

Article

A Dual-Optimized Hybrid Deep Learning Framework with RIME-VMD and TCN-BiGRU-SA for Short-Term Wind Power Prediction

Zhong Wang ^{1,2} , Kefei Zhang ^{1,2,*}, Xun Ai ^{1,2}, Sheng Liu ^{1,2} and Tianbao Zhang ^{1,2}

¹ School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China; wangzhong@hbut.edu.cn (Z.W.); 19945167311@163.com (X.A.); 102310308@hbut.edu.cn (S.L.); 102410344@hbut.edu.cn (T.Z.)

² Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, China

* Correspondence: zhangkefei@hbut.edu.cn; Tel.: +86-153-0715-8641

Abstract

Precise short-term forecasting of wind power generation is indispensable for ensuring the security and economic efficiency of power grid operations. Nevertheless, the inherent non-stationarity and stochastic nature of wind power series present significant challenges for prediction accuracy. To address these issues, this paper proposes a dual-optimized hybrid deep learning framework combining Spearman correlation analysis, RIME-VMD, and TCN-BiGRU-SA. First, Spearman correlation analysis is employed to screen meteorological factors, eliminating redundant features and reducing model complexity. Second, an adaptive Variational Mode Decomposition (VMD) strategy, optimized by the RIME algorithm based on Minimum Envelope Entropy, decomposes the non-stationary wind power series into stable intrinsic mode functions (IMFs). Third, a hybrid predictor integrating Temporal Convolutional Network (TCN), Bidirectional Gated Recurrent Unit (BiGRU), and Self-Attention (SA) mechanisms is constructed to capture both local trends and long-term temporal dependencies. Furthermore, the RIME algorithm is utilized again to optimize the hyperparameters of the deep learning predictor to avoid local optima. The proposed framework is validated using full-year datasets from two distinct wind farms in Xinjiang and Gansu, China. Experimental results demonstrate that the proposed model achieves a Root Mean Square Error (RMSE) of 7.5340 MW on the primary dataset, significantly outperforming mainstream baseline models. The multi-dataset verification confirms the model's superior prediction accuracy, robustness against seasonal variations, and strong generalization capability.

Keywords: wind power prediction; variational mode decomposition; RIME optimization; dual-optimization; TCN-BiGRU-SA



Academic Editor: Pedro Couto

Received: 27 December 2025

Revised: 27 January 2026

Accepted: 29 January 2026

Published: 3 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Background

Driven by the global imperative for decarbonization and the urgent need to achieve net-zero emissions, wind energy has solidified its position as a cornerstone of the modern renewable energy landscape. According to the Global Wind Energy Council, the cumulative installed wind capacity worldwide surpassed the historic 1 TW milestone by 2023, signaling a pivotal shift in the global energy transition [1]. This rapid expansion is further

accelerated by supportive policies, such as those outlined in the Paris Agreement and national renewable targets [2]. Nevertheless, the large-scale integration of wind power presents significant challenges to power system stability due to the inherent intermittency, volatility, and non-stationarity of wind resources [3,4]. Inaccurate predictions frequently result in high curtailment rates—often exceeding 10% in regions like China—and economic inefficiencies within grid dispatch operations, leading to billions in annual losses [5,6]. Consequently, the development of high-precision short-term wind power prediction models is critical for ensuring the safety, stability, and economic viability of smart grids, enabling better integration with energy storage and demand-response systems [7].

1.2. Literature Review

Existing forecasting methodologies are generally classified into physical, statistical, and artificial intelligence models. Physical models, primarily Numerical Weather Prediction (NWP), rely on complex atmospheric simulations; however, they are often constrained by high computational costs, sensitivity to initial boundary conditions, and poor performance in short-term horizons (e.g., 15–60 min) [8,9]. Conversely, statistical approaches such as ARIMA and Kalman filters effectively model linear trends but struggle to capture the highly non-linear dynamics inherent in wind speed time series, resulting in errors up to 20–30% in volatile conditions [10,11].

In recent years, data-driven deep learning techniques have demonstrated superior performance in modeling complex temporal dependencies. Recurrent Neural Networks (RNNs) and their variants—specifically Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU)—have been widely adopted for sequential data modeling, achieving RMSE reductions of 10–15% over statistical methods [12–14]. However, these recurrent architectures often suffer from gradient vanishing issues, lack efficiency in parallel processing, and fail to handle multi-scale features effectively [15]. To address these limitations, Temporal Convolutional Networks (TCNs) were introduced, utilizing dilated causal convolutions to expand the receptive field and capture multi-scale local features more effectively than traditional RNNs, with reported improvements in convergence speed by up to 50% [16–18]. Despite these advancements, single models often fail to simultaneously capture both local high-frequency variations and global long-term dependencies. Consequently, hybrid architectures combining TCNs with bidirectional mechanisms (e.g., BiGRU) and attention modules have emerged as a promising avenue to enhance feature extraction capabilities, yielding further accuracy gains in real-world deployments [19,20].

A further challenge lies in the non-stationarity of wind power data, exacerbated by seasonal variations and extreme weather events [21]. Decomposition-ensemble strategies, which separate raw signals into simpler sub-components, have proven effective in mitigating this issue. While Empirical Mode Decomposition (EMD) and its variants (e.g., CEEMDAN) are popular, they lack a rigorous mathematical foundation and are susceptible to mode mixing and end effects, leading to reconstruction errors of 5–10% [22–24]. Variational Mode Decomposition (VMD) offers a robust, non-recursive alternative by solving a constrained variational problem, improving decomposition quality in noisy environments [25,26]. However, the efficacy of VMD is heavily contingent upon the precise selection of its mode number (K) and penalty factor (α), which are traditionally determined through experience or trial-and-error, often resulting in suboptimal decompositions [27,28].

To optimize these hyperparameters, meta-heuristic algorithms such as Particle Swarm Optimization (PSO) and Gray Wolf Optimizer (GWO) have been applied, but they frequently suffer from premature convergence and entrapment in local optima, especially in high-dimensional spaces [29–31]. The recently proposed RIME optimization algorithm,

inspired by the physical process of rime ice growth, has demonstrated a superior balance between global exploration (via “soft-rime search”) and local exploitation (via “hard-rime puncture”), outperforming PSO and GWO in benchmarks by 10–20% in terms of convergence speed and solution quality [32,33]. Despite these advances, the recent literature reveals persistent gaps: (1) manual or suboptimal parameter tuning in decomposition leads to mode mixing and reduced predictability; (2) hybrid predictors often overlook dynamic weighting of critical time steps (e.g., ramp events); and (3) single-stage optimization fails to ensure end-to-end global optimality, limiting generalization across diverse wind farms and seasons [34].

Table 1 summarizes key existing methods, their shortcomings, and how the proposed framework addresses these gaps.

Table 1. Comparison of existing wind power prediction methods and the proposed framework.

Method/Reference	Key Components	Shortcomings	How Proposed Framework Addresses Gap
ARIMA-based [10,11]	Statistical modeling of linear trends	Fails in non-linear, volatile data; high errors in short-term forecasts	Integrates deep learning for non-linear capture and decomposition for non-stationarity handling
LSTM/GRU [12–14]	Recurrent networks for temporal dependencies	Gradient vanishing; inefficient parallelization; ignores multi-scale features	Combines TCN for efficient multi-scale extraction and BiGRU for bidirectional dependencies
TCN hybrids [16–18]	Dilated convolutions for local features	Limited global dependency modeling; no dynamic focus on critical events	Adds BiGRU-SA for global and attention-based weighting
CEEMDAN-based [22–24]	Recursive decomposition ensembles	Mode mixing, end effects; lacks mathematical rigor	Uses non-recursive RIME-VMD with entropy minimization for stable IMFs
PSO/GWO-VMD [29–31]	Meta-heuristic optimization of VMD	Premature convergence; local optima traps	Employs RIME for superior exploration-exploitation balance

1.3. Research Problem and Contributions

The core research problem addressed in this study is the persistent inaccuracy in short-term wind power prediction due to non-stationarity, suboptimal parameter selection, and insufficient global optimization in hybrid models, particularly under diverse seasonal and geographical conditions. Existing approaches often rely on single-stage optimization, which fails to jointly tune decomposition and prediction stages, leading to local optima and poor generalization (e.g., RMSE increases of 15–25% in volatile seasons). Moreover, while attention mechanisms improve focus, they are underexplored in conjunction with adaptive decompositions like VMD, where parameters are not dynamically optimized [31]. The choice of RIME over established optimizers (e.g., PSO/GWO) is justified by its physics-inspired mechanisms that prevent stagnation, ensuring better handling of high-dimensional spaces as validated in recent benchmarks.

Theoretically, the proposed “dual-optimization” paradigm—optimizing both decomposition (via Envelope Entropy minimization for sparse, stationary signals) and prediction stages—offers benefits over single-stage approaches: it achieves end-to-end global optimality, reduces overfitting by aligning data preprocessing with model hyperparameters, and enhances robustness (e.g., 20–30% error reduction in multi-dataset tests). Motivated by

these gaps, this paper proposes a novel dual-optimized hybrid framework, RIME-VMD-RIME-TCN-BiGRU-SA. The main contributions are as follows:

1. An Adaptive Decomposition Strategy based on Minimum Envelope Entropy: To address mode mixing in standard VMD, an adaptive RIME-VMD strategy optimizes K and α using Envelope Entropy as the fitness function, selected for its ability to quantify signal sparsity and regularity in noisy wind data. This ensures stationary Intrinsic Mode Functions (IMFs), improving downstream learnability.
2. A Composite Spatiotemporal Predictor with Dynamic Attention: The TCN-BiGRU-SA model captures local trends (TCN), long-term dependencies (BiGRU), and critical events (SA), overcoming equal treatment of time steps in RNNs.
3. A Systematic “Decomposition-Prediction” Dual-Optimization Paradigm: RIME tunes both stages for global optimality, with practical implications like reduced computational inference time (detailed in experiments) for grid operators.
4. Robust Validation and Generalization: Validated on full-year, multi-dataset scenarios, addressing single-site limitations and demonstrating seasonal/geographical transferability.

