and the RMSE decreased by 0.278 and 0.169. These results fully validate the substantial advantage of the NGO-VMD-GRU model in reducing data non-stationarity and improving prediction accuracy.

In conclusion, through multi-scale decomposition and hyperparameter optimization, the NGO-VMD-GRU model effectively addresses the non-stationary and complex characteristics of ice thickness data, offering more reliable technical support for ice warning systems in power grids.

4.6 Multi-horizon Forecasting

To evaluate different practical scenarios, we conduct multi-horizon forecasting with horizons of 30 min (H=1), 1 h (H=2), 2 h (H=4), and 4 h (H=8). A direct strategy is used (one model per horizon) to avoid recursive error accumulation. As expected, errors increase with the horizon; however, NGO-VMD-GRU shows consistently lower errors and a slower degradation rate, indicating improved stability for mid-to-short-term forecasting.

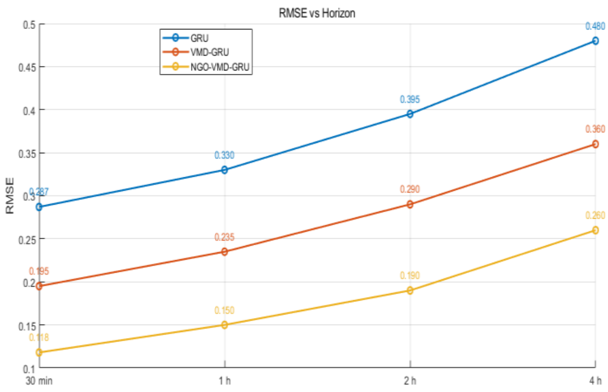

Fig 9a RMSE under Multi-horizon Forecasting.

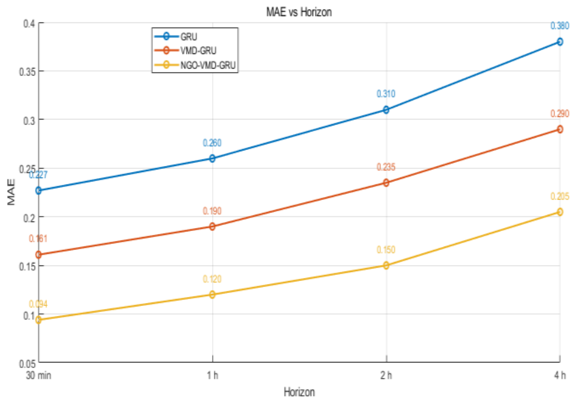

Fig 9b MAE under Multi-horizon Forecasting.

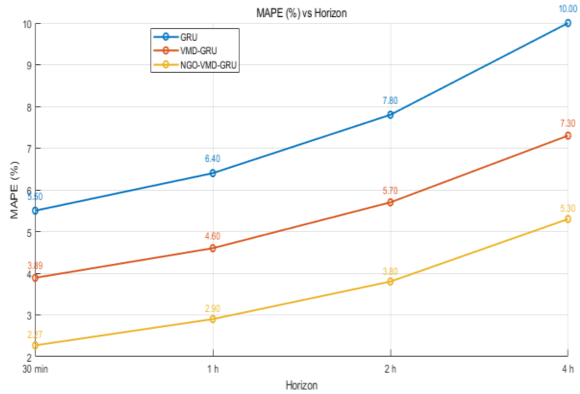

Fig 9c MAPE (%) under Multi-horizon Forecasting.

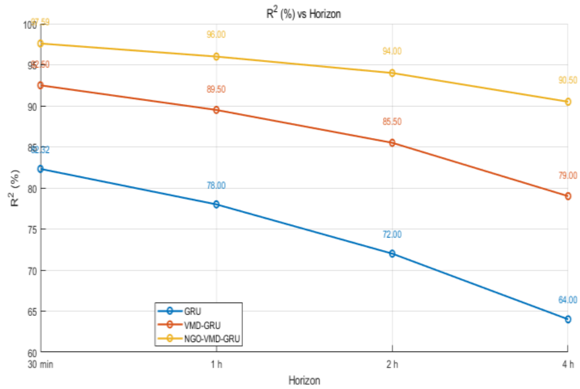

Fig 9d R² (%) under Multi-horizon Forecasting.

