infrastructures

MDPI

Article

Tree-Based Surrogate Model for Predicting Aerodynamic Coefficients of Iced Transmission Conductor Lines

Guoliang Ye [1], Zhiguo Li [1], Anjun Wang [1], Zhiyi Liu [2], Ruomei Tang [2] and Guizao Huang [3,*]

[1] China Southern Power Grid Co., Ltd., Kunming 650032, China
[2] Powerchina Chengdu Electric Power Fittings Co., Ltd., Chengdu 610094, China
[3] College of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China
* Correspondence: gzhuang@swjtu.edu.cn

Abstract

Ultra-high-voltage (UHV) transmission lines are prone to galloping and oscillations under ice and wind loads, posing risks to system reliability and safety. Accurate aerodynamic coefficients are essential for evaluating these effects, but conventional wind tunnel and CFD methods are costly and inefficient for practical applications. To address these challenges, this study develops a surrogate model for rapid and accurate prediction of aerodynamic coefficients for six-bundle conductors. Initially, a CFD model to calculate the aerodynamic coefficients of six-bundle conductors was proposed and validated against wind tunnel experimental results. Subsequently, Latin hypercube sampling (LHS) was employed to generate datasets covering wind speed, icing shape, icing thickness, and wind attack angle. High-throughput numerical simulations established a comprehensive aerodynamic database used to train and validate multiple tree-based surrogate models, including decision tree (DT), random forest (RF), extremely randomized trees (ERTs), gradient boosted decision tree (GBDT), and extreme gradient boosting (XGBoost). Comparative analysis revealed that the XGBoost-based model achieved the highest prediction accuracy, with an R^2 of 0.855 and superior generalization performance. Feature importance analysis further highlighted wind speed and icing shape as the dominant influencing factors. The results confirmed the XGBoost surrogate as the most effective among the tested models, providing a fast and reliable tool for aerodynamic prediction, vibration risk assessment, and structural optimization in UHV transmission systems.

Keywords: transmission lines; ice coating; aerodynamic coefficients; machine learning; surrogate model

Academic Editor: Isam Shahrour

Received: 11 August 2025
Revised: 3 September 2025
Accepted: 9 September 2025
Published: 15 September 2025

Citation: Ye, G.; Li, Z.; Wang, A.; Liu, Z.; Tang, R.; Huang, G. Tree-Based Surrogate Model for Predicting Aerodynamic Coefficients of Iced Transmission Conductor Lines. *Infrastructures* **2025**, *10*, 243. https:// doi.org/10.3390/ infrastructures10090243

1. Introduction

Transmission lines are critical infrastructures within power grids, ensuring stable and continuous electrical supply essential for societal and economic stability. Under extreme environmental conditions involving strong winds and ice accretion, transmission lines usually employing bundled conductors are highly susceptible to pronounced wind-induced vibrations. Such phenomena typically manifest as galloping or sub-span oscillations, which can significantly compromise operational stability and structural integrity [1]. These dynamic instabilities significantly degrade the reliability and safety of electrical systems, often leading to fatigue damage of hardware components, insulator failures, and even widespread power outages, thus posing substantial operational and economic challenges [2,3]. To mitigate these problems, a thorough understanding of the aerodynamic behavior of bundled conductors under diverse weather conditions is essential [4].

Wind-induced vibrations constitute a major threat to the safe operation of transmission lines and are a direct source of economic losses through accelerated component fatigue, increased maintenance interventions, and potential power interruptions. Crucially, the aerodynamic coefficients of bundled or iced conductors form the load foundation for vibration analyses, such as spanning galloping, sub-span oscillations, and aeolian vibration, and their accuracy is a necessary condition for reliable force estimation, stability assessment, and response prediction. Improved fidelity in these coefficients helps reduce lifecycle costs and outage risk, while supporting utilities in meeting evolving expectations for grid resilience under extreme weather.

Traditionally, the aerodynamic behavior of conductors has been extensively investigated using wind tunnel experiments [5–9] and computational fluid dynamics (CFD) simulations [10–12]. Wind tunnel tests provide valuable direct observations but are typically costly, time-consuming, and limited to simplified scenarios [13]. CFD simulations, meanwhile, offer detailed insights into aerodynamic characteristics under various conditions, such as wind speed, attack angle, icing shape, and conductor configurations [10]. For instance, previous studies have systematically examined how icing conditions influence aerodynamic coefficients of bundled conductors, offering insights into the complex interactions between airflow and conductor geometry [14]. Although significant progress has been achieved through these conventional approaches, limitations remain. CFD simulations, in particular, face inherent drawbacks, such as high computational demands, sensitivity to modeling assumptions, and difficulties in providing real-time predictive capabilities required for efficient operational decision-making [15]. Moreover, recent wake-interference studies [9,16] have demonstrated that torsional velocity effects and three-dimensional wake interactions among sub-conductors play a crucial role in large-amplitude galloping, thereby challenging the adequacy of simplified 2D turbulence models, such as the Spalart–Allmaras (S-A) formulation. These findings highlight the need to carefully consider the physical limits of CFD modeling when applied to bundled or iced conductors.

To overcome these challenges, recent studies have increasingly turned toward machine learning (ML)-based methods, leveraging their ability to model complex nonlinear relationships efficiently [17–19]. ML techniques have shown promising results in rapidly predicting aerodynamic characteristics and related dynamic behaviors in transmission lines [20–22]. For example, artificial neural networks and tree-based algorithms have been successfully employed to predict conductor aerodynamic coefficients and dynamic responses under icing and wind-loading scenarios [23]. Nevertheless, performance among different ML techniques varies considerably, and selecting the most suitable modeling approach remains a critical yet controversial issue in practical engineering applications [21].

Recently, machine learning has been applied to the aerodynamic analysis of iced conductors as an efficient alternative to wind tunnel tests. A BP neural network model was used to predict aerodynamic coefficients of iced conductors, showing good agreement with experimental results and confirming its feasibility for galloping studies [24]. In addition, a convolutional neural network (CNN) approach based on composite images was proposed, which achieved high accuracy in predicting drag, lift, and torque coefficients, demonstrating clear advantages over traditional methods [25]. These studies indicate that data-driven approaches for predicting aerodynamic coefficients of iced conductors have become a research hotspot.

This study is structured as follows: Section 2 presents the establishment of a high-fidelity numerical flow field model for six-bundle conductors, and its accuracy is validated against wind tunnel experimental data. The generation of a large-scale aerodynamic coefficient dataset using Latin hypercube sampling, together with the development and optimization of multiple tree-based surrogate models through hyper-parameter tuning

and cross-validation, is described in Section 3. In Section 4, comparative performance evaluations are provided, and a global sensitivity analysis is carried out. Potential engineering applications of the proposed surrogate model are then discussed in Section 5. Finally, Section 6 summarizes the conclusions and outlines future research directions.

2. Numerical Model of Flow Field for Bundled Conductors

This section presents the development and validation of a numerical model for analyzing the flow field around bundled transmission conductors under various wind and icing conditions. The aim is to accurately capture the complex aerodynamic behavior and force distributions experienced by multi-bundle conductors subjected to different environmental parameters. The modeling approach combines computational fluid dynamics (CFD) simulations with systematic parameterization of wind attack angles, icing shapes, and other key factors. The results of this section provide a high-fidelity aerodynamic database, which serves as the foundation for subsequent surrogate model development and data-driven aerodynamic predictions.

2.1. Establishment of the Numerical Model

UHV transmission lines typically employ multi-bundle conductor configurations to meet the requirements for high-capacity, long-distance power transmission. This study focuses on a typical six-bundle conductor type, JL/LB1A-720/50, widely adopted in UHV transmission lines. The detailed material and structural parameters of the conductor are presented in Table 1, with a spacing of 450 mm between individual sub-conductors.

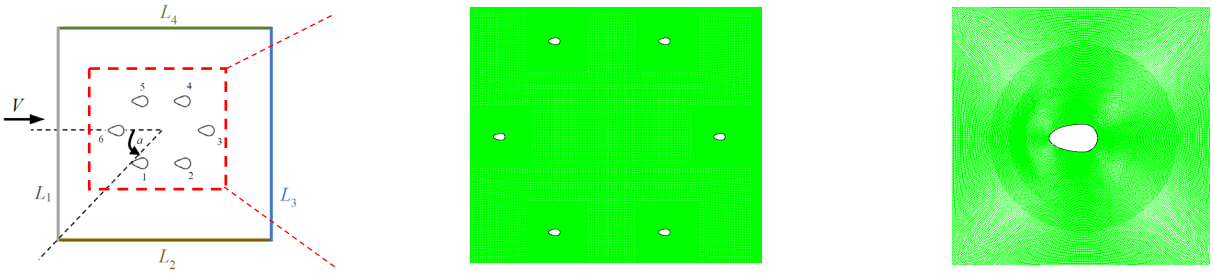
Table 1. Material parameters of conductor lines.

