



ARTICLE

Wind Power Forecasting Utilizing Bidirectional Gated Recurrent Units in Conjunction with Empirical Mode Decomposition and Bayesian Neural Networks

Xiaolan Li^{1,2} and Yanting Wang^{1,2,*}

¹School of Electrical and Information Technology, Yunnan Minzu University, Kunming, China

²Yunnan Key Laboratory of Unmanned Autonomous System, Kunming, China

*Corresponding Author: Yanting Wang. Email: 19372088482@163.com

Received: 20 November 2025; Accepted: 19 January 2026; Published: 18 June 2026

ABSTRACT: To address the operational challenges of power systems with high renewable penetration, this research targets the non-stationarity and stochasticity of wind power. A novel hybrid framework for probabilistic forecasting and risk assessment is proposed. Initially, Empirical Mode Decomposition (EMD) adaptively decomposes the raw power signal into multi-scale Intrinsic Mode Functions (IMFs) and a residual trend, effectively segregating temporal features and reducing complexity. These components are then fused with historical data to form a comprehensive input. The core predictor is a Bidirectional Gated Recurrent Unit (BiGRU) network enhanced with a Temporal Attention (TA) mechanism. The BiGRU captures bidirectional long-term dependencies, while the TA mechanism dynamically focuses on the most influential historical time steps, enabling precise temporal pattern extraction. To quantify uncertainty, a Bayesian Neural Networks (BNNs) layer is integrated, transforming deterministic point forecasts into probabilistic outputs with prediction intervals. Finally, leveraging these probabilistic forecasts, the Value at Risk (VaR) metric is applied to assess potential operational risks under specified confidence levels, translating uncertainty into quantifiable reliability or financial risk. Simulation results confirm the framework's superiority, achieving a normalized Root Mean Square Error (nRMSE) of 15.73% and a normalized Mean Absolute Error (nMAE) of 10.94%, significantly outperforming benchmarks. The innovative integration of signal processing, attentive deep learning, Bayesian inference, and risk theory within a unified model enhances forecasting accuracy, quantifies uncertainty, and enables proactive risk assessment, providing robust decision support for grid dispatch and renewable integration.

KEYWORDS: Short-term wind power forecasting; empirical mode decomposition (EMD); bidirectional gated recurrent unit (Bi-GRU); temporal attention mechanism; Bayesian neural networks (BNNs); predictive uncertainty quantification

1 Introduction

1.1 Background and Significance of Wind Power Forecasting

Renewable energy sources accounted for 29.9% of global electricity generation in 2023, producing 8928 terawatt-hours (TWh), while the remaining 70.1% (20,939 TWh) originated from fossil fuels, nuclear energy, pumped storage, and other non-renewables. This brought the worldwide electricity output across all sources to 29,867 TWh for the year. Simultaneously, the global energy structure is undergoing a rapid transition towards low-carbon alternatives, with wind power emerging as a leading renewable energy source due to its technological maturity and commercial viability. According to IRENA data for 2022, global cumulative wind

power capacity surpassed 900 GW, contributing approximately 24% [1] of total renewable energy generation. However, achieving further integration of wind energy remains challenging due to its inherent variability and intermittency, which introduce significant operational uncertainties in power output and grid scheduling—one of the key challenges widely recognized in renewable energy system research [2–4].

1.2 Existing Forecasting Methods and Technological Limitations

Wind power forecasting methodologies are broadly categorized into physical and statistical approaches [5]. Physical methods, including meteorological and wind field models, utilize Numerical Weather Prediction (NWP) [5] and Geographic Information System (GIS) data to simulate atmospheric processes—among which the characteristics of the atmospheric boundary layer (a key factor affecting near-surface wind fields) play a crucial role in improving simulation accuracy, and recent studies have begun to integrate machine learning techniques to optimize the modeling of such meteorological processes [6]—and topographical effects. While effective for long-term forecasts, these methods require high-quality input data and extensive computational resources [7]. Statistical methods encompass time series models such as ARMA and SARIMA, which rely on historical data patterns [5], as well as traditional machine learning techniques like Support Vector Machine (SVM) that have been applied to distributed wind power forecasting [8]; such data-driven approaches are also widely used in other renewable energy modeling [9], but are constrained by data quality and seasonal variations.

Recent advances in hybrid forecasting models combining signal decomposition techniques with deep learning architectures have demonstrated significant improvements in wind power forecasting—this “decomposition-integration” framework, which shares logical similarities with the two-stage forecasting paradigm (e.g., decomposition + error compensation) validated in time-series prediction tasks [10], outperforming LSTM and SVM in accuracy by approximately 12.5% [11]. Nevertheless, current research often neglects multi-time-scale predictions, seasonal variations, and dynamic feature weighting through attention mechanisms. Moreover, predictive uncertainty and model optimization issues remain poorly addressed in the context of grid scheduling support.

As an important renewable energy source, the inherent variability and intermittency of wind power pose significant challenges to the stable operation of power grids [4,11]. To address these challenges, accurate wind power forecasting is essential. Traditional forecasting methods, such as physical and statistical approaches, although capable of providing certain prediction capabilities, often suffer from limitations including high data quality requirements, substantial computational resource consumption, and difficulties in capturing nonlinear patterns [12].

1.3 Integrated Forecasting and Uncertainty Analysis Framework

To improve the accuracy and reliability of short-term wind power prediction, this study proposes a hybrid framework that integrates Empirical Mode Decomposition (EMD), Bidirectional Gated Recurrent Unit (Bi-GRU), temporal attention mechanism, and Bayesian Neural Network (BNN). First, the proposed method preprocesses the original non-stationary power signals using EMD [13–18]. As an adaptive signal processing technique, EMD can decompose the data into a limited number of stationary Intrinsic Mode Functions (IMFs) in a self-adaptive manner, thereby disentangling complex temporal variations. It is particularly suitable for decomposing non-stationary signals such as wind power time-series data into IMFs [5,19,20], which enhances data analysis, feature extraction, and prediction performance. Subsequently, Hilbert transform analysis is performed on each IMF to extract physically meaningful time-frequency features, including instantaneous amplitude and frequency [13]. These features, combined with historical power data, are used as inputs to the Bi-GRU network to capture bidirectional temporal dependencies