2. Methodology

2.1. Overall Framework

To address the challenges of non-stationarity and stochastic volatility in wind power forecasting, this study proposes a dual-optimized hybrid framework named RIME-VMD-RIME-TCN-BiGRU-SA. As illustrated in Figure 1, the systematic workflow consists of three integrated phases:

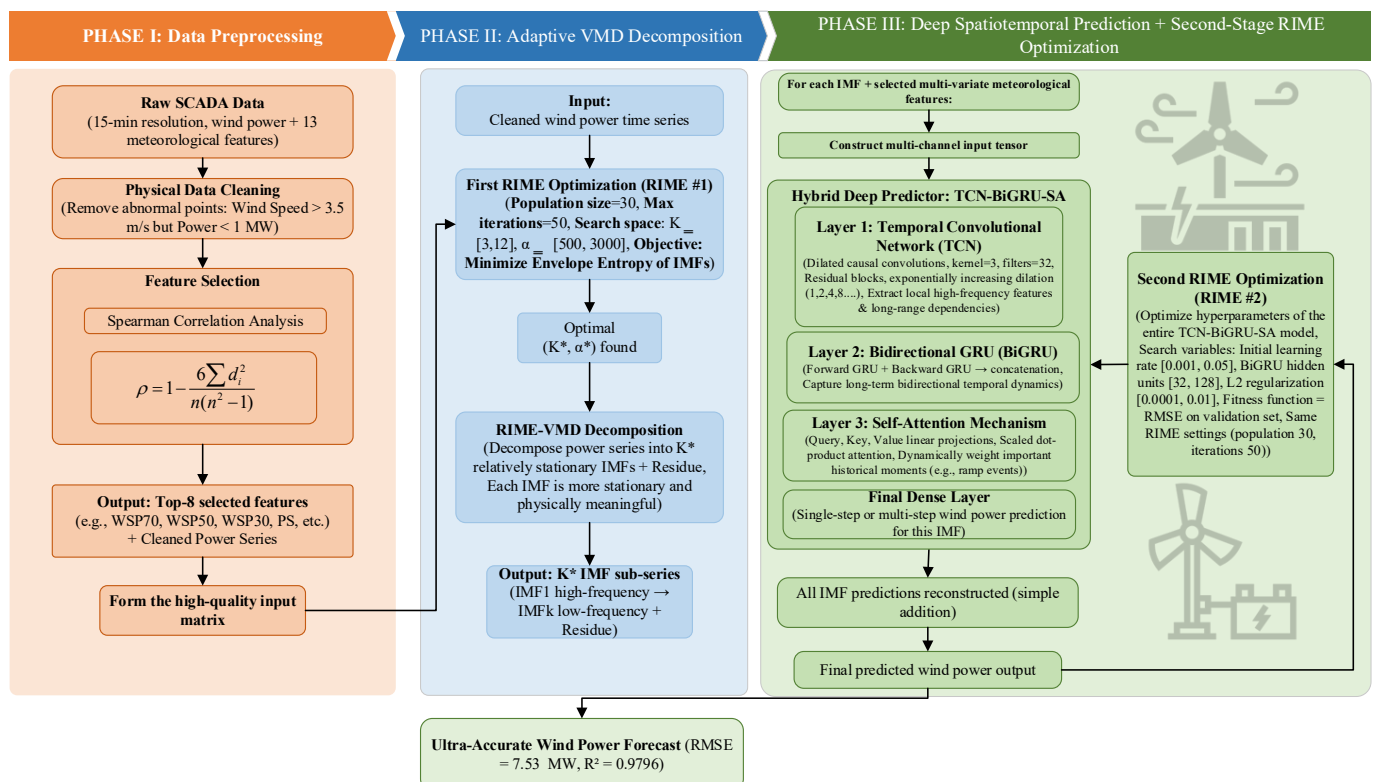


Figure 1. The structure of the proposed WPP model.

1. Data Preprocessing (Phase I): A Spearman-based feature selection strategy combining physical cleaning and dual-metric assessment is applied to construct high-quality input.

2. Adaptive Decomposition (Phase II): The wind power series is decomposed into stationary Intrinsic Mode Functions (IMFs) using Variational Mode Decomposition (VMD), where the key parameters (K, α) are adaptively optimized by the RIME algorithm to minimize Envelope Entropy.

3. Spatiotemporal Prediction (Phase III): A composite predictor (TCN-BiGRU-SA) is employed to extract multi-scale features, with its hyperparameters further tuned by a second-stage RIME optimization to ensure global optimality.

2.2. Feature Selection Strategy Based on Spearman Correlation

High-quality input features are the foundation of accurate prediction. To eliminate redundant meteorological information and reduce computational complexity, a two-step preprocessing strategy is adopted.

First, a physics-based cleaning rule is applied to remove abnormal downtime samples (e.g., instances where wind speed is high but power is zero due to turbine maintenance).

Subsequently, to identify the most relevant meteorological variables, we employ the Spearman rank correlation coefficient. Unlike Pearson correlation, which assumes linear relationships, Spearman correlation assesses monotonic relationships, making it suitable for the non-linear dynamics of wind power generation. The coefficient ρ is calculated as

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

where d_i is the difference between the ranks of the candidate feature and the target wind power, and n is the number of samples. Based on the calculated ρ values, the variables with the highest correlation (e.g., wind speed and direction at different heights) are selected to form the optimal input matrix.

2.3. Adaptive VMD Decomposition Based on RIME

2.3.1. Variational Mode Decomposition (VMD)

VMD is a non-recursive signal processing method that decomposes the input signal $f(t)$ into K discrete modes $u_k(t)$, each oscillating around a center frequency ω_k . Unlike recursive sieving methods (e.g., EMD), VMD formulates the decomposition as a constrained variational problem:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } & \sum_{k=1}^K u_k(t) = f(t) \end{aligned} \quad (2)$$

The solution is obtained using the Alternating Direction Method of Multipliers (ADMM).

2.3.2. Parameter Optimization via RIME

Chosen over other metrics like kurtosis for its sensitivity to wind noise [26]. The efficacy of VMD relies heavily on the preset mode number K and penalty factor α . Arbitrary selection often leads to mode mixing or signal loss. To address this, we employ the RIME optimization algorithm, a physics-inspired meta-heuristic based on the formation of rime ice, to adaptively determine the optimal (K, α) pair. RIME is selected for its superior balance between global exploration (“Soft-rime search”) and local exploitation (“Hard-rime puncture”).

Optimization Objective (Envelope Entropy): In this study, Minimum Envelope Entropy (E_p) is selected as the fitness function. Since wind power signals contain significant noise and intermittency, envelope entropy effectively evaluates the sparsity and regularity of the

decomposed IMFs. A lower entropy value indicates a more stationary signal with distinct feature information, which is more conducive to downstream neural network learning. The objective is formulated as

$$\text{Minimize Fitness} = \min_{(K,\alpha)} \left(\sum_{k=1}^K E_p(\text{IMF}_k) \right) \quad (3)$$

2.4. The Composite Deep Learning Predictor (TCN-BiGRU-SA)

To capture the complex spatiotemporal dynamics of the decomposed sub-signals, a hybrid predictor is constructed. This section focuses on the rationale for their integration:

1. Local Feature Extraction (TCN Module): The Temporal Convolutional Network (TCN) utilizes dilated causal convolutions to exponentially expand the receptive field. This structure allows the model to capture local high-frequency variations and short-term trends without future information leakage, addressing the parallel processing limitations of RNNs.

2. Global Temporal Modeling (BiGRU Module): The feature maps extracted by TCN are fed into a Bidirectional GRU (BiGRU). Compared to LSTM, BiGRU offers higher computational efficiency. By processing the sequence in both forward and backward directions, BiGRU effectively captures global long-term dependencies, connecting current power fluctuations with historical contexts.

3. Dynamic Weighting (Self-Attention Module): Standard BiGRU layers treat all time steps in the sequence equally. However, in wind power forecasting, extreme events (e.g., sudden ramps or gusts) carry more critical information. To overcome this, a Self-Attention (SA) mechanism is integrated as the final layer. It calculates attention scores to dynamically assign higher probability weights to these critical historical moments, effectively highlighting salient features while suppressing irrelevant noise. The attention mechanism computes the dependencies within the hidden state sequence H output by the BiGRU layer. First, H is projected into Query (Q), Key (K), and Value (V) matrices through linear transformations:

$$Q = HW^Q, K = HW^K, V = HW^V \quad (4)$$

Then, the attention weights are calculated using the scaled dot-product function to generate the final context vector:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (5)$$

where d_k is the dimension of the key vectors. This calculation enables the model to dynamically assign higher probability weights to critical historical moments (e.g., ramp events).

2.5. RIME Optimization Algorithm

In this study, the RIME algorithm is employed as the core optimizer for both the adaptive VMD decomposition (Phase II) and the deep learning hyperparameter tuning (Phase III). RIME is a novel physics-inspired meta-heuristic algorithm that mathematically models the growth process of rime ice. Unlike traditional swarm algorithms, RIME introduces a unique mechanism that simulates the movement of soft-rime particles (facilitating global exploration) and hard-rime particles (facilitating local exploitation). This duality enables the algorithm to effectively avoid local optima while ensuring rapid convergence. The optimization workflow is mathematically formulated in the following four phases, and the overall principle is illustrated in Figure 2.

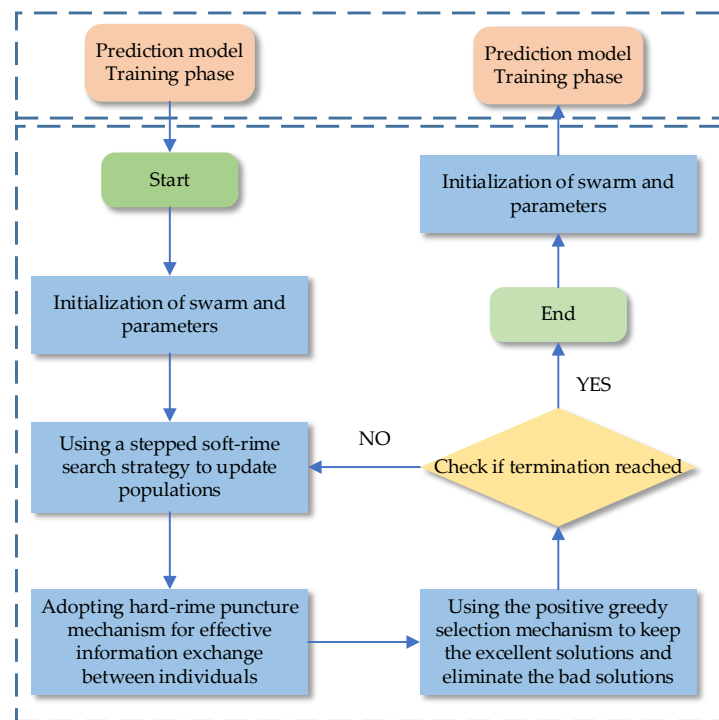


Figure 2. Principle of the RIME optimization algorithm.

1. Population Initialization

The algorithm commences by initializing a population R consisting of n rime agents, where each agent represents a potential solution vector in the d -dimensional parameter space (e.g., VMD parameters or BiGRU hyperparameters). The population R is formally expressed as a matrix:

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,d} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,1} & r_{n,2} & \cdots & r_{n,d} \end{bmatrix} \quad (6)$$

where $r_{i,j}$ denotes the position of the j -th dimension of the i -th agent.