Diameter (mm)	Cross-Section Area (mm^2)	Mass per Unit (kg/km)	Young's Modulus (GPa)
36.20	775.00	2337.90	63.7

To accurately simulate the aerodynamic characteristics, a numerical flow field model was established using the finite-volume method by means of the method in Ref. [10]. The computational domain was designed as a square region of 10 m × 10 m, i.e., $L_1 = L_2 = L_3 = L_4 = 10$ m, to ensure numerical stability and sufficient accuracy, shown in Figure 1. The layout of sub-conductors and boundary conditions is presented in Figure 1a. Wind attack angles ranging from 0° to 180° were analyzed to fully evaluate aerodynamic behavior under various wind directions, with simulations conducted at increments of 5°. To enhance computational efficiency and consistency, the geometry and mesh structure remained unchanged across simulations, while different wind attack angles were implemented by adjusting inflow directions and boundary conditions. It is noted that this study analyzes section-level, quasi-steady samples and, therefore, does not explicitly encode spatial correlations of icing geometry (e.g., along-span variability) nor temporal dependencies in wind loading.

The boundary conditions were defined according to the wind attack angle α, with the left, lower, right, and upper boundaries denoted as L_1, L_2, L_3, and L_4, respectively. The setup principle is as follows: when $\alpha = 0°$, L_1 is specified as the velocity inlet (Inlet), L_3 as the pressure outlet (Outlet), and L_2 and L_4 as symmetry boundaries (Sym). For $0° < \alpha < 90°$, both L_1 and L_2 are designated as Inlets, while L_3 and L_4 act as Outlets. When $45° < \alpha < 135°$, L_1 and L_2 are set as Inlets, and L_3 and L_4 as outlets. At $\alpha = 90°$, L_2 serves as the velocity inlet, L_4 as the pressure outlet, while L_1 and L_3 are imposed as Sym. For $90° < \alpha < 180°$, L_3 and L_2 are used as Inlets, with L_1 and L_4 defined as Outlets. Finally, when $\alpha = 180°$, L_3 is assigned as the Inlet, L_1 as the Outlet, and L_2 and L_4 as Sym. The

detailed boundary condition settings for different wind attack angles are summarized in Table 2.



(**a**) Boundary conditions (**b**) Grid of six-bundle conductors (**c**) Grid of sub-conductor

Figure 1. Flow field model of six-bundle conductor lines.

Table 2. Boundary conditions of the CFD model.

Wind Attack Angle (α)	L_1 (Left)	L_2 (Bottom)	L_3 (Right)	L_4 (Top)
0°	Inlet	Sym	Outlet	Sym
0° < α < 90°	Inlet	Inlet	Outlet	Outlet
45° < α < 135°	Inlet	Inlet	Outlet	Outlet
90°	Sym	Inlet	Sym	Outlet
90° < α < 180°	Outlet	Inlet	Inlet	Outlet
180°	Outlet	Sym	Inlet	Sym

To improve the accuracy of the simulation results while maintaining computational efficiency, a refined mesh was applied in the regions surrounding the sub-conductors, whereas coarser grids were used in the areas farther away, as illustrated in Figure 1b,c. The computational domain was set to 10 m × 10 m, which is about 22 times the cross-sectional size of the six-bundle conductor (450 mm × 450 mm), ensuring that the wake effects dissipate adequately before reaching the domain boundaries. This choice is consistent with previous aerodynamic modeling studies of iced or bundled conductors, where the computational domain was typically selected as 20–25 times the conductor cross-section to minimize boundary influence [10]. The domain was discretized with quadrilateral elements, and several mesh schemes were tested by gradually increasing the number of cells to check mesh convergence. Finally, a mesh with 42,568 cells was demonstrated to be sufficient for stable aerodynamic coefficient predictions. The mesh around each iced sub-conductor was refined to properly resolve the boundary layer, while the outer domain was meshed more coarsely to reduce computational costs.

For turbulence modeling, the S-A one-equation turbulence model was employed due to its robustness, computational efficiency, and capability to accurately capture vortex shedding and boundary layer separation phenomena around conductors at high Reynolds numbers. Pressure–velocity coupling utilized the SIMPLEC algorithm, which accelerates convergence through iterative corrections of the pressure field, effectively reducing mesh dependency [10]. A second-order implicit scheme was chosen for temporal discretization to ensure numerical stability and suppress numerical oscillations. Spatial discretization of the pressure terms adopted a second-order upwind scheme to smooth gradients, while the QUICK scheme was selected for momentum equations to minimize numerical dissipation and improve convection term accuracy. Turbulent transport equations employed third-order discretization schemes to enhance the resolution of turbulent kinetic energy transport. This numerical setup ensures a balanced combination of computational efficiency and accuracy, providing a solid foundation for subsequent aerodynamic analyses.

2.2. Computation of Aerodynamic Coefficients

The aerodynamic forces on conductors, including lift, drag, and torque, constitute key parameters governing their dynamic stability under wind loads, influencing phenomena such as galloping and sub-span oscillations. These aerodynamic forces can be expressed mathematically by:

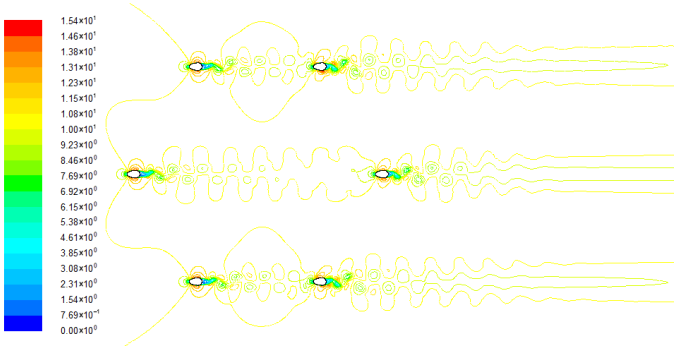
$$C_D = \frac{F_D}{\frac{1}{2}\rho U^2 L d}, \tag{1}$$

$$C_L = \frac{F_L}{\frac{1}{2}\rho U^2 L d}, \tag{2}$$

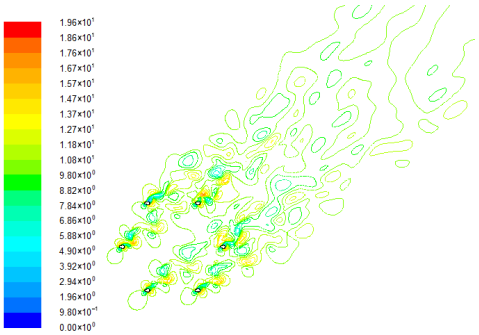
$$C_M = \frac{M}{\frac{1}{2}\rho U^2 L d}, \tag{3}$$

where C_D, C_L, and C_M represent drag, lift, and torque, respectively, ρ is air density, U denotes wind speed, L is the effective length of the conductor model, and d represents conductor diameter. The aerodynamic coefficients C_D, C_L, and C_M are dimensionless and dependent on conductor geometry and environmental conditions.

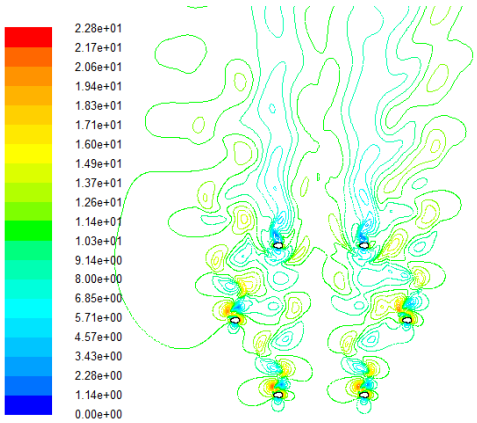
The simulated velocity flow fields at various wind attack angles (0°, 45°, and 90°) are illustrated in Figure 2, in which the white color presents the sub-conductors. At 0°, airflow patterns are symmetric, and flow remains relatively attached, generating stable wake structures downstream. As the wind attack angle increases to 45°, the flow field exhibits noticeable asymmetry, intensified turbulence, and complex vortex shedding around conductors. At 90°, clear and pronounced wake separation is observed, characterized by periodic vortex shedding downstream. These distinct flow characteristics significantly influence the temporal behavior of aerodynamic loading acting upon the conductors.