[21–24]. By integrating bidirectional processing and temporal attention mechanism, the Bi-GRU network can handle uneven temporal correlations in the sequence (e.g., wind fluctuations driven by meteorological conditions) [24,25]. This enables the network to capture contextual dependencies and dynamically assign learnable weights to the features of key time steps, thereby enhancing the extraction of meteorological patterns and power fluctuation patterns. The attention mechanism allows the model to dynamically weight critical time intervals, which is crucial for identifying significant meteorological events and power fluctuation patterns in wind power prediction. Finally, the probabilistic output layer of the BNNs replaces deterministic prediction with Monte Carlo Dropout sampling to quantify the uncertainty of predictions [24,25]. This integrated method, combined with non-parametric kernel density estimation and Value at Risk (VaR) theory, generates dynamic confidence intervals to support improved power grid dispatching and risk management decisions. Deep learning models, especially variants of Recurrent Neural Networks (RNNs) such as Bi-GRU, have demonstrated excellent performance in short-term wind power prediction. For instance, compared with traditional Long Short-Term Memory (LSTM) and Support Vector Machine (SVM), the Bi-GRU model can better capture long-term dependencies when processing sequence data and, in some cases, achieve higher prediction accuracy [26]. This performance enhancement is attributed to the Bi-GRU's capability to process sequence information bidirectionally simultaneously, thereby capturing more comprehensive contextual features. Notably, building upon the Bi-GRU's temporal modeling prowess, Bayesian regularization neural networks (BRNNs) exhibit notable superiority in complex system modeling—specifically in boosting model generalization and mitigating overfitting by integrating probabilistic prior knowledge into the training process. These merits make them particularly well-suited for electrical engineering applications involving uncertain and noisy time-series data, such as wind power signals. For instance, the study [27] on the mon-keypox transmission model employed BRNNs to optimize radial basis deep neural networks; the Bayesian regularization mechanism effectively constrained the model parameter space, enhancing the robustness of epidemic transmission predictions even with sparse and volatile surveillance data. This application paradigm is highly transferable to electrical engineering: in wind power prediction, BRNNs refine temporal sequence forecasting models by suppressing interference from random wind speed fluctuations; in power grid fault diagnosis, they facilitate the establishment of reliable mapping relationships between monitoring signals and fault types, reducing misdiagnoses caused by sensor noise. Inspired by these successful applications, this paper incorporates Bayesian regularization principles into the probabilistic output layer design, capitalizing on its prowess in uncertainty quantification to address the limitations of traditional deterministic wind power prediction models. This integration complements the Bi-GRU's temporal capture capability, thus laying a solid foundation for power grid dispatching and renewable energy integration decisions.

1.4 Main Contributions and Innovations

The main contributions and innovations of this paper are as follows:

(1) The EMD-Hilbert coupling framework is proposed. Through the collaboration of adaptive decomposition and multi-scale time-frequency feature extraction, it breaks through the traditional limitations and improves the interpretability and accuracy of wind power prediction.

(2) A Bi-GRU-TA coupling architecture is proposed, which captures the full context through bidirectional temporal coding and combines dynamic optimization of feature weights to enhance the robustness of wind power prediction under complex weather conditions and cope with extreme weather and fluctuation disturbances.

(3) Design a Bayesian probability output layer, extend deterministic prediction to probability distribution, build an end-to-end uncertainty learning chain, solve the pain point of risk quantification, and provide a basis for power grid dispatching.

These innovations synergistically address the non-stationarity, randomness, and uncertainty issues in wind power prediction, bridging the gap between theoretical modeling and practical operational needs of power systems.

2 Basic Theory

2.1 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is an adaptive and fully data-driven signal processing technology. The aim is to decompose complex non-stationary signals into a set of Intrinsic Mode Functions (IMFs) [28]. Unlike traditional decomposition methods that rely on predefined basis functions, EMD directly adaptively extracts oscillation modes from the data [15], making it particularly suitable for analyzing actual signals such as wind power time series where frequencies vary over time [17,18]. Each IMF must meet two basic conditions: First, the number of extreme points and the number of zero-crossing points must differ by at most one. Second, at any given moment, the average value of its upper and lower envelope lines is zero. The decomposition process of EMD is called sifting. By iteratively extracting the remaining highest-frequency oscillation components in the signal, different intrinsic oscillation modes can be effectively separated without making prior assumptions about the signal structure. This adaptability enables it to handle common nonlinear and non-stationary phenomena in wind power, such as turbulent fluctuations, diurnal variation patterns, and weather-induced variations [13,16,29]. The IMF obtained through decomposition ranges from fine-scale irregular fluctuations to long-term trends, providing multi-scale representations for the original signal and facilitating the improvement of feature extraction and prediction accuracy in subsequent modeling [14,19,30]. The EMD algorithm decomposes the signal $x(t)$ into a series of IMFs and a residual through an iterative screening process. The specific steps are as follows:

(1) For a given signal $x(t)$, identify all its local extremum points (maximum and minimum values).

(2) Through the extreme value point interpolation usually USES cubic spline interpolation on the building envelope $e(t)$ Max and envelope under $e(t)$ min.

(3) Calculate the local average envelope

$$m(t) = \sigma (W_m \cdot [e_{\max}(t), e_{\min}(t)] + b_m) \quad (1)$$

(4) Subtract the average envelope from the original signal to obtain proto-IMF:

$$h(t) = \sigma (W_h \cdot [x(t), m(t)] + b_h) \quad (2)$$

(5) Check whether $h(t)$ meets the IMF conditions. If not, repeat the screening process as a new signal until the conditions are met.

(6) Extract the IMF that meets the conditions and denote it as $c_i(t)$, and define the residuals as:

$$r(t) = \sigma (W_r \cdot [x(t), c_i(t)] + b_r) \quad (3)$$

(7) Repeat the above process for the residual $r(t)$ until the residual becomes a monotonic function or the IMF can no longer be extracted. The original signal can be reconstructed by all IMF and the final residuals:

$$x(t) = \sigma (W_x \cdot [c_1(t), c_2(t), \dots, c_N(t), r_N(t)] + b_x) \quad (4)$$

Here N represents the total number of extracted IMFs, and $r_N(t)$ is the final residual representing the signal trend. As shown in Figs. 1–5, this adaptive decomposition capability enables EMD to separate

non-stationary wind power signals into oscillation components with clear physical meanings, laying the foundation for subsequent analysis.

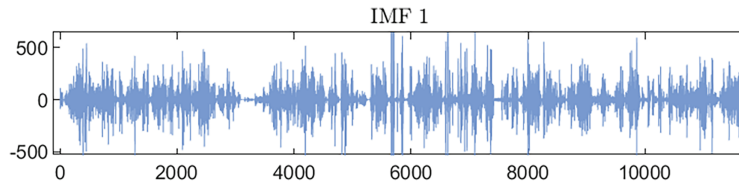


Figure 1: Intrinsic mode function 1 (IMF1).

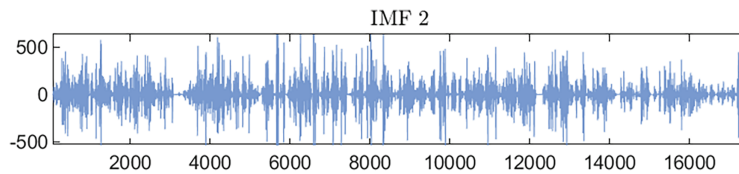


Figure 2: Intrinsic mode function 2 (IMF2).

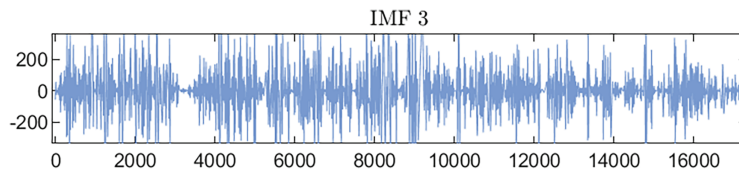


Figure 3: Intrinsic mode function 3 (IMF3).