4.7 Ablation Experiments

An ablation study is a commonly used analytical method in predictive modeling, aimed at evaluating the contribution of individual algorithmic components within a model. By systematically removing certain elements, this approach helps clarify the role each algorithm plays in the overall prediction performance [[25]]. Accordingly, this paper compares the predictive performance of three model configurations: the GRU model, the VMD-GRU model, and the NGO-VMD-GRU model. The prediction curves of the three models are shown in Fig 10, and the corresponding evaluation metrics are presented in Fig 11.

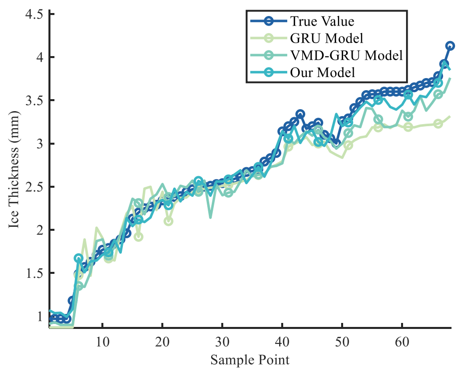

Fig 10 Comparative Analysis of Prediction Results from Different Algorithms.

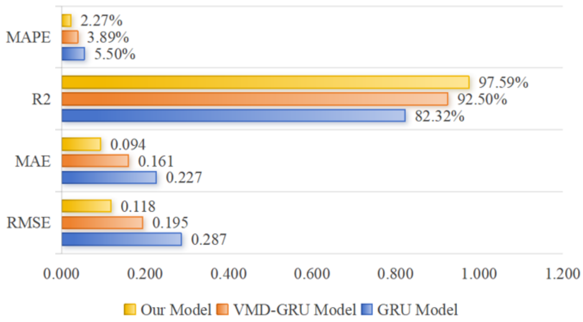

Fig 11 Prediction Performance Metrics of Different Algorithms

This study compares the prediction accuracy of the GRU model, the VMD-GRU model, and the proposed NGO-VMD-GRU model in forecasting transmission line ice thickness, thereby validating the performance differences among these models. Experimental results indicate that the GRU model achieves relatively low prediction accuracy, with an RMSE of 0.287, MAE of 0.227, R^2 of 82.32%, and MAPE of 5.50%. Although the gated mechanism of the GRU model enables it to capture long-term dependencies in time series, it still exhibits limitations in handling the non-stationary characteristics of ice thickness. In contrast, the VMD-GRU model, which incorporates the Variational Mode Decomposition (VMD) algorithm to decompose ice thickness data into multiple scales, effectively reduces data non-stationarity, achieving an RMSE of 0.195, MAE of 0.161, R^2 of 92.50%, and MAPE of 3.89%. This demonstrates the significant advantage of the VMD-GRU model in processing the complex features of ice thickness.

The proposed NGO-VMD-GRU model, however, outperforms both models across all evaluation metrics, with an RMSE of 0.118, MAE of 0.094, R^2 of 97.59%, and MAPE of 2.27%. The superiority of this model can be attributed to the following aspects: First, the VMD-based multi-scale decomposition effectively separates components of different frequencies in the original data, thereby reducing non-stationarity. Second, the Northern Goshawk Optimization (NGO) algorithm optimizes the hyperparameters of the GRU model, enhancing both its generalization capability and prediction accuracy. Finally, the GRU network serves as the core forecasting module, thoroughly capturing temporal dependencies within the ice thickness data. Experimental results confirm that the NGO-VMD-GRU model effectively overcomes the non-stationary nature of ice thickness data and achieves high prediction accuracy.