(**a**) 0°



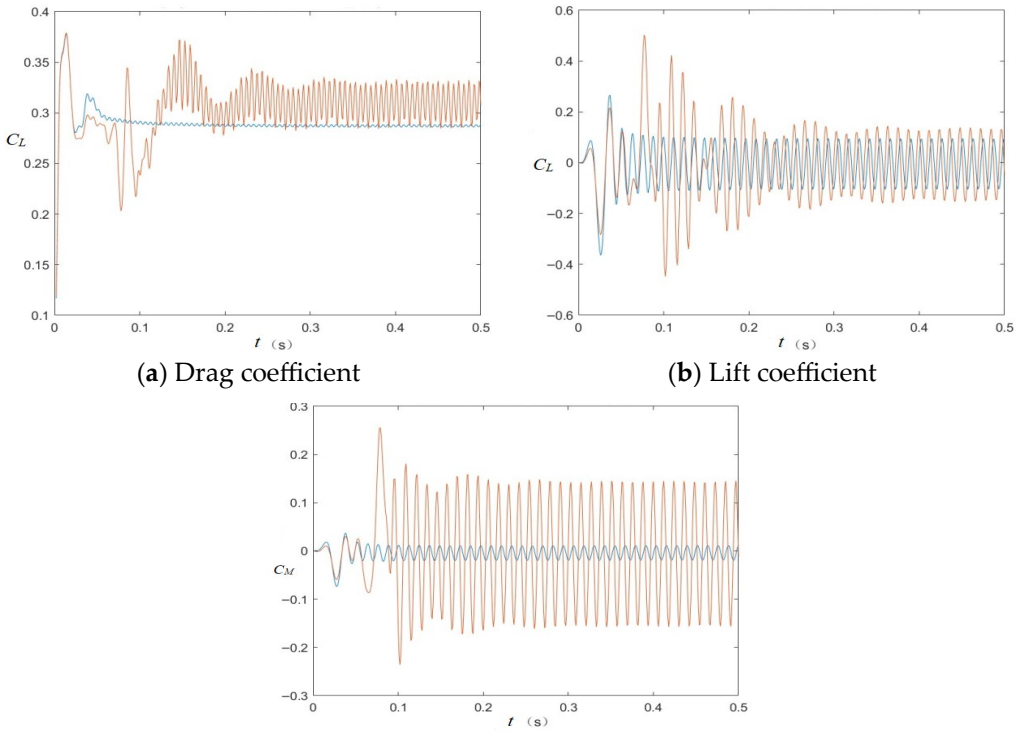
(**b**) 45°



(**c**) 90°

Figure 2. Velocity flow fields at different wind attack angles.

To further quantify these dynamic effects, time series curves of aerodynamic coefficients (drag, lift, and moment) at a 0° wind attack angle are presented in Figure 3. Here, sub-conductor 1 (windward side, blue line) demonstrates relatively stable aerodynamic responses due to direct exposure to incoming airflow. In contrast, sub-conductor 2 (leeward side, orange line), positioned behind sub-conductor 1, exhibits more pronounced oscillations caused by wake-induced turbulence and flow instability. This clear difference highlights the substantial influence of wake interference on aerodynamic performance and underscores the importance of accurately capturing conductor arrangement effects in predictive modeling.



(**a**) Drag coefficient (**b**) Lift coefficient

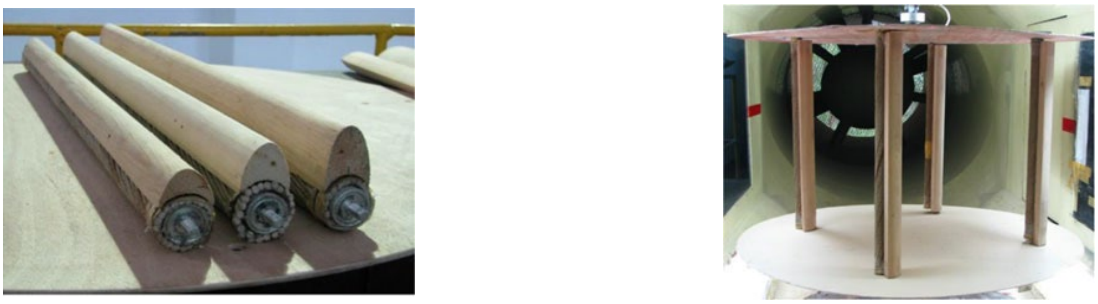
(**c**) Moment coefficient

Figure 3. Time series curve of aerodynamic coefficients at a 0° wind attack angle.

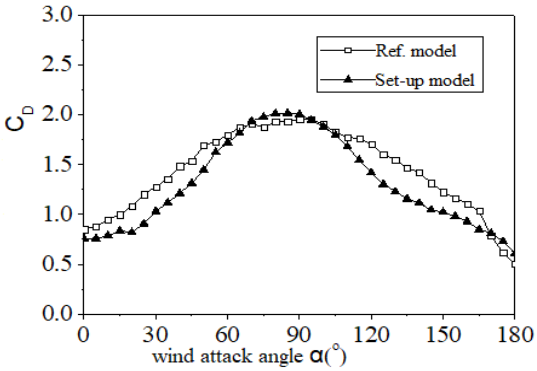
2.3. Verification of the Proposed Model

To ensure the accuracy and reliability of the established numerical model, a comprehensive comparative validation was conducted using published wind tunnel experimental data for bundled conductors. In this study, the aerodynamic performance of a crescent-shaped iced eight-bundle conductor, as reported in Ref. [10], was used as the benchmark for model verification. The validation setup involved an ice thickness of 12 mm, a wind speed of 10 m/s, and wind attack angles ranging from 0° to 180° in 5° increments, replicating the experimental conditions as closely as possible.

The physical arrangement of the wind tunnel test is presented in Figure 4a, where the iced conductor models were mounted in a controlled wind tunnel environment to directly measure aerodynamic forces. Figure 4b presents a direct comparison between the drag coefficient C_D obtained by the numerical model (setup model) and those obtained from the wind tunnel experiments (reference model) across the full range of wind attack angles. The numerical results show good agreement with experimental data, successfully capturing the overall trends, peak values, and characteristic response variations of the drag coefficient as a function of wind attack angle.



(**a**) Wind tunnel test (Cai et al., 2015)



(**b**) Aerodynamic coefficients obtained by the setup model and wind tunnel test (Cai et al., 2015)

Figure 4. Comparative verification of the setup model [10].

While minor discrepancies exist, especially near the peaks, these differences can largely be attributed to inevitable simplifications in the numerical model (such as idealized boundary conditions and mesh resolution) and experimental uncertainties. Importantly, the model faithfully reflects the variation and sensitivity of aerodynamic coefficients with respect to the wind attack angle, thereby confirming its robustness and predictive reliability. This strong consistency between simulation and experiment validates the effectiveness of the developed model and provides a solid foundation for further aerodynamic analyses and the construction of accurate surrogate models for bundled conductors under complex environmental conditions.

3. Establishment of the Prediction Model

3.1. Acquisition of Large-Scale Data Samples

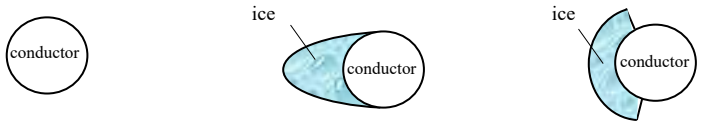
To develop a surrogate model with both high accuracy and strong generalization capability, it is essential to obtain a dataset that uniformly and comprehensively covers the high-dimensional input space. In this study, Latin hypercube sampling (LHS) was employed as the core experimental design strategy. LHS is a stratified, quasi-random sampling technique widely recognized for its efficiency in multi-dimensional parameter studies and its advantage over simple random sampling in achieving good space-filling properties [18].

The basic principle of LHS can be described as follows: Suppose the input parameter space has n_v dimensions and a total sample size of N. For each dimension, the range is divided into N equally probable intervals. A single value is randomly selected from each interval, ensuring that each interval is sampled exactly once per dimension. These values are then randomly paired across all dimensions to form N distinct sample points. Mathematically, for the j-th dimension and the i-th sample, the sample value x_{ij} can be expressed as:

$$x_{ij} = F_j^{-1}\left(\frac{P_{ij} - 1 + u_{ij}}{N}\right), \quad (4)$$

where F_j^{-1} is the inverse cumulative distribution function of the j-th variable, P_{ij} is the i-th permuted interval for dimension j, and u_{ij} is a random number uniformly distributed in [0, 1).

In this research, five key factors affecting the aerodynamic coefficients of six-bundle conductors were selected as input parameters: wind speed, wind attack angle, icing shape, icing thickness, and sub-conductor number. To fully capture engineering scenarios, icing shape (S) included three typical cases: bare conductor, crescent-shaped ice, and sector-shaped ice, which are shown in Figure 5. It is noted that the icing angle of sector-shaped ice is set as 120°. The specific value ranges for each parameter are provided in Table 3. Specifically, icing thickness (T) varies from 0 mm to 50 mm, incremented by 1 mm, reflecting both light and severe icing events. Wind speed (V) ranges from 0 to 30 m/s, encompassing calm conditions up to severe storms. Wind attack angle (a) is considered from 0° to 180° at 5° intervals, enabling the exploration of a wide variety of wind directions relative to the conductor axis. Finally, sub-conductor number (m) covers all six bundle positions to account for possible spatial variability within the bundle structure.



(**a**) Bare conductor (**b**) Crescent-shaped (**c**) Sector-shaped

Figure 5. Typical ice shapes.