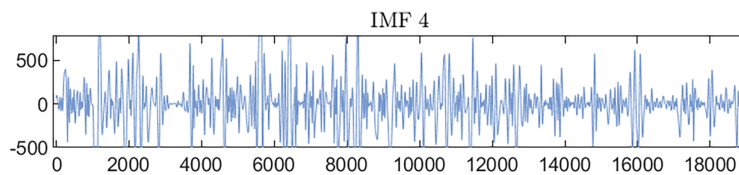


Figure 4: Intrinsic mode function 4 (IMF4).

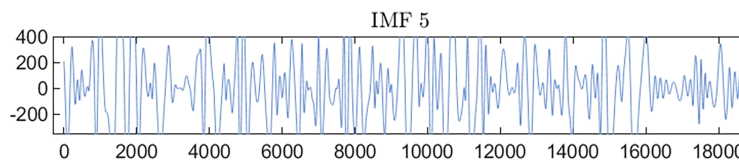


Figure 5: Intrinsic mode function 5 (IMF5).

2.2 Bidirectional Gated Recurrent Unit

2.2.1 Core Formulas of Unidirectional GRU

At each time step t , the GRU regulates the information flow through an Update Gate and a Reset Gate. The Reset Gate r_t determines whether to ignore the information from the historical hidden state, with an

output range of $[0, 1]$ —a value closer to 0 indicates a stronger tendency to ignore historical information, and its calculation formula is as follows:

$$\hat{r}_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (5)$$

The Update Gate z_t balances the retention ratio of the historical hidden state and the update ratio of new information, with an output range of $[0, 1]$ —a value closer to 1 means a greater inclination to retain historical information, and its formula is:

$$\hat{z}_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (6)$$

The Candidate Hidden State \tilde{h}_t computes a candidate state containing new information based on the historical state filtered by the Reset Gate and the current input, with an output range of $[-1, 1]$, and the corresponding formula is:

$$\hat{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \quad (7)$$

The Final Hidden State h_t fuses the historical hidden state and the candidate hidden state through the Update Gate to obtain the final hidden state at the current time step, and its calculation formula is:

$$\hat{h}_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (8)$$

Here are the parameter explanations: x_t is the input vector at time step t ; h_{t-1} is the hidden state at time step $t - 1$; W_r , W_z , W_h are learnable weight matrices; b_r , b_z , b_h are learnable bias vectors; σ denotes the sigmoid activation function (output range $[0, 1]$); \tanh represents the hyperbolic tangent activation function (output range $[-1, 1]$); \odot stands for element-wise multiplication (Hadamard product); and $[h_{t-1}, x_t]$ indicates the concatenation of two vectors.

2.2.2 Bidirectional Fusion Mechanism of BiGRU

The BiGRU is composed of two independent GRUs with opposite directions, which capture temporal information from the forward and reverse directions of the sequence, respectively. It finally obtains features containing complete temporal context through feature fusion, as illustrated in Fig. 6. The Forward GRU processes the sequence in chronological order (from $t = 1$ to $t = T$) and outputs the forward hidden state sequence \vec{h}_t ($t = 1, 2, \dots, T$), focusing on the impact of historical information on the current time step. The Backward GRU processes the sequence in reverse chronological order (from $t = T$ to $t = 1$) and outputs the backward hidden state sequence \overleftarrow{h}_t ($t = 1, 2, \dots, T$), focusing on the impact of future information on the current time step. The final bidirectional hidden state h_t^{BiGRU} of BiGRU at time step t is usually obtained by concatenating the forward and backward hidden states, thereby containing both forward and reverse temporal information, and its calculation formula is:

$$h_t^{\text{BiGRU}} = \left[\vec{h}_t, \overleftarrow{h}_t \right] \quad (9)$$

For an input sequence of length T , the output of BiGRU is a bidirectional hidden state sequence $\{h_1^{\text{BiGRU}}, h_2^{\text{BiGRU}}, \dots, h_T^{\text{BiGRU}}\}$, which can be directly applied to downstream tasks such as classification and sequence labeling in electrical engineering-related time series analysis, such as power load forecasting and fault diagnosis of electrical equipment.

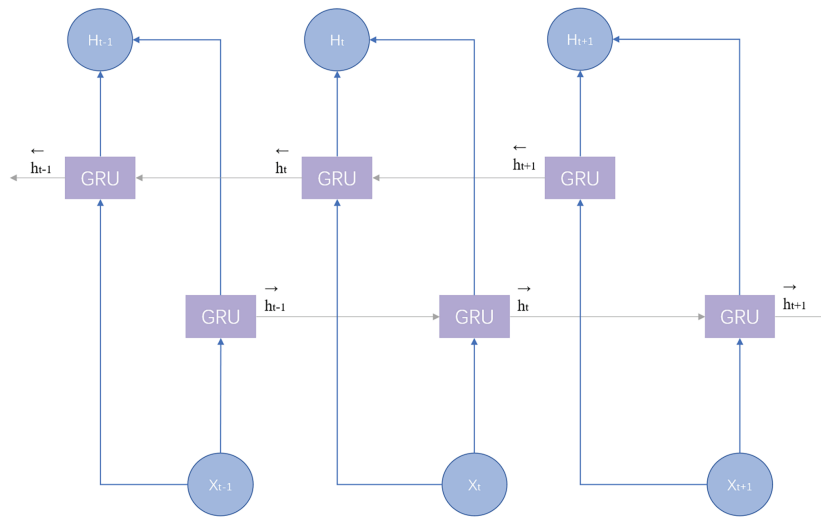


Figure 6: BiGRU structure.

2.3 Bayesian Neural Networks

Bayesian Neural Networks (BNNs) incorporate probability distributions into their weights, extending traditional deterministic neural networks to a probabilistic framework, as illustrated in Fig. 7. Their core advantage lies in quantifying the uncertainty of prediction, which is of great significance in power scenarios such as wind power prediction and grid dispatching—an urgent demand highlighted in the review of wind power forecasting methods [31]. Unlike traditional models that only output point predictions, BNNs use weighted posterior inference to output confidence intervals or predicted probability distributions, providing reliable basis for decisions such as grid reserve capacity configuration. To adapt to power time series data, BNNs can integrate numerical weather forecasts (NWP) with historical data. NWP has been widely proven effective in day-ahead wind power forecasting [32], and its performance can be further enhanced through bias correction technologies [33]; meanwhile, SCADA data (a typical type of historical power data) has been validated as a reliable input for wind power machine learning models [34]. BNNs approximate weight posterior through variational inference or Markov chain Monte Carlo (MCMC), and by integrating signal processing techniques such as Ensemble Empirical Mode Decomposition (EEMD)—a method that has demonstrated value in simplifying wind power signal complexity [35]—a “decomposition–probability prediction” architecture is formed, which can further enhance the prediction accuracy. For problems with high computational complexity, lightweight solutions such as Monte Carlo dropout (MC Dropout) can balance accuracy and efficiency.