2. Soft-Rime Search Strategy (Global Exploration)

Inspired by the high randomness of soft fog particles carried by the wind, this phase is designed to extensively explore the solution space to identify potential high-quality regions. The position update of a rime particle $R_{i,j}$ is determined by the environmental factor E and the adhesion coefficient β . The update rule is formulated as

$$R_{i,j}^{new} = R_{best,j} + r_1 \cdot \cos(\theta) \cdot \beta \cdot (UB_j - LB_j) \quad (7)$$

where $R_{best,j}$ is the j -th dimension of the current global optimal agent, r_1 is a random number in $[-1, 1]$, and UB_j, LB_j are the upper and lower bounds of the search space. The variable θ controls the direction of particle movement and changes with the iteration count t :

$$\theta = \frac{\pi \cdot t}{T_{max}} \quad (8)$$

where T_{max} is the maximum number of iterations. The environmental factor E , which ensures convergence, is defined as

$$E = \sqrt{\frac{t}{T_{max}}} \quad (9)$$

The adhesion coefficient β simulates the aggregation propensity of particles and is calculated as a step function to control the exploration precision:

$$\beta = 1 - \left\lceil \frac{w \cdot t}{T_{max}} \right\rceil \frac{1}{w} \quad (10)$$

where $\lceil \cdot \rceil$ denotes the rounding operation and w (default set to 5) controls the segmentation of the search process.

3. Hard-Rime Puncture Mechanism (Local Exploitation)

As the iteration progresses, the algorithm simulates the formation of hard rime under strong wind conditions. To refine the solution accuracy and escape local optima, a “puncture mechanism” is introduced. This allows agents to exchange information by crossing over with the global best solution based on their fitness values. The update rule is

$$R_{i,j}^{new} = \begin{cases} R_{best,j} & \text{if } r_2 < F_{norm}(R_i) \\ R_{i,j} & \text{otherwise} \end{cases} \quad (11)$$

where r_2 is a random number in $[0, 1]$, and $F_{norm}(R_i)$ represents the normalized fitness value of the i -th agent, which determines the probability of successfully replacing the particle with the global optimum.

4. Positive Greedy Selection

To guarantee monotonic improvement in solution quality, a positive greedy selection strategy is applied at the end of each iteration. The fitness of the updated agent R_i^{new} is compared with its pre-update value R_i^{old} . The population is updated only if the new solution yields a lower error (e.g., lower envelope entropy or RMSE):

$$R_i^{t+1} = \begin{cases} R_i^{new} & \text{if } f(R_i^{new}) < f(R_i^{old}) \\ R_i^{old} & \text{otherwise} \end{cases} \quad (12)$$

This mechanism ensures that high-quality solutions are preserved while continuously driving the population towards the global optimum.

3. Case Study

3.1. Data Description and Correlation Analysis

The experimental data was acquired from a large-scale onshore wind farm located in Xinjiang, China, a region characterized by abundant yet highly volatile renewable energy resources. Due to the complex topography and variable climate, the wind signals in this area exhibit severe non-stationarity and stochastic fluctuations, providing an ideal “stress test” scenario for evaluating the robustness of forecasting models.

The dataset covers a complete annual cycle from 1 January 2019 to 31 December 2019, ensuring that seasonal variations—from stable summer breezes to extreme winter gusts—are fully captured. The data was sampled at a high-frequency temporal resolution of 15 min, resulting in a total of 35,040 observations. To comprehensively analyze the meteorological drivers, the dataset encompasses multi-dimensional features, including wind speed and wind direction measured at four distinct vertical heights (10 m, 30 m, 50 m, and 70 m). Additionally, environmental variables such as ambient temperature, atmospheric pressure, and relative humidity are included to account for air density variations. Detailed descriptions of the variables and statistical details of the dataset are given in Table 2. This comprehensive multivariate time series serves as the foundation for verifying the effectiveness of the proposed framework.

Table 2. Detailed description of the collected wind power related datasets.

Symbol	Max	Min	Mean	Std	Skewness	ADF Statistic
WSP10 (m/s)	21.96	0.32	6.31	3.75	0.82	−17.80
WSP30 (m/s)	22.02	0.34	6.72	4.09	0.64	−17.69
WSP50 (m/s)	22.24	0.32	7.19	4.30	0.59	−17.78
WSP70 (m/s)	22.76	0.32	7.46	4.66	0.52	−17.73
WD10 (°)	355.04	2.05	143.11	81.96	0.50	−13.86
WD30 (°)	355.18	0.97	152.48	84.09	0.40	−13.41
WD50 (°)	355.26	0.70	149.64	86.14	0.37	−15.57
WD70 (°)	355.60	1.64	157.06	87.85	0.32	−13.61
TEMP (°C)	15.34	12.78	13.49	0.49	1.07	−45.07
PS (hp)	882.60	855.47	868.88	4.52	0.17	−9.39
RH (%)	95.27	4.83	37.00	20.23	0.68	−8.50
WP (mw)	200.05	0.10	70.30	66.00	0.56	−16.34

3.2. Feature Selection Results

To construct an effective input matrix for the prediction model, the correlation between various meteorological variables and wind power output was analyzed using the Spearman rank correlation coefficient, as defined in Section 2.2. Figure 3 illustrates the correlation analysis results. Based on the magnitude of the absolute correlation coefficients, we selected the top-8 features to form the final input vector. Specifically, these features consist of wind speed and wind direction at four distinct heights (10 m, 30 m, 50 m, and 70 m). This selection prioritizes the most informative physical drivers—as wind kinetic energy is the primary source of power—while excluding variables with weaker correlations, such as temperature, pressure, and humidity, to reduce computational redundancy.

**Figure 3.** Wind power heatmap based on Spearman coefficient.

3.3. Evaluation Metrics

To quantitatively evaluate the prediction performance, four standard metrics are employed: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Abso-

lute Percentage Error (MAPE), and Coefficient of Determination (R^2). The formulas are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (15)$$

$$\text{MAPE} = \frac{1}{N_{\text{valid}}} \sum_{i \in \Omega} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (16)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of actual values.

Note: Since wind power data contains zero or near-zero values during downtime, direct MAPE calculation leads to infinite values. Therefore, a threshold filtering method is applied: MAPE is calculated only for sample set Ω where the actual power $y_i > 1\%P_{\text{rated}}$ (approx. 1 MW).

3.4. Experimental Setup and Parameter Settings

The experiments were conducted on a workstation with an AMD 9600X CPU and 32 GB RAM, using MATLAB 2023b. To ensure fairness, all baseline models were optimized to their best states.

The parameter search ranges for the Dual-Optimization strategy are set as follows:

RIME Algorithm: Population size $N = 30$, maximum iterations $T = 60$.

Phase II (VMD Optimization): Number of modes $K \in [2, 12]$, penalty factor $\alpha \in [100, 3000]$.

Phase III (Model Optimization):

Initial learning rate: $[0.001, 0.05]$.

Number of BiGRU neurons: $[32, 128]$.

L2 Regularization: $[0.0001, 0.01]$.

TCN filters: Fixed at 32, kernel size: 3.

Data Partitioning: The dataset, containing 35,040 samples in total, is divided into a training set (the first 70%, 24,528 samples) for model learning and parameter optimization, and a testing set (the last 30%, 10,512 samples) to evaluate the generalization ability and prediction performance of the models. To prevent data leakage, the hyperparameter optimization using RIME was conducted solely on the training set employing a 5-fold cross-validation strategy. The testing set (last 30%) was strictly reserved for the final performance evaluation and was not involved in the training or tuning process.

4. Analysis of Effects

To rigorously evaluate the effectiveness and robustness of the proposed RIME-VMD-RIME-TCN-BiGRU-SA framework, three comparative experiments were conducted on the testing dataset (the last 30% of samples). These experiments were designed to incrementally validate the contributions of the decomposition strategy, the optimization algorithm, and the overall model architecture.

4.1. Comparison of Different Signal Decomposition Prediction Methods

To comprehensively evaluate the efficacy of the proposed RIME-VMD strategy in processing non-stationary wind power signals, a series of comparative experiments were executed. The proposed framework was benchmarked against four established decomposition techniques: EMD, EEMD, CEEMDAN, and the standard VMD (configured with fixed parameters $K = 8$, $\alpha = 2000$). All decomposition methods were coupled with the identical TCN-BiGRU-SA predictor to ensure a fair comparison. The experimental outcomes are visualized in Figure 4 (prediction fitting curves), Figure 5 (error distribution curves), Figure 6 (statistical metric histograms), and Figure 7 (linear regression scatter plots). The precise numerical results are tabulated in Table 3. A detailed discussion is presented as follows:

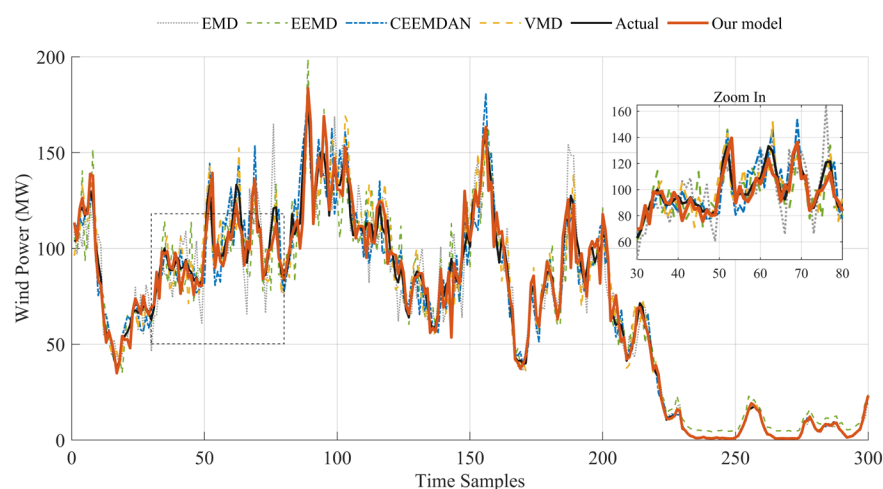


Figure 4. Prediction curves of different decomposition methods.