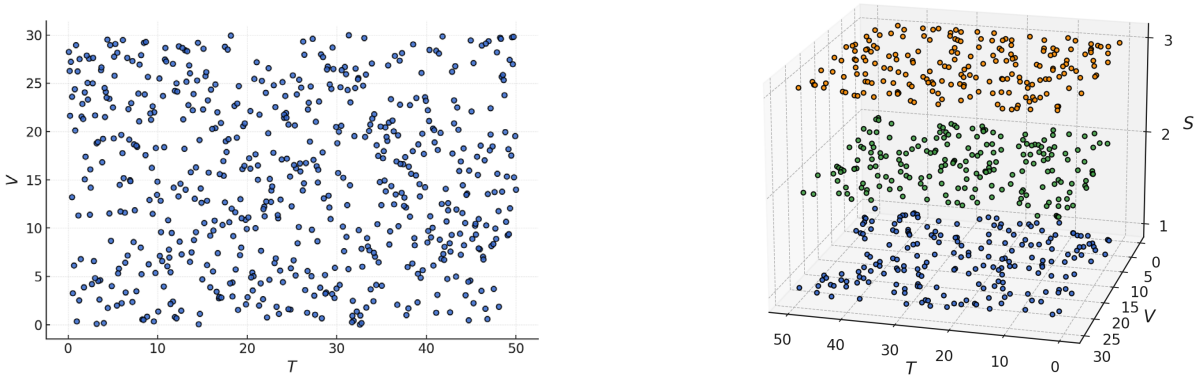
Table 3. Key parameters of aerodynamic characteristics and their value ranges for bundle conductors.

Factor	Label	Unit	Range
Icing shape	S	-	bare, crescent-shaped, sector-shaped
Icing thickness	T	mm	[0, 50, 1] *
Wind speed	V	m/s	[0, 30, 1]
Wind attack angle	a	°	[0, 180, 5]
Sub-conductor number	m	-	1, 2, 3, 4, 5, 6

* [0, 50, 1] represent the range from 0 to 50 with the interval 1.

Based on the Latin hypercube sampling (LHS) strategy, a total of 3580 sample combinations were generated to uniformly explore the five-dimensional design space defined by icing shape, icing thickness, wind speed, wind attack angle, and sub-conductor number. Each sample point represents a unique set of parameter values, and these combinations were used as input cases for high-throughput CFD simulations with the validated model described in Section 2.

The effectiveness of the LHS method in generating well-distributed samples is clearly illustrated in Figure 6. As shown in Figure 6a, the projection of the sample distribution onto the two-dimensional plane of icing thickness (T) and wind speed (V) reveals a highly uniform coverage, with no clustering or significant gaps, ensuring that all regions of the parameter space are adequately represented. Figure 6b extends this visualization to three dimensions among icing thickness (T), wind speed (V), and icing shape (S), further confirming that the sampling strategy achieves excellent space-filling properties across both continuous and discrete variables. The stratified, layered pattern in the third dimension reflects the inclusion of all three typical icing shapes, with samples in each layer uniformly spread over the full range of T and V.



(**a**) Two dimensions between T and V (**b**) Three dimensions among T, V, and S

Figure 6. Illustration of distribution of the sample data.

This comprehensive and well-balanced dataset not only enhances the representativeness and diversity of training and testing samples, but also provides a robust foundation for developing, evaluating, and generalizing the tree-based machine learning models proposed in this study.

3.2. Tree-Based Machine Learning Methods

Tree-based machine learning algorithms offer significant advantages for regression and classification tasks involving complex, nonlinear relationships and heterogeneous data. Their ability to capture feature interactions, robustness to outliers, and interpretability make them particularly suitable for modeling aerodynamic coefficients, which are influenced by multiple coupled factors, such as wind speed, attack angle, and icing conditions. Given our 3580-sample tabular dataset, ensemble tree methods offer strong sample efficiency and stable generalization with modest tuning, whereas neural network alternatives typically require larger datasets and more elaborate regularization to achieve comparable performance. In this study, several representative tree-based regression algorithms were considered, including DT, RF, ERT, GBDT, and XGBoost.

Among these, XGBoost (extreme gradient boosting) stands out for its superior predictive accuracy and computational efficiency. XGBoost is an advanced ensemble learning algorithm built on the framework of gradient boosting decision trees (GBDTs). Its core concept is to construct a series of weak regression trees in an iterative fashion, each learning to correct the residuals of its predecessors. The final prediction is a weighted sum of all trees. Compared to traditional GBDT, XGBoost introduces regularization, second-order gradient optimization, and parallel computation, which together improve both the accuracy and generalization capability of the model—especially when dealing with structured datasets and high-dimensional regression problems.

In the context of aerodynamic coefficient prediction for bundle conductors, XGBoost's segmented tree-based structure effectively captures the strong nonlinearity and feature interactions inherent in the data, while its built-in mechanisms provide resilience to noise and missing values. The objective function of XGBoost can be formulated as:

$$\Gamma = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{i=1}^{K} \Omega(f_k), \tag{5}$$

where $L(y_i, \hat{y}_i)$ denotes the loss function (e.g., mean squared error), and $\Omega(f_k) = \gamma T + 0.5\lambda\|\omega\|^2$ is the regularization term, with T as the number of leaves, ω as the leaf weights, and γ and λ as regularization parameters.

During tree construction, the optimal feature split is determined by maximizing the gain function:

$$\text{Gain} = \frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda} - \gamma, \tag{6}$$

in which $g_i = \partial_{\hat{y}^{(t-1)}} L(y_i, \hat{y}^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} L(y_i, \hat{y}^{(t-1)})$ represent the first and second derivatives (gradients and Hessians) of the loss function for the left and right child nodes, respectively.

The performance of XGBoost is highly dependent on hyper-parameter tuning. Key parameters include the number of trees (n_estimators), learning rate (learning_rate), maximum tree depth (max_depth), subsample ratio (subsample), and minimum child weight (min_child_weight). These parameters jointly control model complexity, learning efficiency, and the ability to generalize, and are typically optimized through grid search and cross-validation.

3.3. Model Training and Testing

Before developing the surrogate prediction model, data preprocessing is necessary to mitigate the effects of varying scales, dimensions, and outliers, thereby ensuring stability and accuracy during training. In this study, Min–Max scaling was utilized to normalize all five-dimensional input parameters and three-dimensional output variables, linearly mapping each feature to the range [0, 1] as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{7}$$

where x is the original value, and x' is the normalized value.

The complete process for establishing the aerodynamic prediction model for six-bundle conductors is comprehensively illustrated in Figure 7. After data preprocessing, the dataset was randomly divided into training and testing subsets at a ratio of 4:1, ensuring that the model's evaluation would be based on previously unseen data and thus provide a reliable assessment of its generalization capability. The training set (80%) was utilized to construct and tune the tree-based surrogate models, with a focus on the XGBoost algorithm due to its superior performance in preliminary comparisons.

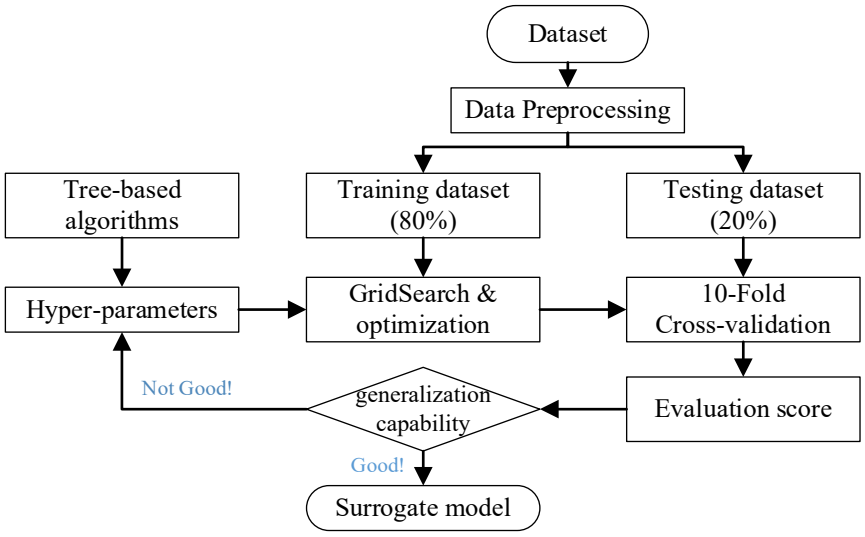


Figure 7. Process illustration of the setup of tree-based surrogate models.