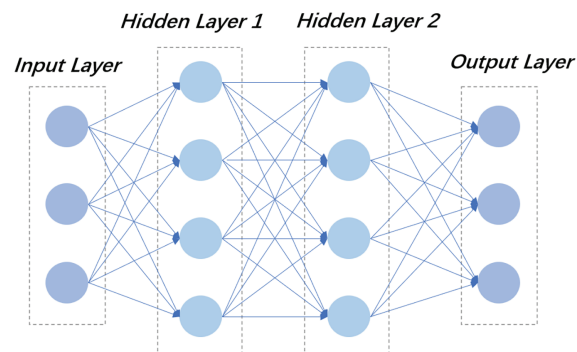


Figure 7: BNNs structure

2.3.1 Core Definition and Formula

(1) Dataset definition

The input and output dataset \mathcal{D} contains N samples, among which the input sample $x_i \in \mathbb{R}^D$ and the output label $y_i \in \mathbb{R}^m$:

$$\hat{\mathcal{D}} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (10)$$

(2) Network parameter definition The set of all learnable parameters of the L -layer neural network θ , including the weight matrices and bias vectors of each layer:

$$\hat{\theta} = \{W_1, b_1, \dots, W_L, b_L\} \quad (11)$$

Here, \hat{W}_l and \hat{b}_l respectively represent the weight matrix and bias vector of the l th layer.

2.3.2 Core Components of Bayesian Neural Network

(1) Prior distribution $p(\theta)$

The prior distribution encodes the initial belief about the weights, usually adopting a zero-mean Gaussian distribution, and the joint prior is the product of the priors of each parameter:

$$p(\hat{\theta}) = \prod_{l=1}^L p(W_l)p(b_l) \quad (12)$$

$$\hat{W}_l \sim \mathcal{N}(0, \sigma_p^2 I) \quad (13)$$

$$\hat{b}_l \sim \mathcal{N}(0, \sigma_p^2 I) \quad (14)$$

Here, σ_p^2 represents the prior variance, and I is the identity matrix.

(2) Likelihood function

It describes the probability of observing output y_i given input x_i and parameter θ . For wind power prediction (a regression task), y_i follows a Gaussian distribution:

$$p(y_i | \hat{x}_i, \theta) = \mathcal{N}(y_i | f(x_i; \theta), \sigma_l^2 I) \quad (15)$$

where $f(x_i; \theta)$ is the predicted value, and σ_l^2 represents the likelihood variance.

(3) Classification tasks (such as discrete output scenarios like equipment fault diagnosis):

$$p(y_i | \hat{x}_i, \theta) = \text{Cat}(y_i | \text{softmax}(f_\theta(x_i))) \quad (16)$$

Here, $f_\theta(x_i)$ represents the deterministic output of the Bayesian neural network, σ_l^2 is the likelihood variance, and $\text{Cat}(\cdot)$ indicates the classification distribution. The joint likelihood of the entire dataset is:

$$p(\hat{\mathcal{D}} | \theta) = \prod_{i=1}^N p(y_i | x_i, \theta) \quad (17)$$

(4) Posterior distribution of weights

According to Bayes' theorem, the updated weight belief distribution based on the data \mathcal{D} is:

$$p(\hat{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \quad (18)$$

The denominator $p(\mathcal{D})$ represents marginal likelihood (evidence), which is usually difficult to calculate directly in deep networks:

$$p(\mathcal{D}) = \int p(\mathcal{D}|\theta)p(\theta)d\theta \quad (19)$$

(5) Prediction distribution

The prediction of the new input x^* requires integrating all possible weight configurations rather than relying on a single set of weights:

$$p(y^*|\hat{x}^*, \mathcal{D}) = \int p(y^*|x^*, \theta)p(\theta|\mathcal{D})d\theta \quad (20)$$

2.3.3 Posterior Approximation Method

Monte Carlo Dropout (MCDO): Treating Dropout as a Bayesian prior, the prediction is obtained by averaging multiple Dropout samplings:

$$p(y^*|x^*, \mathcal{D}) \approx \frac{1}{S} \sum_{s=1}^S p(y^*|x^*, \theta_s) \quad (21)$$

In the formula, θ_s represents the weight samples of the Dropout network, and S is the number of samplings.

3 Comprehensive Prediction Method

3.1 The BiGRU-TA Wind Power Prediction Model

The BiGRU-TA architecture is an advanced hybrid deep learning framework tailored for complex non-stationary wind power time series, as depicted in Fig. 8. By integrating bidirectional sequence processing with adaptive time series feature weighting capabilities, it significantly enhances the predictive modeling performance of wind power [36]. This model precisely aligns with the dynamic influence of meteorological conditions on wind power, providing an efficient and reliable technical solution for short-term wind power prediction.

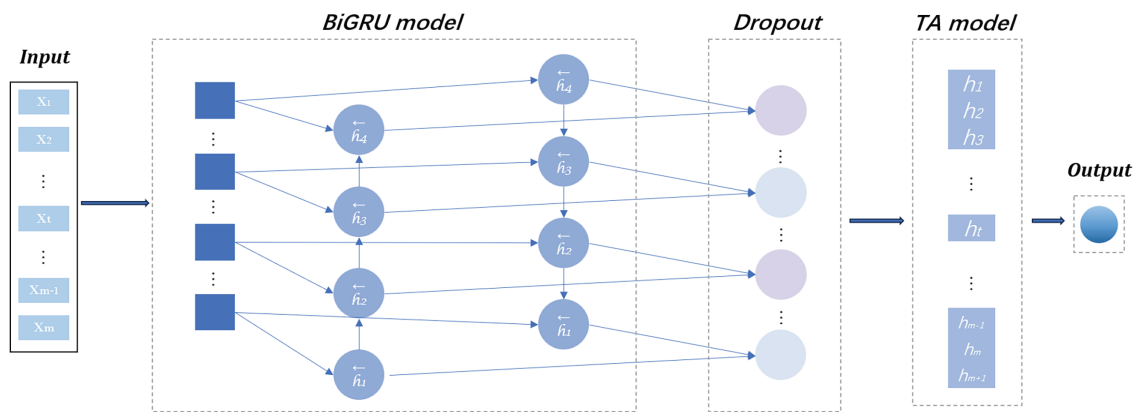


Figure 8: Structure of CNN-BiGRU optimized by temporal attention mechanism.

3.1.1 Core Advantages of the Model and Design Logic

Wind power time series exhibit distinct characteristics of strong volatility, intermittency, and non-stationarity. Events such as sudden wind speed changes, wind turbine start-up/shutdown, and weather system passages lead to significant differences in the importance of information across different time steps, which traditional Recurrent Neural Networks (RNNs) struggle to capture accurately. The BiGRU-TA model addresses this issue through the collaborative design of two core components:

(1) Bidirectional processing advantage: Compared with traditional long Short-Term memory networks (LSTM), BiGRU, through a simple mechanism of update gates and reset gates, more efficiently alleviates the problem of vanishing or explosion gradients, while maintaining lower computational complexity and higher training efficiency [37]. The forward GRU mines historical power trends and meteorological evolution patterns in chronological order (from $t = 1$ to T), while the reverse GRU captures future correlation information in reverse order (from $t = T$ to $t = 1$). The bidirectional fusion enables the model to comprehensively consider the previous trends and subsequent evolution characteristics when predicting the current power output.

(2) Attention focusing ability: The temporal attention mechanism dynamically allocates attention resources [14,28] for different time steps by calculating the learnable weights between hidden states. For critical periods such as seasonal transitions and extreme weather that have a significant impact on wind power, the model automatically assigns higher weights, effectively enhancing the ability to capture local anomalies and long-term dependencies, while avoiding interference from irrelevant information. In addition, this model has a unique advantage of interpretability—through the visualization of attention weights, it can clearly locate the historical periods or meteorological variables that play a decisive role in specific prediction tasks, providing transparent and reliable basis for power grid dispatching decisions. This has significant practical value in the wind power grid connection dispatching scenario.