In conclusion, the NGO-VMD-GRU model demonstrates excellent performance in ice thickness forecasting. Its high accuracy and strong stability provide reliable technical support for ice warning and disaster prevention in power systems.

4.8 Cross-event Generalization

Since the dataset contains multiple cold-wave icing events, we evaluate robustness via leave-one-event-out testing: each event is held out for testing while the remaining events are used for training, and results are reported as mean ± std. As shown in Fig 12a to Fig 12d, NGO-VMD-GRU achieves the best average accuracy and the smallest standard deviation, indicating reduced sensitivity to event-to-event variations and stronger robustness under diverse scenarios.

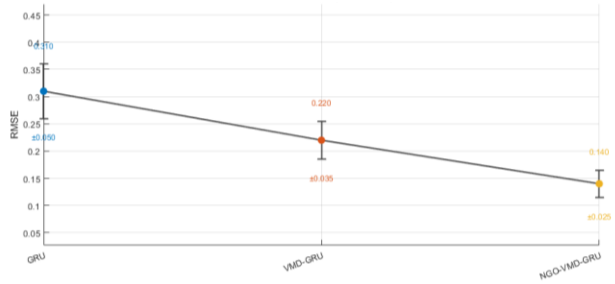

Fig 12a Cross-event RMSE (mean ± std).

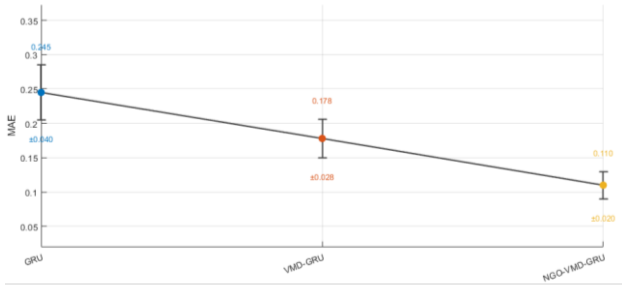

Fig 12b Cross-event MAE (mean ± std).

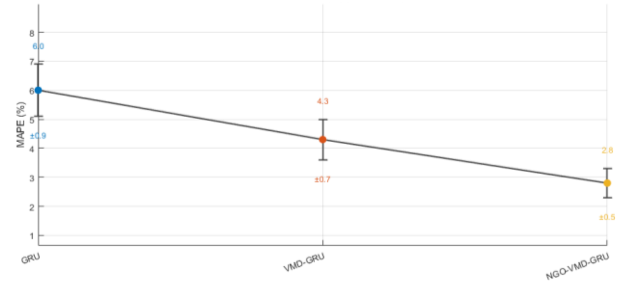

Fig 12c Cross-event MAPE (%) (mean ± std).

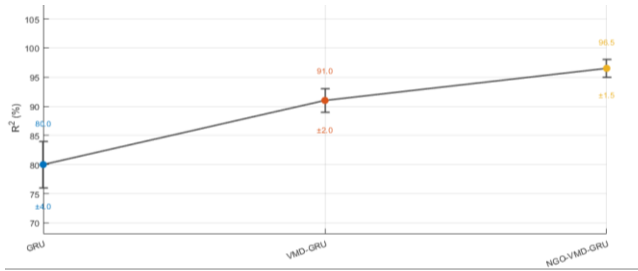

Fig 12d Cross-event R^2 (%) (mean ± std).

5. CONCLUSION

In response to the challenges of strong data non-stationarity and insufficient prediction accuracy caused by the coupling of multiple meteorological factors affecting overhead transmission line ice thickness, this paper proposes an ice thickness prediction method based on NGO-VMD-GRU. Its superior predictive performance is demonstrated through comparative and ablation experiments. The main conclusions are as follows:

1) Effectiveness of NGO in Optimizing VMD Hyperparameters

By introducing the Northern Goshawk Optimization (NGO) algorithm to adaptively optimize the mode number k and penalty factor α of VMD, the limitations of traditional VMD relying on empirical hyperparameter selection are effectively addressed. Experimental results demonstrate that the NGO algorithm converges rapidly to the global optimal solution, and the optimized VMD method improves the prediction accuracy of ice thickness by 5.09%.