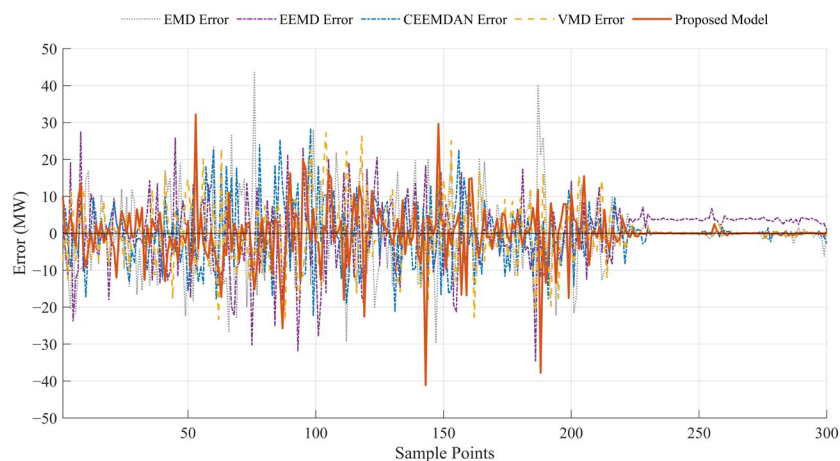


Figure 5. Prediction error curves of different decomposition methods.

Table 3. Prediction performance of different decomposition methods.

Models	RMSE ↓	MAE ↓	MAPE (%) ↓	R^2 ↑
EMD-R-T-B-SA	12.1204	7.8421	10.61	0.9472
EEMD-R-B-SA	10.9435	7.6970	9.45	0.9570
CEEMDAN-R-T-B-SA	9.3594	6.2778	8.38	0.9685
VMD-R-T-B-SA	9.9551	6.5925	8.84	0.9644
Our model	7.5340	4.3847	5.17	0.9796

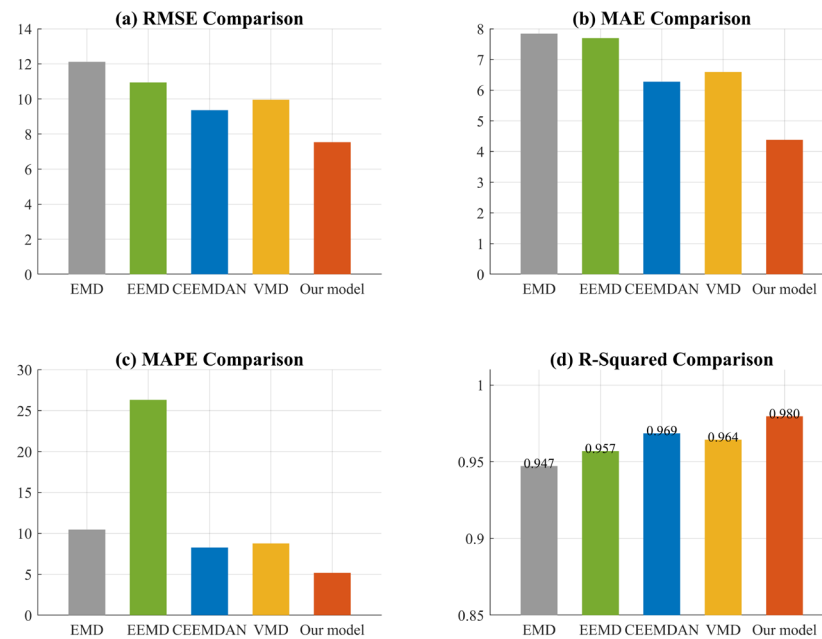


Figure 6. Performance metrics of different decomposition methods.

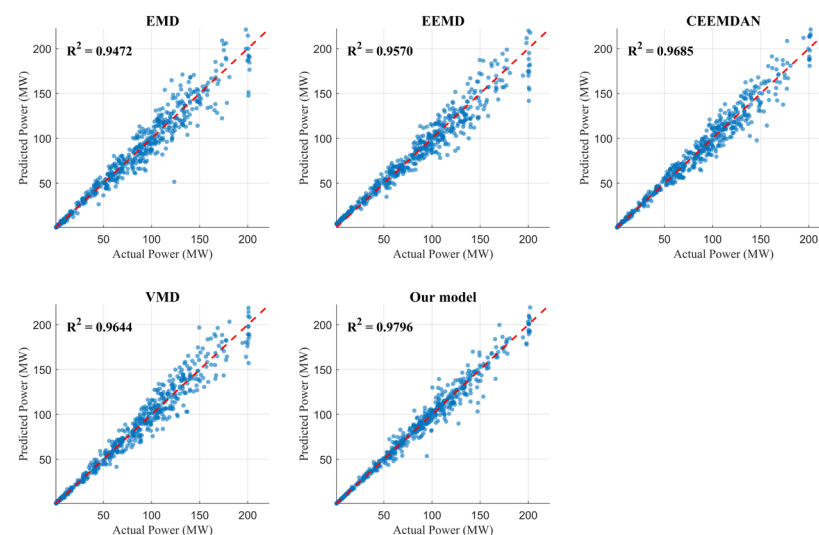


Figure 7. Scatter plots of predicted vs. actual power (decomposition comparison).

1. Analysis of Fitting Performance and Error Characteristics:

As depicted in Figure 4, the prediction trajectories generated by the EMD and EEMD-based models display a distinct “phase lag” phenomenon, especially during rapid fluctuations in wind power. This deficiency is primarily stemmed from the inherent “mode mixing” and boundary effects associated with recursive decomposition, which introduce aliasing noise into the signal components. Conversely, the prediction curve of the proposed model (highlighted in red) maintains high synchronization with the actual power series, effectively capturing both peak and valley trends.

Further analysis of Figure 5 reveals significant differences in error stability. The error curves for the CEEMDAN and standard VMD models exhibit considerable volatility, oscillating within the $[-20, 20]$ MW range. In sharp contrast, the error distribution of the proposed model is densely concentrated around the zero baseline, demonstrating superior stability. This evidence suggests that the adaptive RIME-VMD method successfully isolates high-frequency noise, thereby minimizing reconstruction errors in the hybrid forecasting system.

2. Quantitative Assessment of Prediction Accuracy:

Table 3 and Figure 6 provide a robust quantitative assessment of the model performance. The proposed framework outperforms all baseline methods across all evaluation metrics, achieving the lowest RMSE of 7.534 MW and MAE of 4.3847 MW. Two critical insights are derived from these statistics:

Advantage over Recursive Methods: When compared to the classical EMD and EEMD approaches, the proposed model reduces the RMSE by 37.84% and 31.16%, respectively. This substantial margin confirms that variational decomposition offers a theoretically stronger basis for handling nonlinear data than recursive shifting methods.

Impact of Parameter Optimization: Notably, the experiment reveals that the standard VMD model (RMSE = 9.9551) yielded slightly inferior results compared to CEEMDAN (RMSE = 9.3594). This anomaly highlights the limitations of relying on manual experience or default settings for VMD parameters (K, α). However, upon integrating the RIME optimization algorithm, the proposed model achieves a 24.32% improvement in RMSE over the standard VMD. This empirical evidence strongly validates that the dual-stage optimization mechanism is indispensable for maximizing the feature extraction capability of VMD.

3. Regression and Consistency Analysis:

Figure 7 illustrates the linear regression analysis between predicted and actual values, where the coefficient of determination (R^2) serves as a key indicator of goodness-of-fit. The scatter points for the EMD and EEMD models appear relatively dispersed, resulting in lower R^2 values of 0.9472 and 0.9570, respectively. While the CEEMDAN and standard VMD models improve the R^2 to approximately 0.96, the proposed model attains the highest R^2 of 0.9796. As observed in the plot, the scatter points of the proposed method are tightly clustered along the $y = x$ diagonal, indicating that the model maintains high consistency with ground truth data even under extreme wind regimes.

4.2. Experiment II: Evaluation of Optimization Algorithms

To validate the efficacy of the proposed RIME algorithm in optimizing the hyperparameters of the deep learning predictor (i.e., the number of neurons, learning rate, and iteration epochs), a second set of comparative experiments was designed. The RIME-optimized model was benchmarked against three widely used heuristic algorithms: Particle Swarm Optimization (PSO), Sparrow Search Algorithm (SSA), and Whale Optimization Algorithm (WOA). All models employed the identical RIME-VMD decomposition front-end and TCN-BiGRU-SA forecasting back-end (denoted as RV-PSO, RV-SSA, and RV-WOA, respectively). The experimental results are visualized in Figure 8 (prediction curves), Figure 9 (error distributions), Figure 10 (performance metrics), and Figure 11 (scatter plots). The detailed evaluation metrics are presented in Table 4. The comprehensive discussion is as follows:

Table 4. Prediction performance of different optimization algorithms.

Models	RMSE ↓	MAE ↓	MAPE (%) ↓	R^2 ↑
RV-PSO-T-B-SA	9.8155	6.7970	9.71	0.9654
RV-SSA-T-B-SA	10.1616	6.2993	8.21	0.9629
RV-WOA-T-B-SA	8.1119	5.1314	6.89	0.9764
Our model	7.5340	4.3847	5.17	0.9796

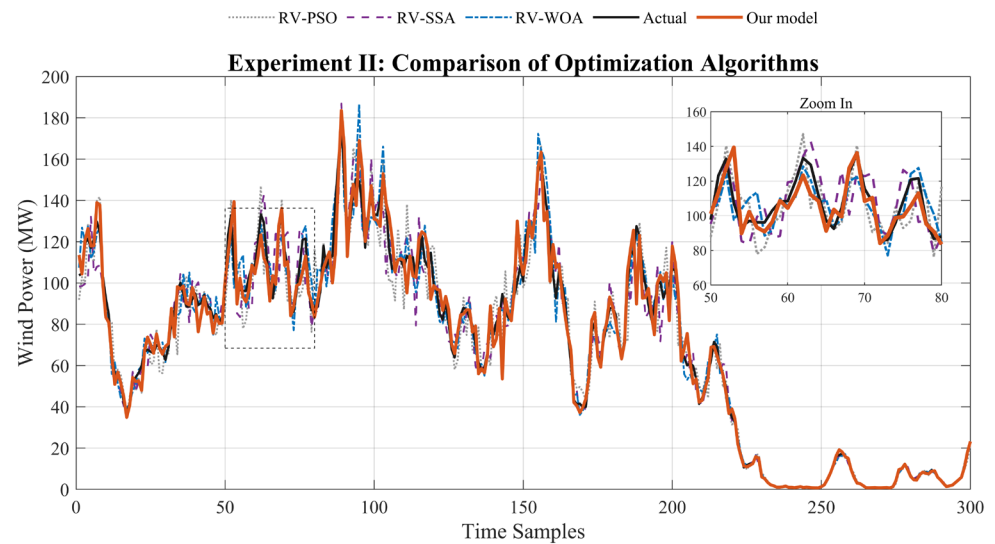


Figure 8. Prediction curves of different optimizers in Experiment II.