3.1.2 Model Structure and Prediction Process

The model takes feature input to bidirectional time series modeling to attention weighting to prediction output as its core process, with each link closely connected to meet the requirements of wind power prediction:

(1) Multi-source feature fusion input: The model input layer integrates two types of core features—one is the historical wind power series, covering power output data at different time granularities, reflecting the short-term fluctuations and periodic patterns of wind power; The second type is numerical weather forecast

(NWP) data, which includes key meteorological parameters such as wind speed, wind direction, and air density, making up for the deficiency of single power data in depicting meteorological driving factors. After all the features are standardized, they are organized into the input tensors of the adapted model in time steps.

(2) BiGRU temporal feature extraction: After the multi-source fusion features are sent into the BiGRU layer, the forward and reverse GRU processes the sequences in parallel, respectively, and outputs the forward hidden state sequence \vec{h}_t and the reverse hidden state sequence \overleftarrow{h}_t . The bidirectional hidden state $h_t^{\text{BiGRU}} = [\vec{h}_t, \overleftarrow{h}_t]$ containing the complete temporal context is obtained through vector stitching, laying the foundation for subsequent feature screening.

(3) Time-attention weighted optimization: After the bidirectional hidden state sequence enters the temporal attention layer, the model generates the correlation score e_t through the trainable interaction between the query vector (query) and the key vector (key), and then obtains the attention weight α_t through softmax normalization. Ultimately, the context vector c_t is generated by weighted aggregation of value vectors (value), which condenses the time series information [11] that has the greatest influence on the current prediction.

(4) Predicted output and performance enhancement: After the context vector is processed by the fully connected layer and the activation function, the predicted value of wind power is output. To further enhance the robustness of prediction, in practical applications, signal preprocessing techniques are often combined to form a decomposition-prediction paradigm. For instance, after denoising the original wind speed signal using variational modal decomposition (VMD), BiGRU models the temporal dynamics of each mode, and XGBoost is used for residual correction. Significantly improve the data quality of NWP and the accuracy of final power prediction [37].

3.1.3 Model Application Performance and Limitations

In short-term wind power prediction practice, this model has demonstrated significant performance advantages: compared with classic models such as independent LSTM and Support vector machine (SVM), its ability to integrate NWP data with dynamic feature selection of attention can more accurately adapt to the inherent volatility and intermittency of wind resources, effectively reducing prediction errors. Compared with advanced architectures such as CNN-ABiLSTM and Transformer, BiGRU maintains high prediction accuracy while having better computational efficiency, and is particularly suitable for real-time power grid dispatching scenarios [21]. In complex climate environments such as hot deserts, hybrid models represented by VMD-BiGRU significantly reduce prediction errors through hierarchical modeling, verifying the effectiveness of the collaborative optimization of signal preprocessing and deep time series modeling [38]. It is worth noting that BiGRU performs well in short-term predictions, but it still has limitations when dealing with multi-time-scale changes, seasonal transitions, and extreme weather events. This study will further integrate Bayesian uncertainty quantification to enhance the model's support capacity for smart grids [39].

3.2 BNNs Uncertainty Prediction Model

As the uncertainty quantification module of the hybrid prediction framework, the BNN-based uncertainty model takes the output of the BiGRU-TA model as its input and quantifies prediction uncertainty through a Bayesian probability framework, as depicted in Fig. 9.

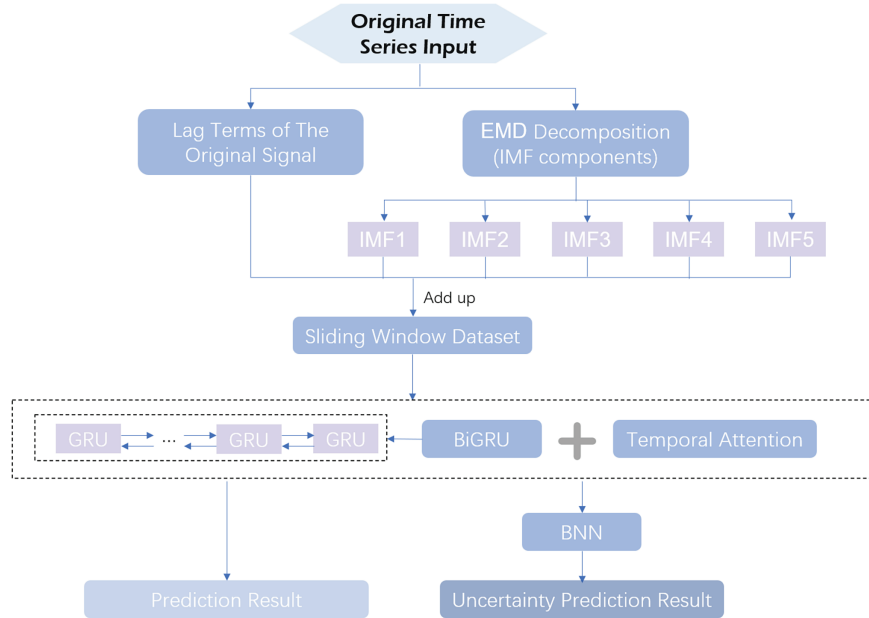


Figure 9: BiGRU-TA-BNN flowchart.

This solves the critical limitation of traditional point prediction models, which only provide a single predicted value without risk references—an essential function for improving the safety and economy of power grid operations. The model employs the Monte Carlo Dropout (MCDO) method as its core, simplifying the complex posterior inference process of Bayesian Neural Networks (BNNs) without significant computational overhead, making it suitable for engineering scenarios with real-time requirements. In terms of structure, the model is built on a fully connected network, with the Dropout mechanism introduced in hidden layers. Within the Bayesian framework, this Dropout mechanism is equivalent to a prior distribution of network weights—each weight has a certain probability of being set to zero during training, analogous to sampling weights from a specific prior distribution. The input is the high-dimensional predictive feature vector output by the BiGRU-TA model (containing rich temporal features and key wind power patterns), while the output is the parameter of the wind power probability distribution, providing a direct basis for subsequent uncertainty quantification.

The model's training process is tightly integrated with the MCDO method to achieve implicit learning of weight distributions. During training, the Dropout probability of hidden layers is set to 0.2 and remains active throughout. This ensures each weight has a 20% probability of being inactivated in each training iteration, which not only prevents overfitting but also constructs the weight prior distribution. The model uses negative log-likelihood loss to measure the discrepancy between the predicted probability distribution and actual power values, optimizing network parameters (including fully connected layer weights and Dropout parameters) through iterative minimization of this loss to implicitly learn the weight posterior distribution.

Uncertainty quantification is realized through statistical analysis of the prediction set, generating two core indicators: the final point prediction value and the uncertainty characterization index. The former is the mean of the prediction set, calculated as:

$$\bar{y}^* = \frac{1}{S} \sum_{s=1}^S y_s^* \quad (22)$$

This mean serves as the final wind power point prediction, reducing random errors in individual predictions. It can be cross-validated with the BiGRU-TA model's output to assess prediction reliability—high consistency indicates reliable results, while significant deviations prompt inspection of input features (e.g., NWP or historical power data).