A grid search strategy, implemented via the GridSearchCV module from Scikit-learn, was employed to systematically optimize key hyper-parameters. This process was combined with 10-fold cross-validation, allowing the model to be trained and validated on multiple splits of the training data. Through this approach, issues such as overfitting and underfitting could be effectively identified and mitigated, resulting in robust model performance across varying data partitions. The primary hyper-parameters subjected to tuning included the learning rate, maximum tree depth, subsample ratio, minimum child weight, and regularization terms, all of which play a critical role in balancing model complexity and predictive accuracy. The detailed search ranges and the final selected optimal values for these hyper-parameters are provided in Table 4. The convergence behavior of the XGBoost-based surrogate model during training is illustrated in Figure 8. As the epoch number increases, the cost function (loss) decreases rapidly in the initial stages and gradually stabilizes, indicating effective learning and optimization of the model parameters. This steady decline in loss demonstrates the model's ability to fit the training data efficiently without signs of overfitting, further validating the appropriateness of the selected hyper-parameters and the overall robustness of the training procedure.

Table 4. Model hyper-parameter adjustment process based on the XGBoost algorithm.

Hyper-Parameters	Range	Best Hyper-Parameter
n_estimators	50~1000	300
learning_rate	0.01~0.1	0.05
max_depth	2~8	6
subsample	0.8~1.0	0.9
min_child_weight	0.8~1.0	0.88

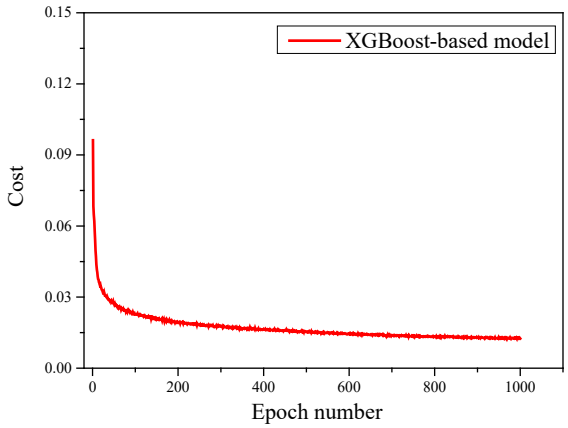


Figure 8. Training process of the XGBoost-based surrogate model.

Following optimization, the best-performing surrogate model was further evaluated using the independent testing dataset (20%) to assess its predictive accuracy, stability, and generalization capacity. The entire modeling workflow, including dataset partitioning, hyper-parameter tuning, model training, cross-validation, and performance evaluation, is clearly visualized in Figure 7, offering a transparent and reproducible process for surrogate model development in aerodynamic prediction tasks.

It should be noted that the same modeling and hyper-parameter optimization procedure described above was also applied to all other tree-based algorithms evaluated in this study, including DT, RF, ERT, and GBDT. For each algorithm, an independent grid search combined with 10-fold cross-validation was performed using the training dataset to identify the optimal hyper-parameter settings, ensuring a fair and rigorous comparison among different models. Due to space limitations, the specific optimal parameters for

these additional models are not listed in detail in this paper, but the overall workflow and optimization principles are consistent with those outlined for the XGBoost-based model.

After selecting the optimal hyper-parameters, the model was retrained on the entire training dataset, applying regularization constraints and feature weighting to further enhance its predictive accuracy and robustness. Finally, the generalization capability of the trained model was quantitatively assessed using the independent test dataset. Three widely used regression evaluation metrics were employed: coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE), calculated as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i^2)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i^2)^2}, \tag{8}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i^2)^2, \tag{9}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{10}$$

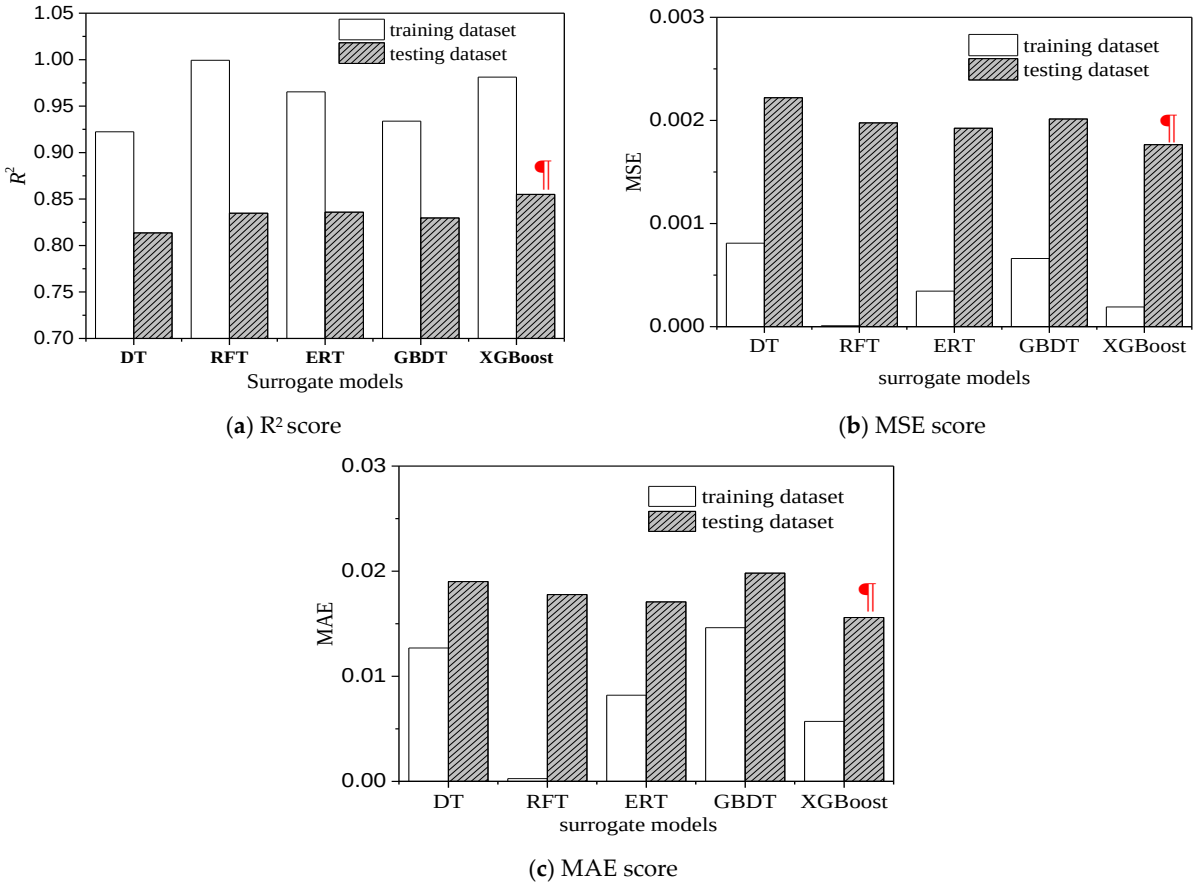
where n is the number of test samples, y_i and $\hat{y}_i$ are the observed and predicted aerodynamic coefficients, respectively, and $\overline{y}$ is the mean of observed values. An R^2 value approaching 1 indicates high predictive accuracy, whereas lower values of MSE and MAE reflect improved model precision. These metrics collectively provide a comprehensive basis for evaluating model accuracy, stability, and generalization performance.

4. Discussion

To determine the optimal tree-based surrogate model for accurately predicting aerodynamic coefficients of six-bundle conductors, five representative tree-based algorithms—decision tree (DT), random forest (RF), extremely randomized trees (ERTs), gradient boosted decision tree (GBDT), and extreme gradient boosting (XGBoost)—were systematically evaluated. Each algorithm reflects a distinct modeling philosophy: DT represents the simplest single-tree structure, RF utilizes ensemble learning through feature perturbation under the Bagging framework, ERT incorporates complete randomness in feature and threshold selection to enhance robustness, GBDT applies gradient boosting in a sequential manner to reduce bias, and XGBoost extends GBDT with second-order optimization, regularization, and efficient parallelization.

The comparative evaluation was based on three widely used regression metrics—coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE)—on both training and testing datasets. The results, summarized in Figure 9, highlight clear performance differences among the models. RF achieved the highest R^2 (0.999) and the lowest MSE (9×10^{-6}) on the training set, but this near-perfect fit indicates overfitting, as reflected by its reduced generalization on the test set (R^2 = 0.835, MSE = 0.001976). ERT also exhibited very high training accuracy (R^2 = 0.965) due to its randomized splitting strategy but suffered a notable performance drop on the test set (R^2 = 0.836, MSE = 0.001925), confirming its tendency toward overfitting.

GBDT delivered balanced but less competitive results, with moderate training accuracy (R^2 = 0.934) and reduced test performance (R^2 = 0.829), suggesting that its sequential boosting process is more sensitive to noise and parameter tuning. DT, as expected for a single-tree model, had the lowest accuracy on both datasets (R^2 = 0.922 on training, 0.814 on testing), indicating its limited capacity to capture complex nonlinear relationships.



(**a**) R^2 score

(**b**) MSE score

(**c**) MAE score

Figure 9. Comparison of scores of different prediction models. (The red one presents the best).