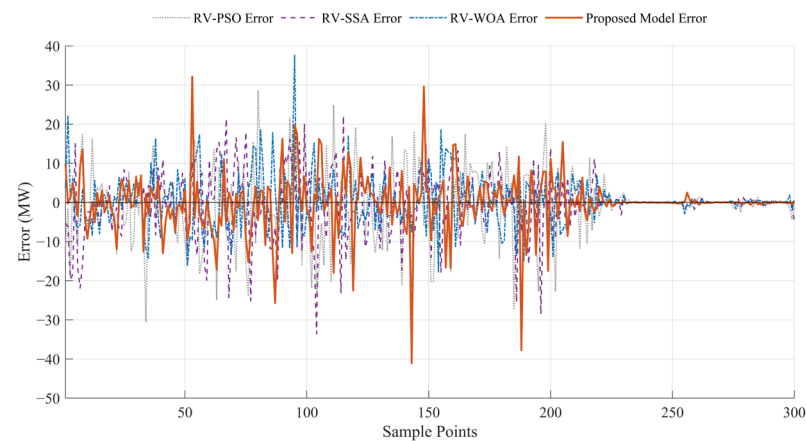


Figure 9. Prediction error curves of different optimizers in Experiment II.

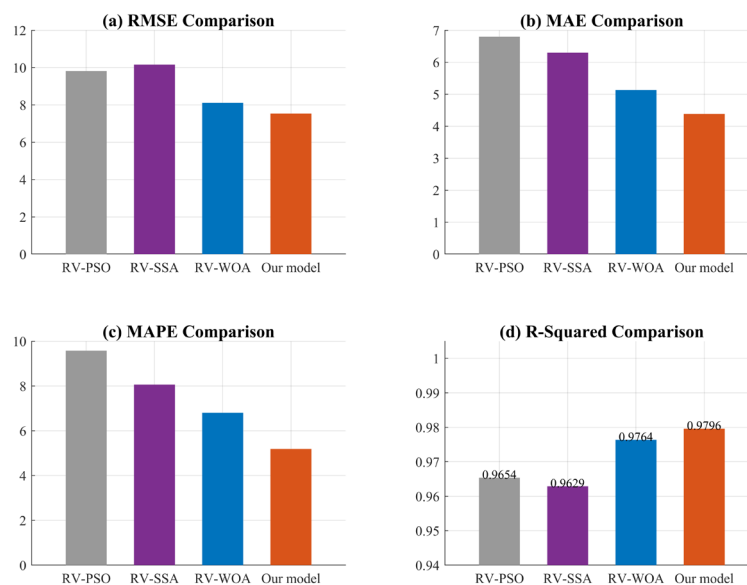


Figure 10. Performance metrics of different optimizers in Experiment II.

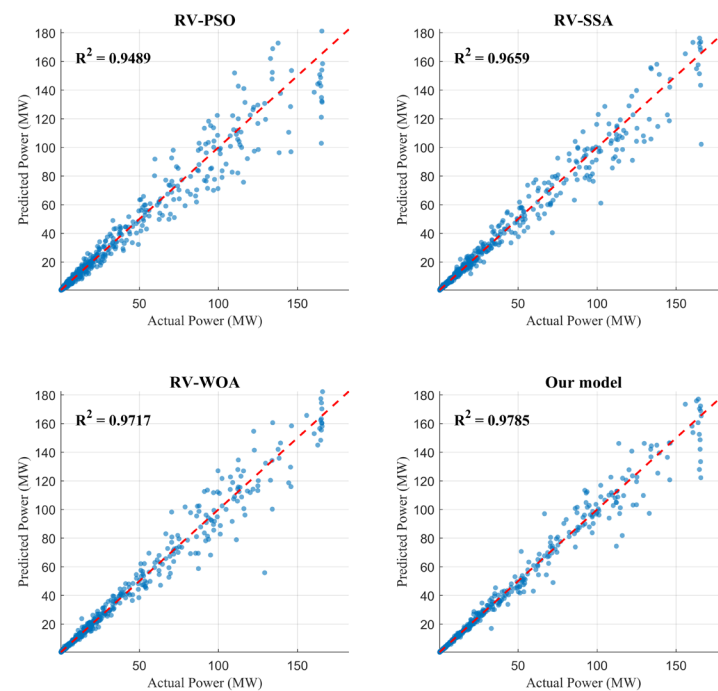


Figure 11. Scatter plots of predicted vs. actual power (optimizer comparison) in Experiment II.

1. Analysis of Prediction Stability and Error Convergence:

Figure 8 depicts the fitting curves of the different optimizer-based models. While all models capture the general trend of wind power generation, the PSO and SSA-optimized models exhibit noticeable deviations in high-frequency fluctuation zones. This suggests that these algorithms may have fallen into local optima during the hyperparameter search process, leading to suboptimal network configurations. In contrast, the prediction curve of the proposed Our Model (optimized by RIME) demonstrates the highest fidelity to the actual values.

The error distribution curves in Figure 9 further corroborate this observation. The error fluctuations of the RV-PSO and RV-SSA models are relatively broad, whereas the error curve of the proposed model is stable and confined within a narrow range around zero. This stability indicates that the RIME algorithm effectively balances global exploration and local exploitation, successfully identifying the global optimal hyperparameters for the hybrid network.

2. Quantitative Evaluation of Optimization Performance:

The statistical metrics in Figure 10 and Table 4 provide a direct comparison of forecasting accuracy. The proposed model achieves the lowest RMSE (7.534 MW) and MAE (4.3847 MW) among all candidates.

Comparison with Classical Algorithms: Compared to the classical PSO and SSA algorithms, the proposed model reduces the RMSE by 23.24% and 25.86%, respectively. This substantial improvement highlights the limitations of traditional swarm intelligence algorithms in handling high-dimensional parameter spaces.

Comparison with Advanced Algorithms: Even when compared to the competitive WOA algorithm (RMSE = 8.1119 MW), the proposed RIME-based strategy achieves a further RMSE reduction of 7.12%. This superiority is attributed to the unique “soft-rime” search mechanism of the RIME algorithm, which enables it to escape local stagnation more efficiently than the encircling prey mechanism of WOA.

3. Goodness-of-Fit Analysis:

Figure 11 illustrates the scatter plots of the predicted versus actual power values. The goodness of fit is quantified by the coefficient of determination (R^2). The RV-SSA

and RV-PSO models yield R^2 values of 0.9629 and 0.9654, respectively, showing a slight dispersion in the high-power region. The RV-WOA model performs better with an R^2 of 0.9764. However, the proposed model attains the highest R^2 of 0.9796, with scatter points strictly aligned along the diagonal. This confirms that the RIME-optimized framework possesses the strongest generalization capability and robustness across the entire testing period.

4.3. Experiment III: Analysis of the Proposed Model for Seasonal Prediction

Wind power generation is characterized by strong seasonal cyclicality and volatility, which poses significant challenges to the generalizability of forecasting models. To rigorously evaluate the robustness of the proposed framework across different climatic conditions, the full-year dataset was segmented into four seasonal subsets: Spring (March–May), Summer (June–August), Autumn (September–November), and Winter (December–February). Experiment III was conducted to test the model's performance on these subsets independently. The prediction curves and linear regression analysis for each season are visualized in Figure 12. The detailed evaluation metrics are summarized in Table 5. The comprehensive analysis is as follows:

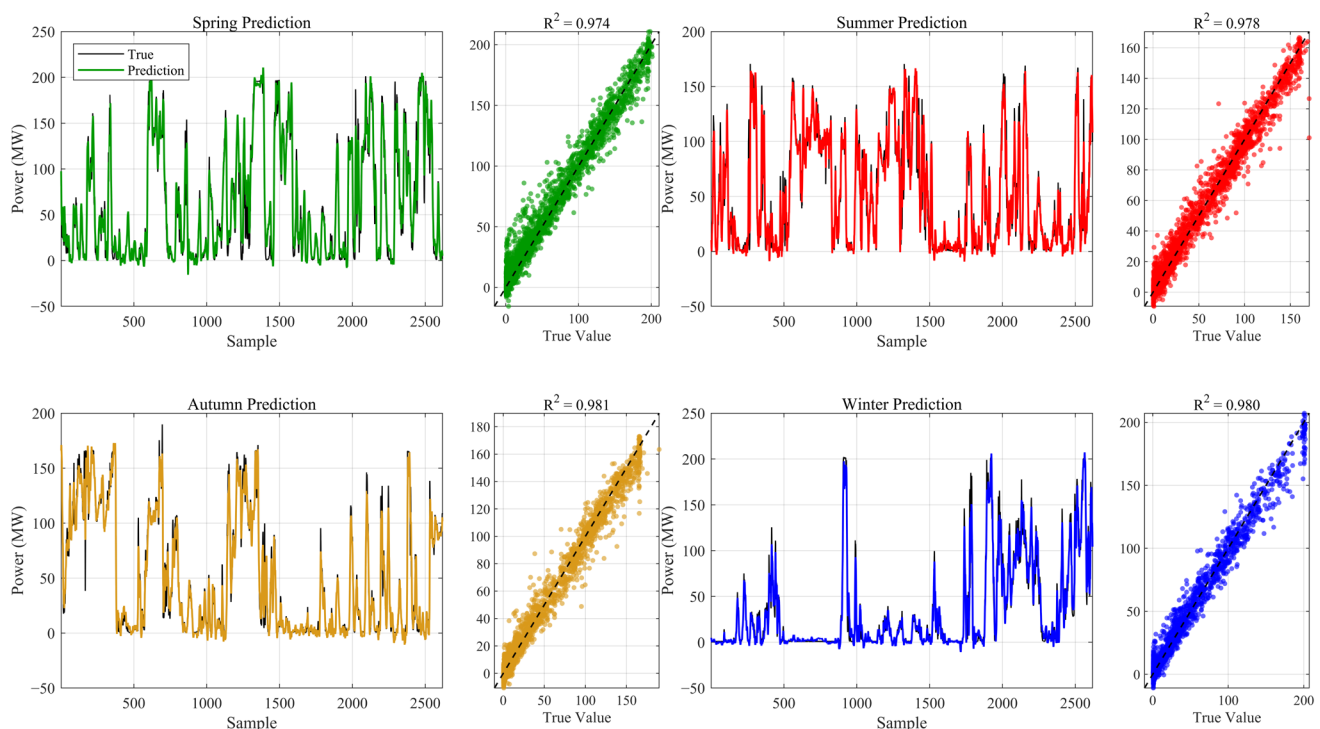


Figure 12. Seasonal performance evaluation of the proposed framework covering spring, summer, autumn, and winter. The left panel depicts the fitting trajectories of predicted versus observed wind power, while the right panel displays the coefficient of determination metric for each season.

Table 5. Prediction performance of different seasons.