The latter indicator is the prediction set's standard deviation, characterizing the magnitude of prediction uncertainty, calculated as:

$$\sigma^* = \sqrt{\frac{1}{S} \sum_{s=1}^S (y_s^* - \bar{y}^*)^2} \quad (23)$$

The standard deviation σ^* directly reflects the dispersion of the prediction set: larger values indicate more scattered predictions and greater uncertainty, while smaller values indicate concentrated predictions and higher reliability. This indicator has clear practical significance—larger standard deviations typically correspond to complex meteorological conditions (e.g., sudden wind speed/direction changes) or high data noise (e.g., turbine fault-induced abnormal power data), reminding dispatchers to increase reserve capacity. Smaller standard deviations indicate stable conditions and reliable data, supporting confident grid operation planning based on the point prediction.

Additionally, combining the prediction set's mean and standard deviation with non-parametric kernel density estimation allows the model to generate dynamic prediction intervals (e.g., 95% confidence intervals). These intervals intuitively illustrate the potential range of wind power output, providing detailed risk information for grid dispatching, energy market transactions, and turbine maintenance scheduling—effectively mitigating operational risks from wind power prediction errors.

4 Case Study

Based on a comprehensive dataset containing 52,704 consecutive samples, four critical operational parameters were extracted for wind power forecasting and analysis, as illustrated in Fig. 10. The wind speed data, recorded in meters per second (m/s), is located at column index 2, representing the primary driving force of power generation. Wind direction measurements in degrees ($^{\circ}$), positioned at column index 16, provide essential information for understanding inflow conditions and wake effects. The active power output in kilowatts (kW), identified at column index 62, serves as the target variable for prediction models. Additionally, nacelle ambient temperature in degrees Celsius ($^{\circ}\text{C}$), found at column index 94, offers crucial insights into environmental conditions affecting turbine performance and power curve characteristics. This high-resolution dataset, collected at 10-min intervals over approximately one year, enables robust modeling of turbine behavior under varying operational conditions and supports the development of accurate forecasting algorithms that account for both meteorological influences and turbine response characteristics. The substantial sample size ensures statistical significance for training complex machine learning architectures while capturing seasonal patterns and transitional weather phenomena relevant to wind power generation.

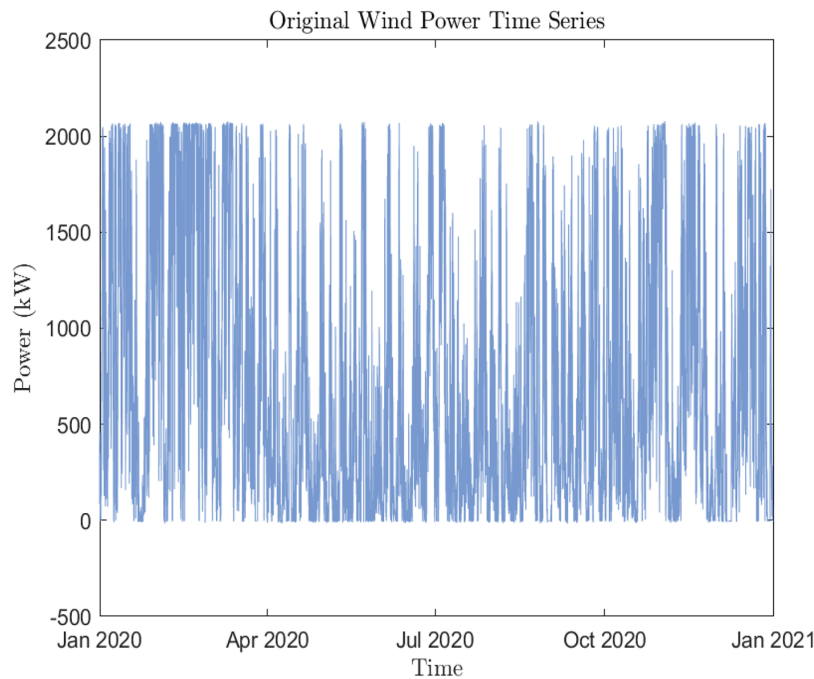


Figure 10: Original wind power time series.

4.1 Deterministic Power Forecasting Using Bi-GRU with Attention Mechanism

The Bi-GRU (Bidirectional Gated Recurrent Unit) model integrated with an attention mechanism was employed for deterministic wind power forecasting. The results demonstrate superior performance in capturing temporal dependencies and feature importance within the wind power time-series data. The attention mechanism effectively highlighted critical time steps, enhancing the model's predictive accuracy in tracking the dynamic fluctuations of wind power output.

The proposed forecasting framework demonstrates compelling efficacy in short-term wind power prediction for the investigated turbine with a rated capacity of 2050 kW. The model's performance was rigorously evaluated, yielding a normalized Mean Absolute Error (nMAE) of 10.94% and a normalized Root Mean Square Error (nRMSE) of 15.73%. These error metrics, expressed as a percentage of the turbine's rated capacity, indicate a high degree of forecasting precision, with the average prediction deviation being within approximately 11% of the installed capacity—an essential benchmark for ensuring reliability in power system operations and renewable energy integration.

Furthermore, the model exhibits exceptional explanatory power, as evidenced by a coefficient of determination (R^2) of 0.9278, as illustrated in Fig. 11. This statistically significant value indicates that the proposed architecture successfully captures and explains 92.78% of the variance observed in the wind power generation time series. As visualized in the scatter plot of predicted vs. actual wind power values, such a high R^2 signifies that the model's predictions closely align with real-world measurements, with the regression line approaching the perfect prediction line. This outstanding predictive performance stems from the synergy between the adaptive signal decomposition technique and the subsequent deep learning architecture, which effectively handles the non-stationary characteristics and complex meteorological dependencies inherent in wind power data, thereby facilitating its application in grid integration, energy market transactions, and operational decision-making within electrical power systems.