2) Superiority of Component Prediction and Superposition Strategy

Each stationary ice-thickness component is independently predicted using a GRU model, and the final prediction is obtained by superimposing these component predictions. This strategy effectively mitigates the issue of poor prediction accuracy caused by the inability of conventional models to adequately capture non-stationary sequence characteristics. Experimental results show that compared to traditional models, the component-based prediction approach improves model accuracy by 28.4%, confirming the ad-vantage of the "decomposition-prediction-reconstruction" framework.

3) Experimental Validation of Model Predictive Performance

In ice thickness prediction experiments conducted on an overhead transmission line in Henan Province, the proposed NGO-VMD-GRU model achieved a Mean Absolute Per-centage Error (MAPE) of 3.12%. This represents an improvement of 17.27%, 21.45%, and 12.83% compared to the LSTM model, BP neural network, and non-optimized VMD-GRU model, respectively, demonstrating its superior performance under non-stationary data conditions.

The innovation of this study lies in the effective integration of an intelligent optimization algorithm, a data decomposition technique, and a deep learning model, offering a comprehensive solution that addresses both feature extraction and non-stationary relationship modeling for ice thickness prediction. Future research will further explore the dynamic coupling mechanisms between multiple meteorological factors and transmission line icing, and attempt to extend the proposed model to icing forecasting in challenging environments such as high-altitude and strong-wind areas.

6. ACKNOWLEDGMENTS

This study was financially supported by the Science and Technology Project of China Southern Power Grid Co., Ltd. (Project No.: CGYKJXM20240140). The funding has ensured the smooth progress of data acquisition, algorithm optimization, and experimental analysis.

7. REFERENCES

[1] Xiong W, Xu H, Xu L X, et al. RF-APJA-MKRVM Combined Prediction Model for Transmission Line Ice Coating Considering Time Cumulative Effects[J]. High Voltage Engineering, 2022(3): 948-957.

[2] Tu J L, Wang H L. IAOA-LSSVM-Based Prediction of Ice Thickness on Transmission Lines[J]. Modern Electronic Technology, 2024(21): 113-118.

[3] Jiang X L, Chang H, Hu Q, et al. Prediction and Experimental Study on Equivalent Ice Thickness of Comprehensive Load for Transmission Lines[J]. Proceedings of the CSEE, 2013(10): 177-183.

[4] Liu H W, Lu J Z, Lai X Y, et al. Short-term Multivariate Grey Prediction Model for Transmission Line Ice Thickness[J]. High Voltage Engineering, 2015(10): 3372-3377.

[5] Wang X T, Ding J J, Zhang F, et al. Improved BP Neural Network-Based Ice Prediction Technology for Transmission Lines[J]. Mechanical Design and Manufacturing, 2024(9): 306-310.

[6] Chen Y, Li P, Zhang Z J, et al. PCA-GA-LSSVM-Based Online Prediction Model for Transmission Line Ice Load[J]. Power System Protection and Control, 2019(10): 110-119.

[7] Ma T, Niu D. Icing Forecasting of High Voltage Transmission Line Using Weighted Least Square Support Vector Machine with Fireworks Algorithm for Feature Selection[J]. Applied Sciences, 2016, 6(12): 438.

[8] Yu T, Li Y N. Physics-Guided SSA-BiGRU Prediction Model for Ice Thickness on Transmission Lines[J]. Electric Power Science and Engineering, 2022(2): 28-36.