Among all candidates, XGBoost achieved the best trade-off between fitting accuracy and generalization, with R^2 values of 0.981 (training) and 0.855 (testing), and the lowest MAE (0.0156) on the test set. These results demonstrate that XGBoost's advanced regularization and optimization mechanisms effectively mitigate overfitting while maintaining high predictive precision. The improved generalization capability of XGBoost makes it the most suitable choice for aerodynamic coefficient prediction in scenarios involving diverse and nonlinear interactions among environmental and structural parameters.

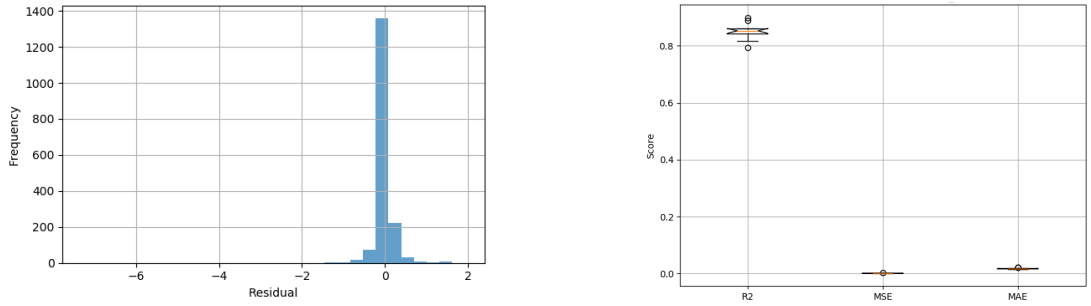
The overall comparison of the five tree-based models is summarized in Table 5. Although RF and ERT achieve near-perfect training accuracy, their substantial drop in testing performance reveals a pronounced overfitting tendency. DT and GBDT provide moderate results but fail to match the predictive precision of advanced ensemble methods. XGBoost consistently outperforms other models in test-set accuracy and error metrics, demonstrating superior generalization and robustness for aerodynamic coefficient prediction tasks.

Table 5. Summary of performance comparison among tree-based models.

Model	Training R^2	Testing R^2	Generalization Assessment
DT	0.922	0.814	Limited capacity, underfitting
RFT	0.999	0.835	Overfitting tendency
ERT	0.965	0.836	Overfitting tendency
GBDT	0.934	0.83	Sensitive to noise, unstable
XGBoost	0.981	0.855	Best trade-off, high accuracy

To evaluate the reliability of the XGBoost-based model, an uncertainty quantification analysis was conducted, and the results are shown in Figure 10. As presented in Figure 10a, the residual histogram indicates that prediction errors are highly concentrated around zero,

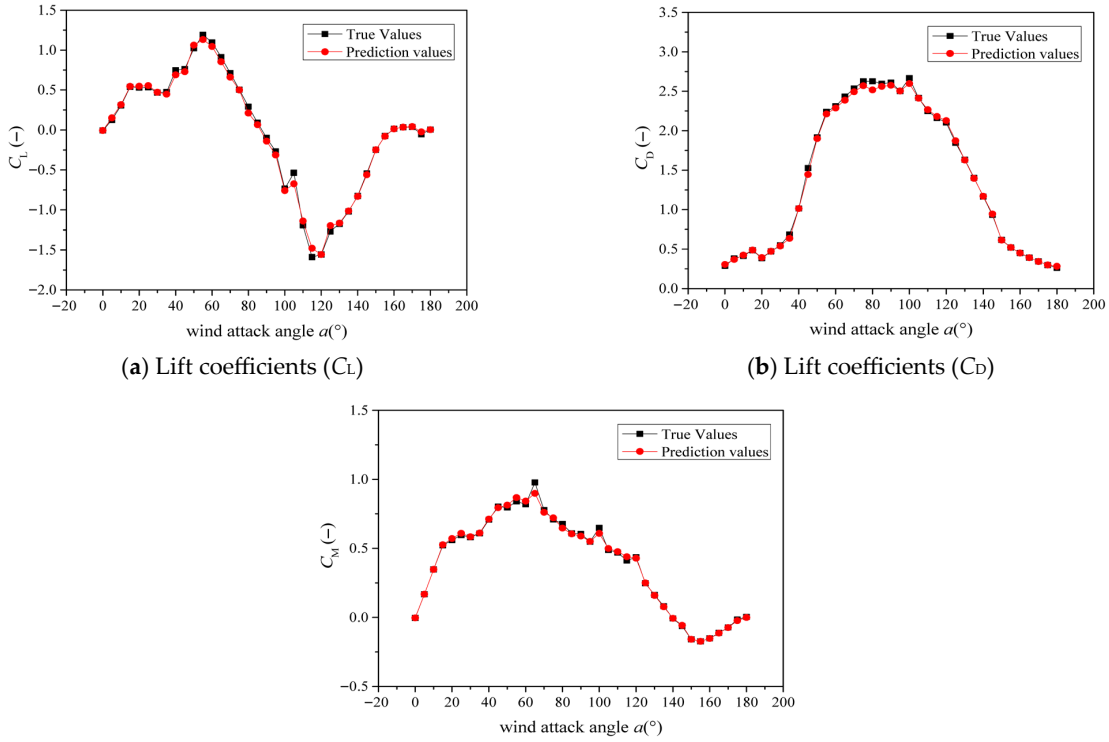
with only a few deviations on the negative side, suggesting that the model exhibits low bias and good overall calibration. Figure 10b further demonstrates the stability of model performance across different validation folds: the R^2 values remain consistently high (close to 0.85), while both MSE and MAE show very small dispersion. These results confirm that the proposed tree-based surrogate not only achieves accurate predictions but also provides reliable variance estimates, thereby supporting its robustness in aerodynamic coefficient modeling.



(**a**) Residual histogram for the test predictions (**b**) Cross-validated distributions

Figure 10. Uncertainty quantification via prediction intervals and variance estimates.

To further validate the predictive capability of the XGBoost-based surrogate model, Figures 11 and 12 present detailed comparisons between the predicted and true aerodynamic coefficients under two representative operating conditions. Figure 10 corresponds to a crescent-shaped ice case with a wind velocity of 10 m/s and an icing thickness of 20 mm, while Figure 11 corresponds to a sector-shaped ice case with a wind velocity of 30 m/s and the same icing thickness. For both scenarios, the variation of lift coefficient (C_L), drag coefficient (C_D), and moment coefficient (C_M) with wind attack angle (a) is shown for sub-conductor 1.



(**a**) Lift coefficients (C_L) (**b**) Lift coefficients (C_D)

(**c**) Moment coefficients (C_M)

Figure 11. Comparison of the true values and prediction values obtained by a surrogate model at the sub-conductor 1 with crescent ice, 10 m/s wind velocity, and 20 mm icing thickness.