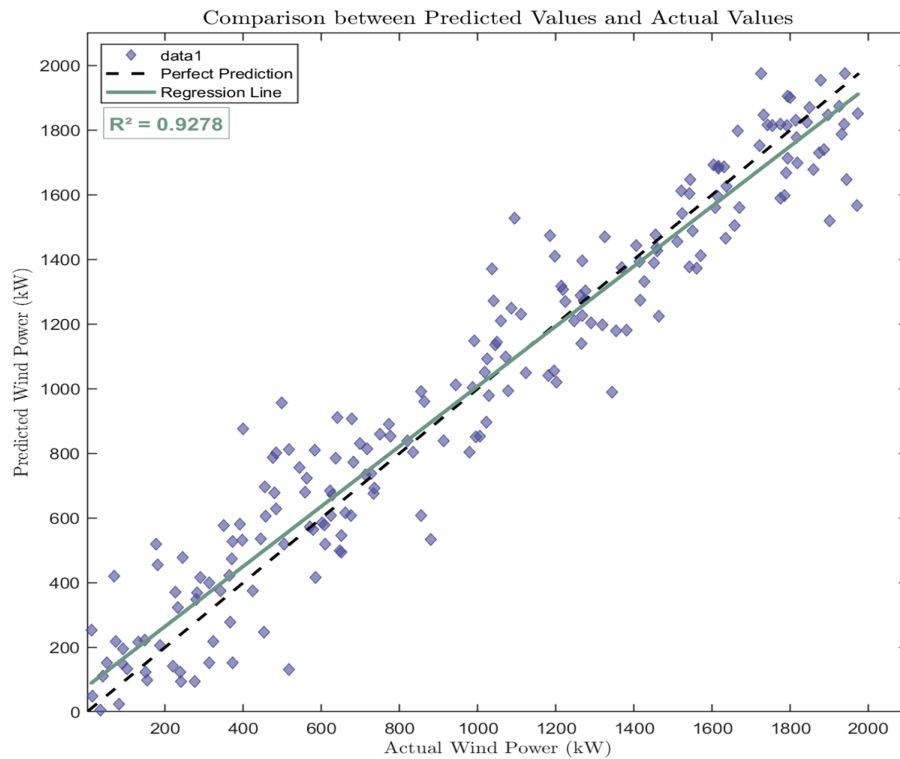


Figure 11: Predicted vs. actual values.

The Fig. 12, consisting of a histogram and a corresponding density curve, provides a statistical perspective on the accuracy of the wind power forecasting model, which is of great significance for evaluating the model's applicability in electrical power system operations. The horizontal axis of the graph represents the prediction error (kW), while the vertical axis reflects the frequency of error occurrence and the probability density. It can be observed that the prediction errors are approximately normally distributed, centered around a mean value close to zero, with most error values clustered within the range of -500 to 500 kW. The standard deviation of the error distribution is quantified at around 331.89 kW, indicating that the model's prediction deviations are relatively controllable. This error profile aligns with the operational requirements of wind power integration, as small and centralized errors help reduce the risk of power imbalance in the electrical grid and facilitate effective power dispatch planning.

The Fig. 13 plot complements the statistical analysis by illustrating the model's dynamic tracking capability over the forecasting horizon, a key performance indicator for short-term wind power prediction in renewable energy systems. The vertical axis of this plot denotes power output (kW), covering the full capacity range of the wind turbine from 0 to 2500 kW, and the horizontal axis represents the time step size, spanning 0 to 2000 to reflect the temporal resolution of the prediction. The blue solid line in the plot corresponds to the actual wind power values, capturing the real-time fluctuations caused by variable wind conditions and turbine operational states; the pink dashed line represents the predicted wind power values generated by the model. Throughout the entire time series, the predicted values closely follow the trend of the actual values, accurately capturing both the peak power outputs (approaching the turbine's rated capacity of 2050 kW) and the low-power operating periods. This tight alignment demonstrates the model's proficiency in capturing the temporal dependencies and non-stationary characteristics of wind power data, ensuring that the forecast results can support reliable grid integration, energy trading decisions, and operational optimization of the wind power system.

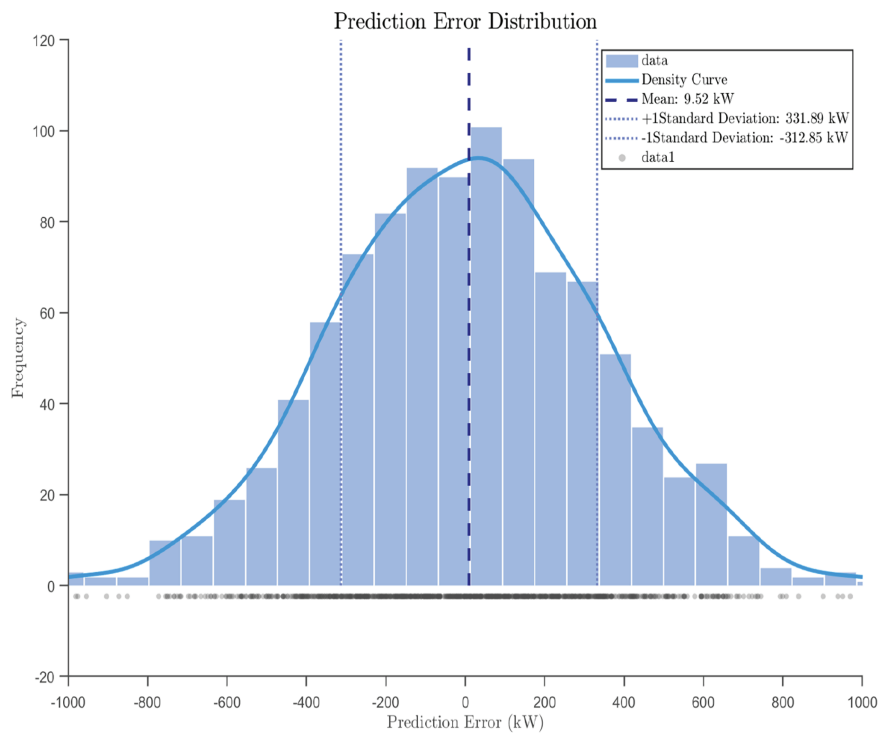


Figure 12: Prediction error distribution.

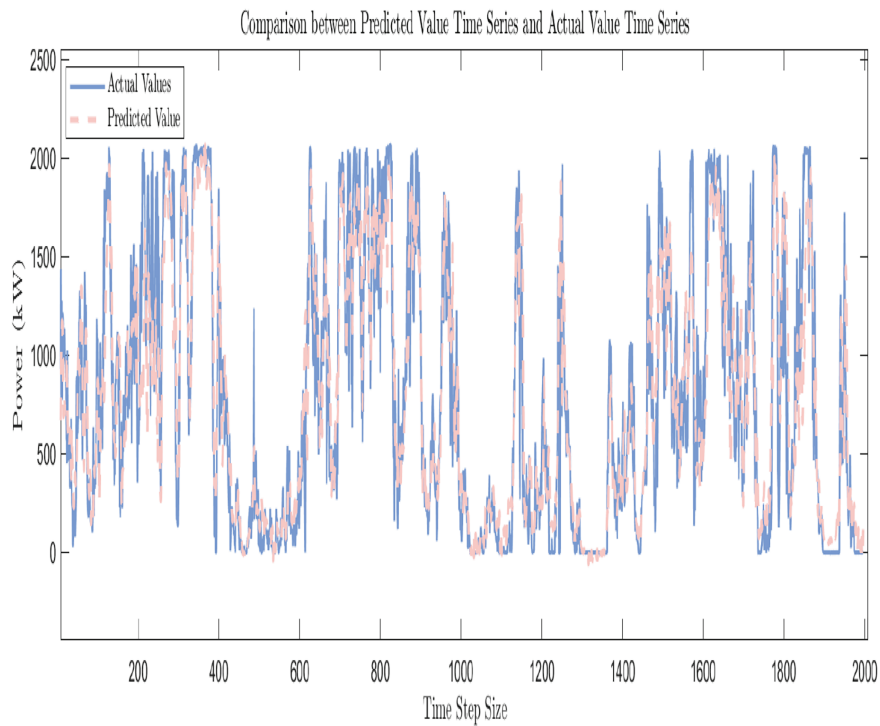


Figure 13: Time series comparison: predicted vs. actual values.