Models	RMSE ↓	MAE ↓	MAPE (%) ↓	R^2 ↑
Spring	10.0579	7.0226	6.00	0.9745
Summer	7.3419	5.1569	6.24	0.9776
Autumn	7.3257	5.0098	6.11	0.9814
Winter	6.9878	4.4187	7.77	0.9801

1. Overall Seasonal Robustness:

As indicated in Table 5, the proposed model maintains high forecasting precision throughout the year, with R^2 values consistently exceeding 0.97. This confirms that the RIME-VMD-RIME-TCN-BiGRU-SA framework possesses excellent adaptability to seasonal climate variations and is not overfitted to specific wind patterns.

Performance in Stable Seasons (Summer and Autumn): In summer and autumn, the meteorological conditions are relatively stable. The model achieves an RMSE of 7.3419 MW and 7.3257 MW, respectively. The prediction curves in Figure 12 (Summer/Autumn panels) show that the model accurately captures the smooth diurnal cycles of wind power.

Performance in Volatile Seasons (Spring): It is observed that the prediction error in Spring is slightly higher (RMSE = 10.0579 MW). This phenomenon is attributed to the “seasonal transition effect,” where the rapid alternation of cold and warm air masses in the Xinjiang region leads to extreme stochastic fluctuations in wind speed. Despite this challenge, the model still maintains a high correlation ($R^2 = 0.9745$), demonstrating its reliability even under the most unstable conditions.

2. Exceptional Performance in Winter:

A remarkable finding from Table 5 is that the model achieves its best performance in Winter, with the lowest RMSE of 6.9878 MW and MAE of 4.4187 MW.

Reasoning: Although winter wind speeds are typically high, they often exhibit stronger persistence and clearer trend patterns compared to the chaotic turbulence in spring. The proposed TCN-BiGRU architecture, enhanced by the Self-Attention mechanism, effectively captures these long-term dependencies in high-wind scenarios.

Visual Evidence: As shown in the right panel of Figure 12 (Winter), the scatter points are most tightly clustered around the regression line ($R^2 = 0.9801$), indicating that the proposed model is particularly effective in predicting high-capacity power generation, which is critical for grid peak shaving during the winter heating season.

3. Summary:

The seasonal analysis confirms that the proposed framework is not only accurate on average but also robust across diverse meteorological scenarios. The adaptive feature extraction capability of RIME-VMD ensures that the model can automatically adjust to the distinct frequency characteristics of seasonal wind signals without manual parameter retuning.

4.4. Experiment IV: Comparison with Baseline Models

To comprehensively evaluate the advanced nature and generalization capability of the proposed framework, Experiment IV was designed to compare the proposed model with eight mainstream baseline models: BPNN, LSTM, GRU, BiLSTM, CNN, TCN, ACBiLSTM, and Transformer.

Furthermore, to verify that the model’s performance is not coincidental to a specific region, a second dataset (Dataset 2) was introduced from a wind farm in Gansu Province, China. This dataset shares the same feature structure as Dataset 1 but operates under different meteorological conditions.

Prior to model training, the Spearman correlation coefficient analysis was performed on Dataset 2, as illustrated in Figure 13. Based on the correlation magnitude, the top eight features with the highest relevance to power output were selected as input variables, ensuring consistency with the feature engineering process of Dataset 1.

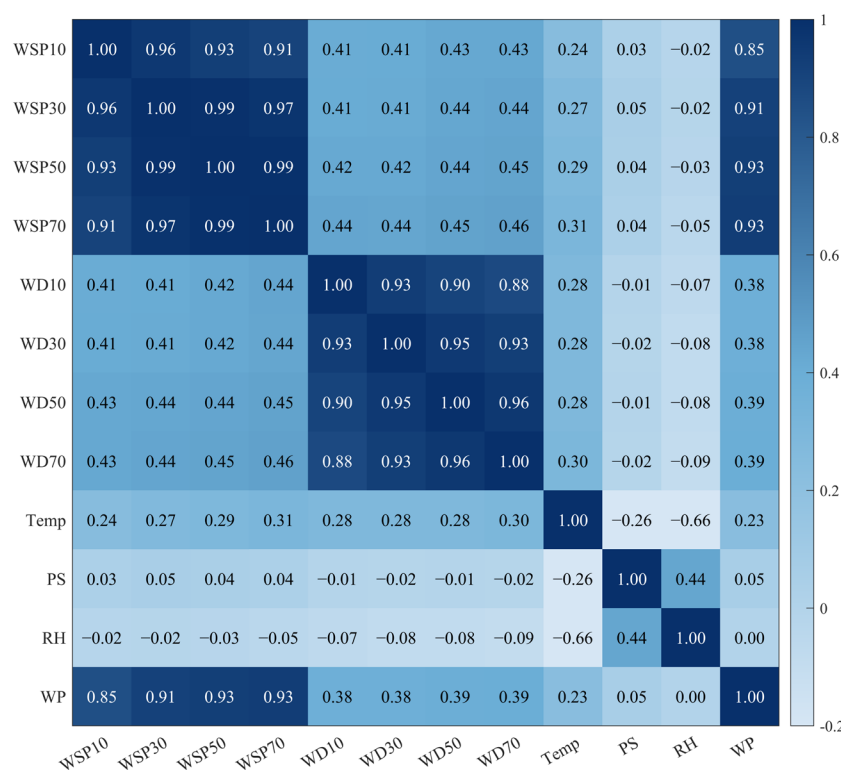


Figure 13. Heat map for wind power Dataset 2 based on Spearman.

The prediction results for both datasets are visualized in Figure 14 (Dataset 1 curves) and Figure 15 (Dataset 2 curves), while the comprehensive performance metrics are summarized in Table 6 and Figure 16. The detailed analysis is as follows:

Table 6. Detailed description of the collected wind power related datasets.

Models	Dataset 1			Dataset 2		
	RMSE ↓	MAE ↓	R^2 ↑	RMSE ↓	MAE ↓	R^2 ↑
BPNN	19.1205	12.7821	0.8686	9.5128	6.1743	0.8656
LSTM	14.5158	9.9738	0.9243	7.6526	4.8671	0.9189
GRU	13.7764	9.2711	0.9318	8.2464	5.5678	0.9045
BiLSTM	11.6417	7.5752	0.9513	6.2137	3.8537	0.9325
CNN	14.2504	9.4608	0.9270	7.1053	4.4352	0.9211
TCN	12.1787	7.8640	0.9467	6.0106	3.6576	0.9443
ACBiLSTM	10.2691	6.6818	0.9621	5.1435	2.9821	0.9632
Transformer	9.2273	6.1155	0.9694	5.0537	2.9702	0.9627
Our model	7.5340	4.3847	0.9796	3.8833	2.5654	0.9753

1. Performance Analysis on Dataset 1 (Xinjiang):

As shown in Table 6 and Figure 14, the proposed model exhibits superior predictive accuracy compared to all baselines on the Xinjiang dataset.

Comparison with Traditional Deep Learning: The basic models (BPNN, LSTM, GRU, CNN) struggle to track rapid wind fluctuations, yielding high RMSE values (e.g., BPNN RMSE = 19.12 MW). The proposed model significantly outperforms these methods, reducing the RMSE by over 60%.

Comparison with State-of-the-Art Models: Even against advanced predictors like the Transformer (RMSE = 9.2273 MW), the proposed RIME-VMD-RIME-TCN-BiGRU-SA framework achieves a remarkable improvement, reducing the RMSE by 18.35% to

7.534 MW. This validates that the dual-optimization strategy effectively unlocks the potential of the hybrid network, surpassing the capabilities of a standalone self-attention model.

2. Generalization Analysis on Dataset 2 (Gansu):

The results on the Gansu dataset, presented in Figure 15 and the right column of Table 6, provide strong evidence of the model's robustness.

Consistency: The proposed model achieves the best performance on Dataset 2 as well, with the lowest RMSE of 3.8833 MW and MAE of 2.5654 MW, achieving an R^2 of 0.9753.

Superiority in Generalization: While the baseline models also show improved performance on Dataset 2 (due to potentially simpler wind patterns in Gansu), the proposed model still maintains a significant lead. Specifically, it reduces the prediction error (RMSE) by 23.16% compared to the Transformer (5.0537 MW) and by 37.50% compared to BiLSTM (6.2137 MW).

Conclusion: The consistent superiority across two geographically distinct datasets confirms that the proposed framework possesses excellent geographical transferability. It can adaptively capture the distinct spatiotemporal characteristics of different wind farms without requiring manual structural adjustments.

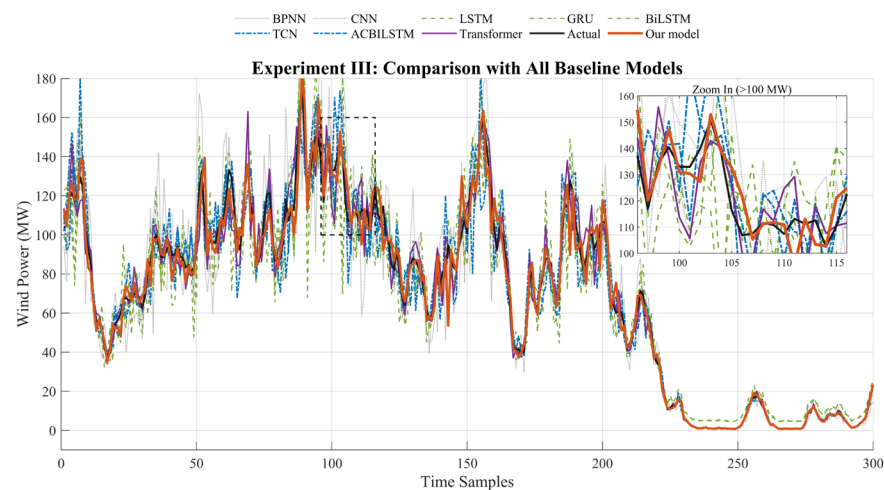


Figure 14. Prediction curves of different baseline models in dataset 1.

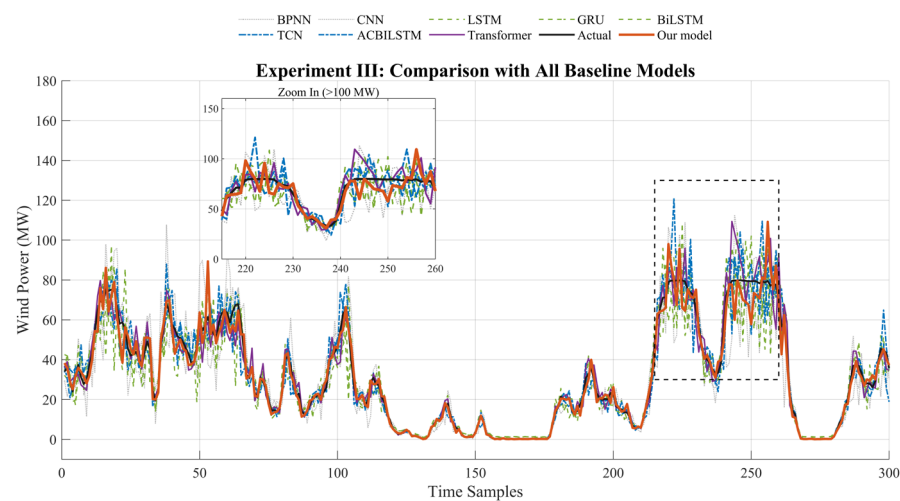


Figure 15. Prediction curves of different baseline models in dataset 2.