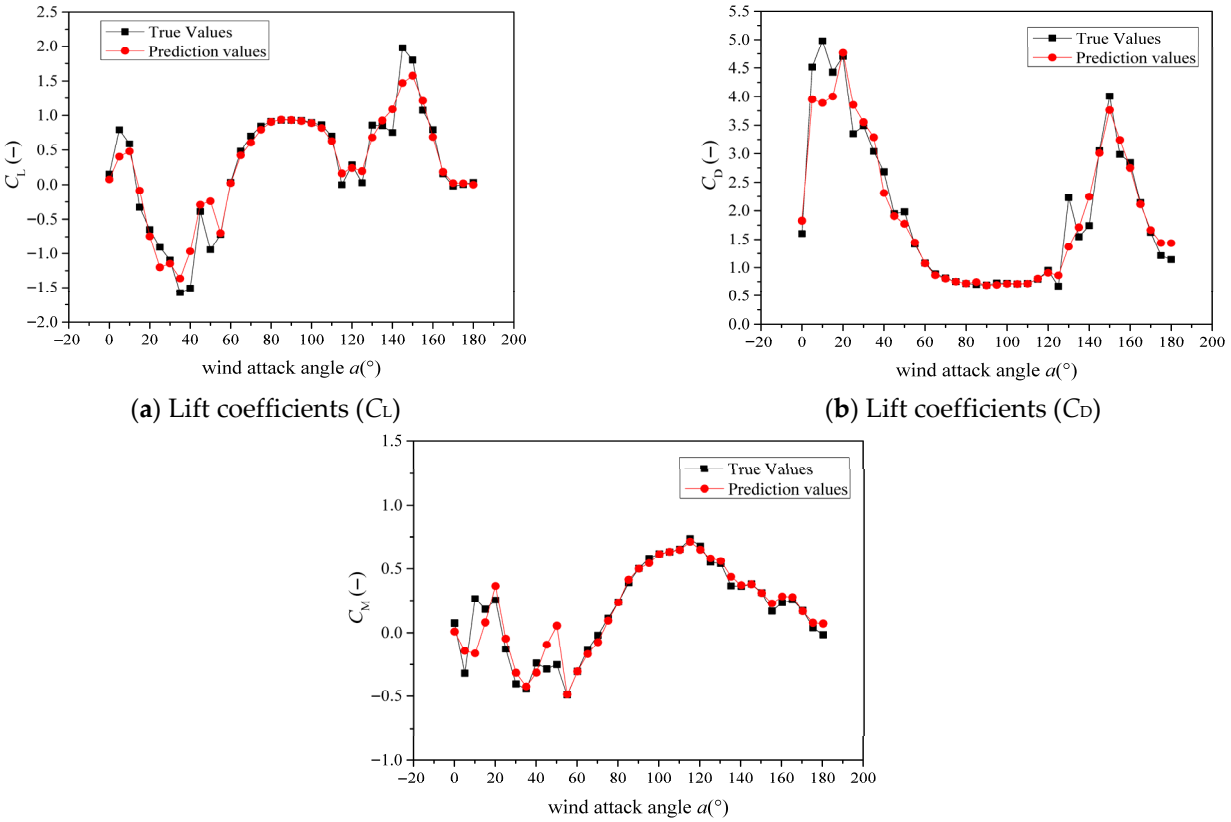


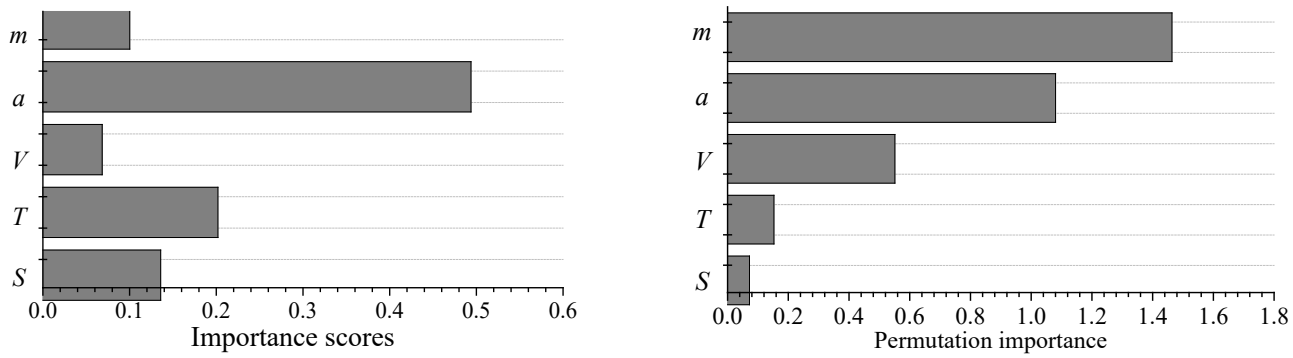
Figure 12. Comparison of the true values and prediction values obtained by a surrogate model at the sub-conductor 1 with sector ice, 30 m/s wind velocity, and 20 mm icing thickness.

Across the entire range of a from 0° to 180°, the predicted curves (red dots) closely follow the measured true values (black squares), accurately reproducing both the amplitude and phase of the aerodynamic coefficient variations. Minor local deviations are observed at certain angles, particularly near peak and valley regions, which may be attributed to local flow separation complexities not fully captured in the training data. Nonetheless, the overall agreement remains high, with the surrogate model successfully capturing the key aerodynamic features under different icing shapes and wind speeds.

These results confirm that the XGBoost-based model is capable of delivering accurate and stable predictions of aerodynamic coefficients under diverse and challenging conditions, further supporting its applicability in practical engineering analyses of wind-induced vibrations in iced bundled conductors.

Feature importance analysis is a critical step in understanding the internal decision-making process of machine learning models, as it quantifies the relative contribution of each input variable to the model's predictive performance. This not only enhances the interpretability of the surrogate model but also provides valuable guidance for prioritizing key parameters in subsequent aerodynamic studies and engineering applications.

Within the XGBoost framework, feature importance evaluation revealed, as shown in Figure 13a, that wind attack angle (a) is the most influential factor, contributing approximately 49.38% to the prediction of aerodynamic coefficients. This is followed by icing thickness (T, 20.19%), icing shape (S, 13.58%), sub-conductor number (m, 10.01%), and wind speed (V, 6.83%). These results indicate that the aerodynamic response of six-bundle conductors is dominated by the wind attack angle, while icing-related parameters also exert a significant impact, and wind speed plays a comparatively minor role under the considered scenarios.



(**a**) Importance scores (**b**) Permutation importance

Figure 13. Analysis of the importance characteristics of the XGBoost-based surrogate model.

To provide a more robust and unbiased interpretation beyond the gain-based metric, permutation importance analysis was also performed. As shown in Figure 13b, this approach identifies sub-conductor number (m) and wind attack angle (a) as the two dominant factors, with wind speed (V) ranking third, whereas icing thickness (T) and icing shape (S) contribute relatively less. The difference arises because gain-based importance reflects model-internal splitting criteria, which tend to favor continuous variables with more potential thresholds (such as wind attack angle), whereas permutation importance directly measures the impact of perturbing each feature on prediction accuracy. The consistency between both methods in highlighting wind attack angle and sub-conductor effects confirms their dominant physical role, while the divergence in the relative ranking of other variables underscores the necessity of combining multiple importance measures. This joint analysis not only reduces potential bias associated with a single metric but also provides a more comprehensive understanding of the aerodynamic drivers of iced bundled conductors.

By integrating this feature importance analysis with the results from model performance evaluations, the XGBoost-based surrogate model is confirmed to be the most effective choice among the investigated tree-based algorithms. It demonstrates high predictive accuracy, strong generalization capability, and clear interpretability, making it particularly suitable for aerodynamic coefficient prediction of bundled conductors under complex wind and icing conditions.

It should be noted that the present surrogate model was trained and validated entirely on CFD-generated data, which have been benchmarked against published wind tunnel measurements to ensure credibility. Nevertheless, direct experimental validation of the surrogate itself against independent wind tunnel or field measurements is still lacking. This limitation will be addressed in future work by designing dedicated wind tunnel campaigns and exploring field monitoring data, which will enable a more rigorous verification of surrogate predictions and provide a deeper quantification of predictive uncertainty under real icing events. In addition, the surrogate achieves inference at the second level per sample on a standard workstation, representing a speedup of several orders of magnitude compared with full CFD simulations (hour level). By contrast, the training process remains computationally intensive and is carried out offline using a precomputed CFD database. Future work will, therefore, focus on integrating physical experiments with data-driven approaches to develop real-time prediction capability.

5. Potential Applications of the Developed Surrogate Model

The tree-based surrogate model established in this study provides a powerful and efficient tool for predicting aerodynamic coefficients of six-bundle conductors under diverse wind and icing conditions. Compared with traditional wind tunnel testing and high-fidelity

CFD simulations, the surrogate model offers a substantial reduction in computational cost while maintaining high prediction accuracy, making it suitable for a variety of practical and research applications.

The model enables rapid aerodynamic coefficient prediction for typical transmission line configurations, allowing engineers to efficiently evaluate conductor aerodynamics under numerous environmental scenarios. This capability is particularly valuable in the early stages of line design, where multiple configurations and loading conditions must be assessed in a short time. In addition, the surrogate model serves as a key component for wind-induced vibration analysis, including galloping, sub-span oscillations, and wake-induced vibrations. By providing fast and reliable aerodynamic inputs, it facilitates dynamic simulations and stability assessments for conductors under complex environmental loads.

Beyond design-stage applications, the model can be integrated into fast monitoring and risk assessment systems. Coupled with online meteorological and structural monitoring data, the surrogate model can provide near-instantaneous aerodynamic estimates, enabling timely identification of adverse conditions and supporting preventive control measures. Moreover, the modeling framework developed in this study is scalable and adaptable. With appropriate retraining using additional datasets, the surrogate model can be extended to other conductor types, bundle configurations, or even different transmission line components, providing a generalizable methodology for aerodynamic performance evaluation across the power transmission sector.

Overall, the developed surrogate model bridges the gap between high-accuracy aerodynamic analysis and engineering efficiency, offering significant potential for both academic research and practical engineering applications. Nevertheless, due to the scope of this work, certain aspects—such as the influence of D-shaped icing and non-uniform icing—have not yet been considered and will be the subject of future in-depth investigations.

6. Conclusions

This study developed a large-scale aerodynamic coefficient dataset for six-bundle conductors based on high-throughput numerical simulations and established a multi-input, multi-output aerodynamic prediction model using several tree-based algorithms. It is concluded that:

(1) Among the five evaluated tree-based surrogate models (DT, RF, ERT, GBDT, and XGBoost), the XGBoost-based model consistently achieved the highest predictive accuracy and strongest generalization capability. This was evidenced by its superior R^2 values and lower MSE and MAE on the testing dataset, indicating a balanced trade-off between fitting precision and robustness.