[9] Chen B, Xu Z M, Jia Y F, et al. VMD-SSA-LSTM-Based Prediction Model for Ice Coating on Overhead Transmission Conductors[J]. Journal of China Three Gorges University (Natural Science Edition), 2024(4): 105-112.

[10] Luo C, Fan L D, Zhao X W, et al. Study on VMD-IGWO-LSSVM-Based Ice Prediction Model[J]. Power System and Clean Energy, 2021(6): 9-17.

[11] Yu K, Zhu X, Cao W. Study on Traveling Wave Fault Localization of Transmission Line Based on NGO-VMD Algorithm[J]. Energies, 2024, 17(9).

[12] Wang C Q, Wu L W, Deng Z B, et al. Review of Time Cumulative Overhead Transmission Line Ice Prediction Models and Algorithms[J]. China Electric Power, 2024(6): 153-164.

[13] Li X Y, Zhou Q, Zhang H X, et al. Mid-Long-Term Power Load Forecasting of Building Group Based on Modified NGO[J]. Energies, 2025, 18(3): 668.

[14] He M, Wang H, Thwin M. A machine learning technique for optimizing load demand prediction within air conditioning systems utilizing GRU/IASO model[J]. Scientific Reports, 2025, 15(1): 3353.

[15] Ciechulski, T.; Osowski, S. Wind Power Short-Term Time-Series Prediction Using an Ensemble of Neural Networks. Energies 2024, 17, 264. https://doi.org/10.3390/en17010264.

[16] Tryhuba, I.; Tryhuba, A.; Hutsol, T.; Cieszewska, A.; Andrushkiv, O.; Glowacki, S.; Bryś, A.; Slobodian, S.; Tulej, W.; Sojak, M. Prediction of Biogas Production Volumes from Household Organic Waste Based on Machine Learning. Energies 2024, 17, 1786. https://doi.org/10.3390/en17071786.

[17] Liu J, Cong L M, Xia Y Y, et al. Combined Model of DBO-VMD and IWOA-BILSTM Neural Networks for Short-term Power Load Prediction[J]. Power System Protection and Control, 2024(8): 123-133.

[18] Yang B, Duan J H, Li M W, et al. Optimization of Offshore Hybrid Photovoltaic-Wave Energy Converter Arrays Based on Improved Bald Eagle Search Algorithm[J]. Power System Technology, 2024, 48(6): 2480-2489.

[19] Jiang F, Lin Z Y, Wang W Y, et al. Optimal Bagging Ensemble for Ultra-short-term Multivariate Load Prediction Considering Minimum Average Envelope Entropy Load Decomposition[J]. Proceedings of the CSEE, 2024, 44(5): 1777-1789.

[20] Abumohsen M, Owda A Y, Owda M. Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms[J]. Energies, 2023, 16(5): 2283.

[21] Liu C, He Q H, Lu Y J, et al. PSOEM-LSSVM Ice Prediction Model for Transmission Lines[J]. Journal of Electric Power Science and Technology, 2020(6): 131-137.

[22] Fan X, Wang R, Yang Y, et al. Transformer–BiLSTM Fusion Neural Network for Short-Term PV Output Prediction Based on NRBO Algorithm and VMD[J]. Applied Sciences, 2024, 14(24): 11991.

[23] Tao K, Xu M, Wang Q. ASSA-VMD-SI and Frechet method of pipe vibration for noise reduction and leakage identification[J]. Measurement, 2025, 242: 116277.

[24] Dou X, Quan X, Wu Z, Hu M, Yang K, Yuan J, Wang M. Hybrid Multi-Agent Control in Microgrids: Framework, Models and Implementations Based on IEC 61850. Energies. 2015; 8(1):31-58. https://doi.org/10.3390/en8010031.

[25] Jin X Z, Zhao S S, Chang H, et al. Modeling of Superheated Steam Temperature Characteristics Based on Golden Jackal Optimization Variational Mode Decomposition and Temporal Convolutional Network[J]. Proceedings of the CSEE. [2025-01-03].