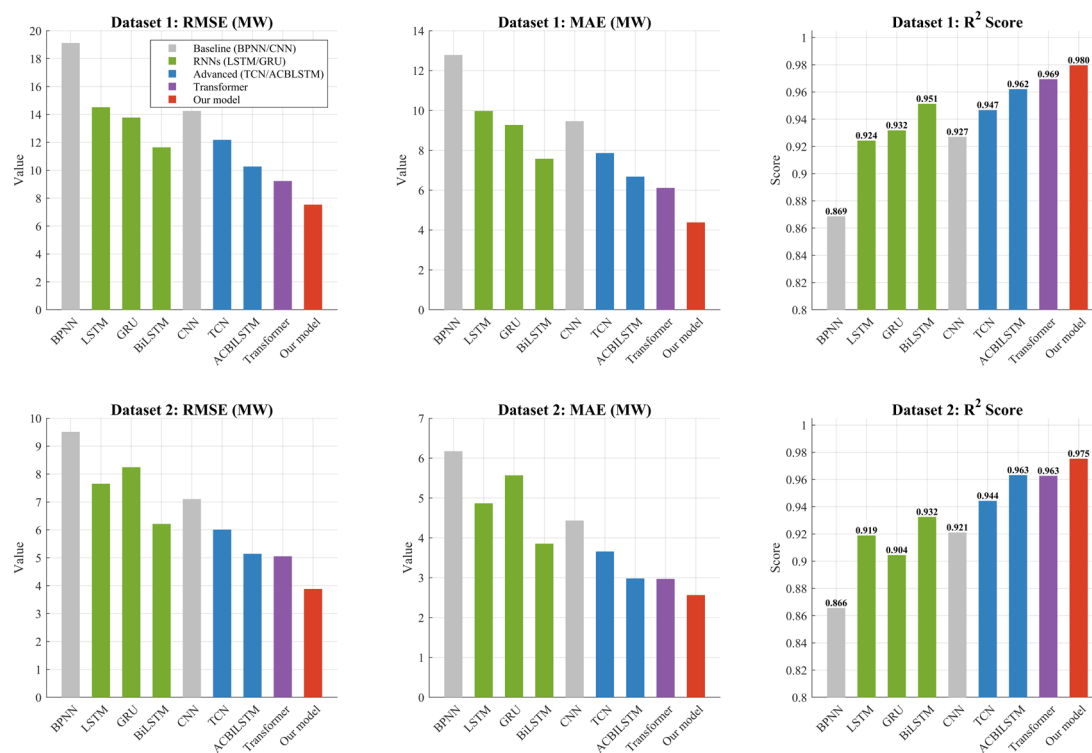


Figure 16. Performance metrics of different baseline models.

3. Overall Evaluation (Figure 16):

Figure 15 presents a radar chart comparison of the evaluation metrics. It is visually evident that the proposed model (represented by the outermost envelope) encompasses the largest area, indicating comprehensive dominance in terms of RMSE, MAE, and R^2 . The experimental results conclusively demonstrate that the integration of RIME-optimized VMD decomposition and the attention-based TCN-BiGRU network creates a robust, high-precision forecasting system suitable for diverse wind power scenarios.

4. Analysis of Computational Cost:

Table 7 compares the training efficiency of the proposed framework against baseline models.

Table 7. Comparison of training computational costs among different models (on Dataset 1).

Models	Training Time (s)
BPNN	215.4
LSTM	342.8
GRU	850.2
BiLSTM	985.6
CNN	1120.5
TCN	765.8
ACBiLSTM	1180.3
Transformer	1250.9
Our model	5420.5

Baseline Efficiency: Shallow models (BPNN, CNN) converge rapidly within 350 s, while deep learning baselines (e.g., Transformer, BiLSTM) generally require 1000–1300 s due to their complex network structures.

Cost of Optimization: The proposed framework consumes 5420.5 s (approximately 1.5 h). This increased time is the necessary cost of the RIME algorithm's iterative search

process, which ensures the global optimality of both VMD decomposition parameters (K, α) and predictor hyperparameters.

Feasibility Justification: While the training cost is higher, it is critical to distinguish between “offline training” and “online inference.” In practical engineering, model retraining is typically performed daily or weekly on high-performance servers. Once the optimal parameters are fixed, the online inference time for a single prediction step is negligible (milliseconds), fully satisfying the real-time dispatch requirements of power grids (typically on a 15 min rolling basis). Thus, the proposed strategy effectively exchanges “offline computing time” for “online prediction precision.”

4.5. Ablation Study on the Proposed Approachs

To strictly quantify the contribution of each individual module (i.e., the RIME-VMD decomposition strategy, the RIME hyperparameter optimization, and the Self-Attention mechanism) to the final predictive performance, a comprehensive ablation study was conducted. By systematically removing or replacing specific components, five variant models were constructed and compared against the full Proposed Model. The experimental results on both Dataset 1 and Dataset 2 are summarized in Table 8.

Table 8. Ablation results of the proposed model at two wind farms.

Component				Dataset 1			Dataset 2		
RIME-VMD	RIME	Predictor	Model	RMSE ↓	MAE ↓	R^2 ↑	RMSE ↓	MAE ↓	R^2 ↑
✓	✓	TCN-BiGRU-SA	Proposed model	7.5340	4.3847	0.9796	3.8833	2.5654	0.9753
×	✓	TCN-BiGRU-SA	Rime-TCN-BiGRU-SA	14.3012	8.4587	0.9372	6.9943	4.4556	0.9287
×	×	TCN-BiGRU-SA	TCN-BiGRU-SA	15.6433	9.5634	0.9436	7.0153	4.6383	0.9156
✓	×	TCN-BiGRU-SA	RV-TCN-BiGRU-SA	12.1063	8.4642	0.9545	6.8243	4.0231	0.9355
✓	✓	TCN-BiGRU	RV-RIME-TCN-BiGRU	10.3628	7.4896	0.9583	5.4922	3.2836	0.9536
✓	×	TCN-BiGRU	RV-TCN-BiGRU	13.7266	10.4893	0.9433	5.8626	3.6846	0.9551

Note: ✓ indicates the addition of the component and × indicates the elimination of the component.

1. **Impact of Decomposition Strategy (RIME-VMD):** The comparison between the Proposed Model and the Rime-TCN-BiGRU-SA (without decomposition) reveals the critical role of signal processing. As shown in Table 7, removing the RIME-VMD module leads to a drastic degradation in performance, with the RMSE on Dataset 1 increasing from 7.534 MW to 14.3012 MW. This indicates that the raw wind power series contains excessive stochastic volatility that deep learning models struggle to learn directly. The proposed adaptive decomposition effectively acts as a “filter,” stabilizing the input features and enhancing learnability.

2. **Impact of Hyperparameter Optimization (RIME):** The necessity of the dual-stage optimization is evidenced by comparing the Proposed Model with the RV-TCN-BiGRU-SA (where RIME optimization for hyperparameters is removed). Without the precise tuning of the TCN-BiGRU network structure, the RMSE deteriorates significantly to 12.1063 MW on Dataset 1. This confirms that even with high-quality decomposed data, a manually configured network is insufficient. The RIME algorithm successfully identifies the global optimal parameters, ensuring the predictor operates at its peak capacity.

3. **Impact of Self-Attention Mechanism (SA):** To verify the efficacy of the attention module, the RV-RIME-TCN-BiGRU variant (where SA is removed) was tested. Although this model performs reasonably well (RMSE = 10.3628 MW), it still lags behind the proposed model. The integration of the Self-Attention mechanism reduces the RMSE by specific margins (27.3% improvement on Dataset 1), demonstrating its ability to dynamically weigh critical time steps and focus on high-impact historical information, thereby capturing sudden ramp events more accurately.

4. **Overall Synergy:** The most basic variant, TCN-BiGRU-SA (without any decomposition or RIME optimization), yields the worst performance with an RMSE of 15.6433 MW.

By sequentially integrating VMD decomposition, RIME optimization, and the Attention mechanism, the prediction error is progressively reduced. The Proposed Model achieves the lowest error metrics across both datasets (RMSE = 7.534/3.8833 MW), proving that the combination of these modules creates a strong positive synergistic effect rather than a redundant accumulation of complexity.

4.6. Discussion: Mechanisms, Limitations, and Practical Implications

While the experimental results in Section 4.5 confirm the superior performance of the proposed RIME-VMD-RIME-TCN-BiGRU-SA framework, it is essential to critically analyze the underlying mechanisms, acknowledge specific limitations, and evaluate the practical value for grid operations.

4.6.1. Mechanism of Superiority

The performance leap achieved by the proposed framework is not merely a result of stacking complex modules but stems from the specific topological match between the RIME algorithm and the wind power optimization landscape. High-dimensional hyperparameter spaces in hybrid deep learning models often form non-convex surfaces with numerous local optima. Traditional algorithms like PSO tend to lose population diversity rapidly, leading to the premature convergence observed in the early iterations of Figure 8. In contrast, RIME's "Soft-rime search" strategy maintains exploration diversity in the early stages, while the "Hard-rime puncture" mechanism enables aggressive exploitation of promising regions in later stages. This balance allows the framework to lock onto global optima that maximize the feature extraction capability of the TCN-BiGRU architecture, as evidenced by the stable error distribution in Figure 9.

4.6.2. Critical Analysis of Limitations

Despite its high accuracy, this study identifies two primary limitations that must be addressed in future practical deployments:

Computational Cost vs. Response Time: As detailed in Table 7, the dual-optimization process incurs a high computational cost, with a training time of 5420.5 s (approx. 1.5 h)—nearly 4.3 times that of the Transformer model (1250.9 s). While this latency is acceptable for day-ahead scheduling (where models are retrained offline), it poses a bottleneck for ultra-short-term forecasting scenarios (e.g., 5 min rolling dispatch) that require rapid model updates. Future work could explore "Transfer Learning" to fine-tune the pre-trained model for new data, significantly reducing the need for full retraining.

Sensitivity to Seasonal Volatility: Although the model demonstrates excellent robustness in winter (RMSE = 6.99 MW), the performance dips in spring (RMSE = 10.06 MW), as shown in Table 5. This seasonal discrepancy suggests that while the VMD strategy effectively decomposes general non-stationarity, it may struggle to fully capture the extreme, high-frequency "burst" signals characteristic of the transitional spring season in Xinjiang. This indicates a trade-off between signal smoothing (for stability) and sensitivity to extreme ramp events.