(2) Global sensitivity analysis revealed that wind attack angle (a) was the most influential factor (55.38%), followed by icing thickness (T, 20.19%), icing shape (S, 13.58%), sub-conductor number (m, 10.01%), and wind speed (V, 6.83%). These findings confirmed that the aerodynamic responses of six-bundle conductors are governed by strong multi-factor coupling effects, with a and T exerting dominant influences.

(3) Compared to conventional CFD simulation workflows, the proposed XGBoost-based surrogate model achieved high-accuracy aerodynamic coefficient predictions within seconds. This substantial improvement in computational efficiency enabled rapid evaluation of aerodynamic behavior and supports engineering analyses of wind-induced vibration (galloping, sub-span oscillations, etc.) in UHV transmission lines.

It should be noted that, due to the scope of the present study, certain factors, such as 3D models, D-shaped icing, non-uniform icing distributions, and typhoons with complex wind

loads, were not considered. These aspects will be the focus of future research to further enhance the applicability and robustness of the proposed surrogate modeling framework.

Author Contributions: Conceptualization, G.Y. and Z.L. (Zhiyi Liu); methodology, G.H.; software, Z.L. (Zhiguo Li); validation, Z.L. (Zhiyi Liu); investigation, R.T. and G.Y.; data curation, G.Y.; writing—original draft preparation, G.Y.; writing—review and editing, G.H.; visualization, A.W.; supervision, G.H.; project administration, Z.L. (Zhiguo Li); funding acquisition, Z.L. (Zhiguo Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Funding Project of China Southern Power Grid, grant number CGYKJXM20220333.

Data Availability Statement: Some of the data generated or analyzed during this study are included in this published article. Additional datasets are available from the corresponding author upon reasonable request.

Acknowledgments: The authors appreciate the constructive comments and suggestions from the anonymous reviewers, which helped improve the quality of this manuscript.

Conflicts of Interest: Guoliang Ye, Zhiguo Li and Anjun Wang are employed by China Southern Power Grid Co., Ltd.; Zhiyi Liu and Ruomei Tang are employed by Powerchina Chengdu Electric Power Fittings Co., Ltd. The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UHV	Ultra-high voltage
LHS	Latin hypercube sampling
S-A	Spalart–Allmaras
DT	Decision tree
RF	Random forest
ERT	Extremely randomized tree
GBDT	Gradient boosted decision tree
XGBoost	Extreme gradient boosting

References

1. Papailiou, K.O. *Overhead Lines*; Springer Nature: Paris, France, 2017.
2. Van Dyke, P.; Havard, D.; Laneville, A. Effect of Ice and Snow on the Dynamics of~Transmission Line Conductors. In *Atmospheric Icing of Power Networks*; Farzaneh, M., Ed.; Springer Netherlands: Dordrecht, The Netherlands, 2008; pp. 171–228.
3. Cosmai, U.; Van Dyke, P.; Mazzola, L.; Lillien, J.-L. Conductor Motions Chapter 10. In *Overhead Lines, CIGRE Green Books, K.O. Papailiou*; CIGRE: Paris, France, 2014.
4. Lilien, J.-L.; Van Dyke, P.; Asselin, J.-M.; Farzaneh, M.; Halsan, K.; Havard, D. *State of the Art of Conductor Galloping, A Complementary Document to "Transmission Line Refernce Book—Wind-Induced Conductor Motion Chapter 4: Conductor Galloping" Based on EPRI Research Project 792*; CIGRE: Paris, France, 2005.
5. Stengel, D.; Thiele, K.; Clobes, M.; Mehdianpour, M. Aerodynamic damping of nonlinear movement of conductor cables in wind tunnel tests, numerical simulations and full scale measurements. *J. Wind Eng. Ind. Aerodyn.* **2017**, *169*, 47–53. [CrossRef]
6. Cai, M.; Zhou, L.; Lei, H.; Huang, H. Wind Tunnel Test Investigation on Unsteady Aerodynamic Coefficients of Iced 4-Bundle Conductors. *Adv. Civ. Eng.* **2019**, *2019*, 2586242. [CrossRef]
7. Lou, W.J.; Huang, C.R.; Huang, M.F.; Yu, J. An aerodynamic anti-galloping technique of iced 8-bundled conductors in ultra-high-voltage transmission lines. *J. Wind Eng. Ind. Aerodyn.* **2019**, *193*, 103972. [CrossRef]
8. Flaga, A.; Pistol, A.; Krajewski, P.; Flaga, Ł. Aerodynamic and aeroelastic wind tunnel model tests of overhead power lines in triangular configuration under different icing conditions. *Cold Reg. Sci. Technol.* **2020**, *170*, 102919. [CrossRef]
9. Matsumiya, H.; Nishihara, T.; Yagi, T. Aerodynamic modeling for large-amplitude galloping of four-bundled conductors. *J. Fluids Struct.* **2018**, *82*, 559–576. [CrossRef]
10. Cai, M.; Yan, B.; Lu, X.; Zhou, L. Numerical Simulation of Aerodynamic Coefficients of Iced-Quad Bundle Conductors. *IEEE Trans. Power Delivery* **2015**, *30*, 1669–1676. [CrossRef]

11. Liu, X.; Zou, M.; Wu, C.; Yan, B.; Cai, M. Galloping Stability and Aerodynamic Characteristic of Iced Transmission Line Based on 3-DOF. *Shock Vib.* **2020**, *2020*, 8828319. [CrossRef]
12. Liang, H.; Zhang, Z.; Kong, X.; Yang, D.; Lie, J.; Du, B. Galloping behavior of insulated overhead transmission line based on aerodynamic analysis. *Sci. Rep.* **2025**, *15*, 2669. [CrossRef]
13. Zhou, L.S.; Yan, B.; Zhang, L.; Zhou, S. Study on galloping behavior of iced eight bundle conductor transmission lines. *J. Sound Vib.* **2016**, *362*, 85–110. [CrossRef]
14. Yan, B.; Liu, X.; Lv, X.; Zhou, L. Investigation into galloping characteristics of iced quad bundle conductors. *J. Vib. Control* **2016**, *22*, 965–987. [CrossRef]
15. Liu, X.H.; Yan, B.; Zhang, H.Y.; Zhou, S. Nonlinear numerical simulation method for galloping of iced conductor. *Appl. Math. Mech.-Engl. Ed.* **2009**, *30*, 489–501. [CrossRef]
16. Matsumiya, H.; Yagi, T.; Macdonald, J.H.G. Unsteady aerodynamic force modelling for 3-DoF-galloping of four-bundled conductors. *J. Fluids Struct.* **2022**, *112*, 103581. [CrossRef]
17. Absi, G.N.; Mahadevan, S. Multi-fidelity approach to dynamics model calibration. *Mech. Syst. Sig. Process.* **2016**, *68–69*, 189–206. [CrossRef]
18. Huang, G.; Wu, G.; Yang, Z.; Chen, X.; Wei, W. Development of surrogate models for evaluating energy transfer quality of high-speed railway pantograph-catenary system using physics-based model and machine learning. *Appl. Energy* **2023**, *333*, 120608. [CrossRef]
19. Li, Z.; Wu, G.; Huang, G.; Guo, Y.; Zhu, H. Study on Icing Prediction for High-Speed Railway Catenary Oriented to Numerical Model and Deep Learning. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 1189–1200. [CrossRef]
20. Liu, Y.; Guo, Y.; Wang, B.; Zhang, X.; Huang, G.; Zhang, G.; Liu, Y.; Deng, Q.; Wu, G. Pollution morphology characteristics on a superhydrophobic surface and its pollution flashover voltage in DC electric field. *High Voltage* **2021**, *7*, 564–574. [CrossRef]
21. Huang, G.; Wu, G.; Guo, Y.; Liang, M.; Li, J.; Dai, J.; Yan, X.; Gao, G. Risk assessment models of power transmission lines undergoing heavy ice at mountain zones based on numerical model and machine learning. *J. Clean. Prod.* **2023**, *415*, 137623. [CrossRef]
22. Wen, N.; Yan, B.; Mou, Z.; Huang, G.; Yang, H.; Lv, X. Prediction models for dynamic response parameters of transmission lines after ice-shedding based on machine learning method. *Electr. Power Syst. Res.* **2022**, *202*, 107580. [CrossRef]
23. Mou, Z.; Yan, B.; Yang, H.; Cai, D.; Huang, G. Prediction model for aerodynamic coefficients of iced quad bundle conductors based on machine learning method. *R. Soc. Open Sci.* **2021**, *8*, 210568. [CrossRef]
24. Liang, J.; Cai, M.; Wang, Q.; Zhou, L.; Liu, J.; Min, G.; Huang, H. Aerodynamic Characteristics Analysis of Iced Conductor Based on BP Neural Network. *Buildings* **2023**, *13*, 64. [CrossRef]
25. Chen, S.; Bao, X.; Lou, W.; Wen, Z.; Wu, H. Prediction of aerodynamic coefficients of iced conductors based on composite image and convolutional neural network. *J. Wind Eng. Ind. Aerodyn.* **2023**, *238*, 105430. [CrossRef]