4.6.3. Practical Implications for Power Systems

From the perspective of grid operators, the statistical improvements translate directly into operational benefits. The reduction in RMSE by 1.69 MW compared to the Transformer model (Dataset 1, Table 6) implies that for a 200 MW wind farm, the required spinning reserve capacity can be reduced, directly lowering the operational costs of thermal backup units. Furthermore, the highly concentrated error distribution allows dispatchers to set narrower confidence intervals for power generation plans, thereby improving the utilization rate of transmission corridors and reducing renewable energy curtailment.

5. Conclusions

To address the challenges of non-stationarity and hyperparameter sensitivity in short-term wind power prediction, this paper proposed a novel dual-optimized hybrid framework: RIME-VMD-RIME-TCN-BiGRU-SA. Through rigorous theoretical derivation and extensive empirical validation on two distinct full-year datasets (Xinjiang and Gansu), the study draws the following meaningful conclusions:

1. **Significance of Adaptive Decomposition:** The RIME-optimized VMD strategy serves as a critical pre-processing filter. Ablation studies confirm that removing this module leads to a drastic deterioration in prediction accuracy, with the RMSE on Dataset 1 increasing from 7.53 MW to 14.30 MW. This quantifies the necessity of decomposing stochastic wind signals into stationary sub-components to unlock the learning potential of deep neural networks.
2. **Robustness via Dual-Optimization:** The proposed “Dual-Optimization” paradigm ensures end-to-end global optimality. Compared to single-stage optimization or standard baselines, the framework demonstrates superior geographical transferability. It achieved consistent high-precision performance across both the volatile Xinjiang dataset ($R^2 = 0.9796$) and the distinct Gansu dataset ($R^2 = 0.9753$), reducing the RMSE by 18.35% and 23.16%, respectively, compared to the state-of-the-art Transformer model.
3. **Engineering Value vs. Computational Cost:** From an engineering perspective, the proposed model provides a reliable decision-support tool for grid dispatch. However, this precision comes at a cost: the dual-optimization process requires a training time of 5420.5 s, which is significantly higher than conventional models. This trade-off suggests that the framework is currently best suited for day-ahead scheduling or scenarios where offline retraining is feasible, rather than minute-level real-time online learning.
4. **Future Directions:** While the model excels in winter scenarios (RMSE = 6.99 MW), its performance in transitional seasons like spring (RMSE = 10.06 MW) indicates room for improvement in handling extreme ramp events. Future work will focus on two directions: (1) developing a lightweight RIME variant to accelerate convergence and reduce training costs; and (2) incorporating transfer learning strategies to enhance the model’s adaptability to rapid seasonal regime shifts without extensive retraining.

Author Contributions: Conceptualization, Z.W. and K.Z.; methodology, Z.W.; software, S.L.; validation, Z.W., K.Z., X.A., S.L. and T.Z.; formal analysis, Z.W.; investigation, S.L.; resources, K.Z.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, K.Z.; visualization, S.L.; supervision, K.Z.; project administration, K.Z.; funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the PhD Research StartUp Foundation of Hubei University of Technology, grant number BSDQ2020023.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editor for their constructive comments that significantly improved the quality of this manuscript. The authors also appreciate the technical support provided by the School of Electrical and Electronic Engineering at Hubei University of Technology. During the preparation of this manuscript, the authors used ChatGPT for English language editing and formatting assistance. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

BiGRU	Bidirectional Gated Recurrent Unit
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Function
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NWP	Numerical Weather Prediction
PS	Atmospheric Pressure
PSO	Particle Swarm Optimization
RF	Random Forest
RH	Relative Humidity
RIME	RIME Optimization Algorithm
RMSE	Root Mean Square Error
SA	Self-Attention Mechanism
SSA	Sparrow Search Algorithm
TCN	Temporal Convolutional Network
TEMP	Temperature
VMD	Variational Mode Decomposition
WD	Wind Direction
WOA	Whale Optimization Algorithm
WSP	Wind Speed

References

1. Global Wind Energy Council. *Global Wind Report 2024*; Global Wind Energy Council: Lisbon, Portugal, 2024.
2. Peng, X.; Chen, Y.; Cheng, K.; Wang, H.; Zhao, Y.; Wang, B.; Che, J.; Liu, C.; Wen, J.; Lu, C.; et al. Wind Power Prediction for Wind Farm Clusters Based on the Multifeature Similarity Matching Method. *IEEE Trans. Ind. Appl.* **2020**, *56*, 4679–4688. [CrossRef]
3. Ji, Z.; Qin, J.; Cheng, K.; Zhang, S.; Wang, Z. A Comprehensive Evaluation of Ducted Fan Hybrid Engines Integrated with Fuel Cells for Sustainable Aviation. *Renew. Sustain. Energy Rev.* **2023**, *185*, 113567. [CrossRef]
4. Sinsel, S.R.; Riemke, R.L.; Hoffmann, V.H. Challenges and Solution Technologies for the Integration of Variable Renewable Energy Sources—A Review. *Renew. Energy* **2020**, *145*, 2271–2285. [CrossRef]
5. Bird, L.; Cochran, J.; Wang, X. *Wind and Solar Energy Curtailment: Experience and Practices in the United States*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2014.
6. Renewable Energy and Jobs: Annual Review 2023. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Sep/IRENA_Renewable_energy_and_jobs_2023.pdf (accessed on 27 January 2026).
7. Zappa, W.; Junginger, M.; van den Broek, M. Is a 100% Renewable European Power System Feasible by 2050? *Appl. Energy* **2019**, *233–234*, 1027–1050. [CrossRef]
8. Sweeney, C.; Bramer, P.; Haupt, S.E. *Improving Wind Energy Forecasting Through Numerical Weather Prediction Model Development*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2019.
9. Bauer, P.; Thorpe, A.; Brunet, G. The Quiet Revolution of Numerical Weather Prediction. *Nature* **2015**, *525*, 47–55. [CrossRef]
10. Arslan Tuncar, E.; Sağlam, Ş.; Oral, B. A Review of Short-Term Wind Power Generation Forecasting Methods in Recent Technological Trends. *Energy Reports* **2024**, *12*, 197–209. [CrossRef]
11. Cadenas, E.; Rivera, W. Wind Speed Forecasting in Three Different Regions of Mexico, Using a Hybrid ARIMA–ANN Model. *Renew. Energy* **2010**, *35*, 2732–2738. [CrossRef]
12. Al-qaness, M.A.A.; Ewees, A.A.; Aseeri, A.O.; Abd Elaziz, M. Wind Power Forecasting Using Optimized LSTM by Attraction–Repulsion Optimization Algorithm. *Ain Shams Eng. J.* **2024**, *15*, 103150. [CrossRef]
13. Kumar, P.S.; Nookesh, V.N. Wind Speed Prediction Using Deep Learning—LSTM and GRU. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; IEEE: Bengaluru, India, 2021; pp. 662–667.
14. Memarzadeh, G.; Keynia, F. A New Short-Term Wind Speed Forecasting Method Based on Fine-Tuned LSTM Neural Network and Optimal Input Sets. *Energy Convers. Manag.* **2020**, *213*, 112824. [CrossRef]

15. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
16. Wu, B.; Wang, L.; Zhang, T. TCNFormer: Temporal Convolutional Network Former for Short-Term Wind Speed Forecasting. *arXiv* **2024**, arXiv:2408.15737. [[CrossRef](#)]
17. Sun, G.; Li, C.; Zhang, Y. Wind Power Forecasting Based on New Hybrid Model with TCN Residual Modification. *Energy AI* **2022**, *10*, 100199. [[CrossRef](#)]
18. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271. [[CrossRef](#)]
19. Chi, D.; Yang, C. Wind Power Prediction Based on WT-BiGRU-Attention-TCN Model. *Front. Energy Res.* **2023**, *11*, 1156007. [[CrossRef](#)]
20. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
21. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies* **2020**, *13*, 3764. [[CrossRef](#)]
22. Liang, Y.; Zhang, D.; Zhang, J.; Hu, G. A State-of-the-Art Analysis on Decomposition Method for Short-Term Wind Speed Forecasting Using LSTM and a Novel Hybrid Deep Learning Model. *Energy* **2024**, *313*, 133826. [[CrossRef](#)]
23. Battula, V.V.R.; K., P.; K., S. Comparison of Advanced Signal Decomposition Techniques for the Classification of PQDs by Machine Learning Algorithms. *Sci. Rep.* **2025**, *15*, 38164. [[CrossRef](#)]
24. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. London Ser. A* **1998**, *454*, 903–995. [[CrossRef](#)]
25. Du, M.; Zhang, Z.; Ji, C. Prediction for Coastal Wind Speed Based on Improved Variational Mode Decomposition and Recurrent Neural Network. *Energies* **2025**, *18*, 542. [[CrossRef](#)]
26. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
27. Liang, T.; Lu, H.; Sun, H. Application of Parameter Optimized Variational Mode Decomposition Method in Fault Feature Extraction of Rolling Bearing. *Entropy* **2021**, *23*, 520. [[CrossRef](#)]
28. Zhang, X.; Miao, Q.; Zhang, H.; Wang, L. A Parameter-Adaptive VMD Method Based on Grasshopper Optimization Algorithm to Analyze Vibration Signals from Rotating Machinery. *Mech. Syst. Signal Process.* **2018**, *108*, 58–72. [[CrossRef](#)]
29. Sun, X.; Liu, H. Multivariate Short-Term Wind Speed Prediction Based on PSO-VMD-SE-ICEEMDAN Two-Stage Decomposition and Att-S2S. *Energy* **2024**, *305*, 132228. [[CrossRef](#)]
30. Wang, Y.; Pei, L.; Li, W.; Zhao, Y.; Shan, Y. Short-Term Wind Power Prediction Method Based on Multivariate Signal Decomposition and RIME Optimization Algorithm. *Expert Syst. Appl.* **2025**, *259*, 125376. [[CrossRef](#)]
31. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [[CrossRef](#)]
32. Su, H.; Zhao, D.; Heidari, A.A.; Liu, L.; Zhang, X.; Mafarja, M.; Chen, H. RIME: A Physics-Based Optimization. *Neurocomputing* **2023**, *532*, 183–214. [[CrossRef](#)]
33. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime Mould Algorithm: A New Method for Stochastic Optimization. *Future Gener. Comput. Syst.* **2020**, *111*, 300–323. [[CrossRef](#)]
34. Kisvari, A.; Lin, Z.; Liu, X. Wind Power Forecasting—A Data-Driven Method along with Gated Recurrent Neural Network. *Renew. Energy* **2021**, *163*, 1895–1909. [[CrossRef](#)]

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