We compared its performance with three benchmark models: CNN-LSTM, Standard LSTM, Random Forest, MLP, SVR.

Based on the comparative performance analysis presented in [Table 1](#), the proposed Bi-GRU-TA model demonstrates superior forecasting capabilities among all evaluated approaches. The model achieves the lowest normalized error metrics, with a nMAE of 10.94%, nRMSE of 15.73%, and an exceptionally high R^2 of 92.78%—indicating that it explains over 92% of the variance in wind power output, a remarkable result for this task. The substantial performance gap between Bi-GRU-TA and traditional methods is particularly evident when comparing with Support Vector Regression (SVR), which exhibits nearly threefold higher error rates (nMAE: 28.07%, nRMSE: 34.01%) and a much lower R^2 of 61.18%. This pronounced disparity underscores the limitations of conventional regression techniques in capturing the complex temporal dynamics and nonlinear patterns inherent in wind power generation data. Among deep learning alternatives, the Bi-GRU-TA architecture maintains a consistent advantage, achieving approximately 11%–30% lower error rates and a notably higher R^2 (92.78%) compared to CNN-LSTM (84.11%), Standard LSTM (75.16%), and MLP (66.37) models. This performance enhancement can be attributed to the synergistic integration of bidirectional processing, which captures both forward and backward temporal dependencies, coupled with the temporal attention mechanism that dynamically weights influential time steps. The attention mechanism's ability to focus on critical temporal segments enables more effective learning of the underlying physical relationships between meteorological conditions and power output fluctuations, thereby boosting the model's explanatory power as reflected in its high R^2 score. The results validate the hypothesis that explicitly modeling temporal dependencies through sophisticated recurrent architectures with attention mechanisms provides significant advantages for wind power forecasting tasks, where accurate prediction requires understanding both short-term fluctuations and long-term patterns in the time-series data. The high R^2 of Bi-GRU-TA further confirms its capability to robustly capture the complex variability in wind power generation.

Table 1: Model performance comparison.

Model Name	nMAE (%)	nRMSE (%)	R^2 (%)
Bi-GRU-TA	10.94	15.73	92.78
CNN-LSTM	12.35	17.82	84.11
Standard LSTM	15.71	20.12	75.16
Random Forest	14.56	19.25	68.41
MLP	12.27	16.32	66.37
SVR	28.07	34.01	61.18

4.2 Probabilistic Forecasting Using Bayesian Neural Networks (BNNs)

To enhance the reliability of overall power prediction for wind power plants, this study employs Bayesian neural Networks (BNNs) to conduct quantitative uncertainty prediction for five wind turbines of different models and installation locations within the plant (denoted as F1–F5). This targeted modeling strategy not only ADAPTS to the output characteristics of each wind turbine affected by terrain and turbulence intensity differences, It also provides refined decision-making basis for the dispatching of power plant clusters, which is of great significance to the operation of the power system and the grid connection of renewable energy. As shown in the [Figs. 14–18](#) graph, the BNNs framework outputs dual results of point prediction and interval prediction for each wind turbine: point prediction is presented as a prediction mean curve, and interval prediction is characterized by a 95% confidence interval to represent statistical uncertainty. This output mode enables independent assessment of the predicted reliability of each wind turbine, facilitating operation and maintenance personnel as well as dispatchers to simultaneously grasp the expected output and potential deviation range of a single wind turbine. Overall, the actual power values of the five wind turbines were highly

consistent with the predicted average for most time steps, and the confidence intervals effectively covered the inherent volatility of wind power, providing key support for risk decision-making in grid dispatching, energy market transactions, and power plant operation planning.

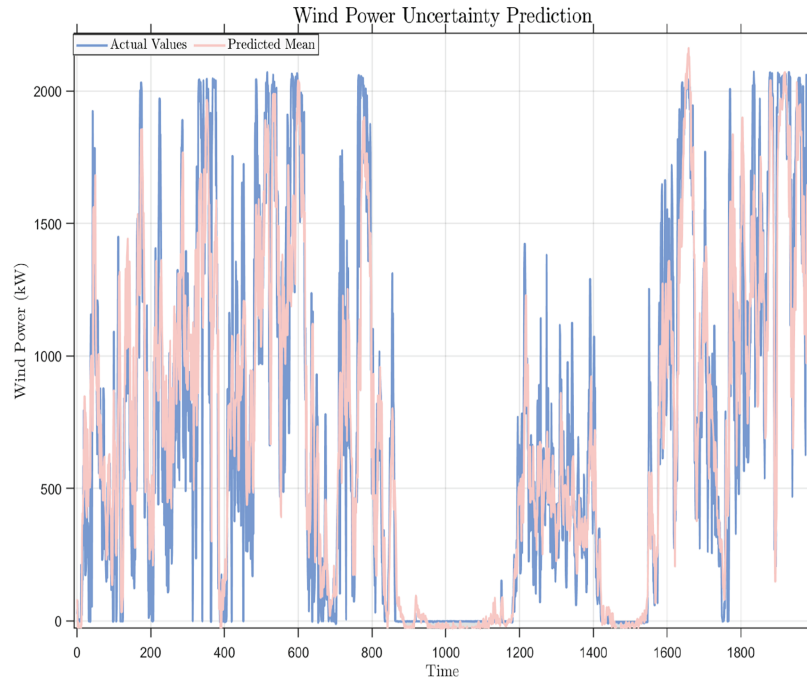


Figure 14: Uncertainty prediction 1.

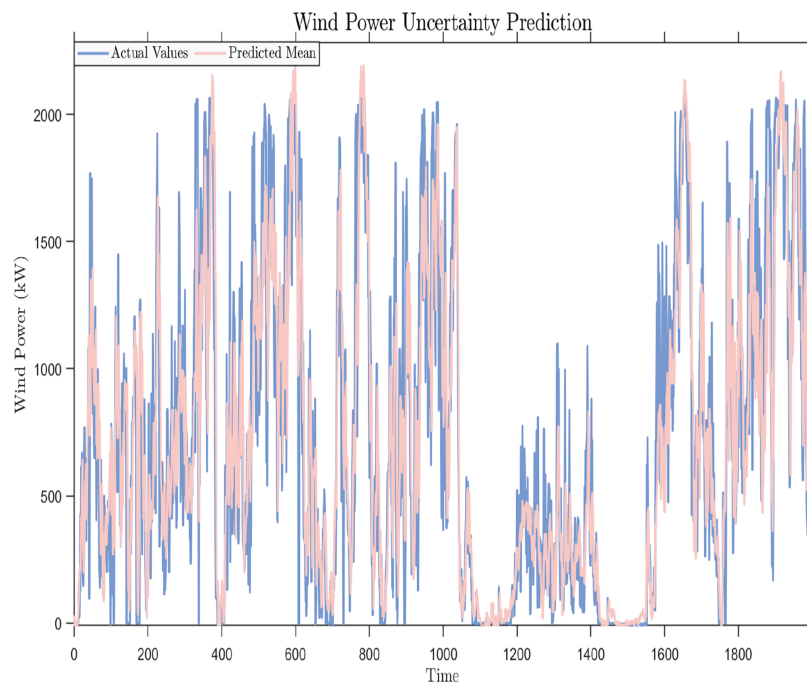


Figure 15: Uncertainty prediction 2.

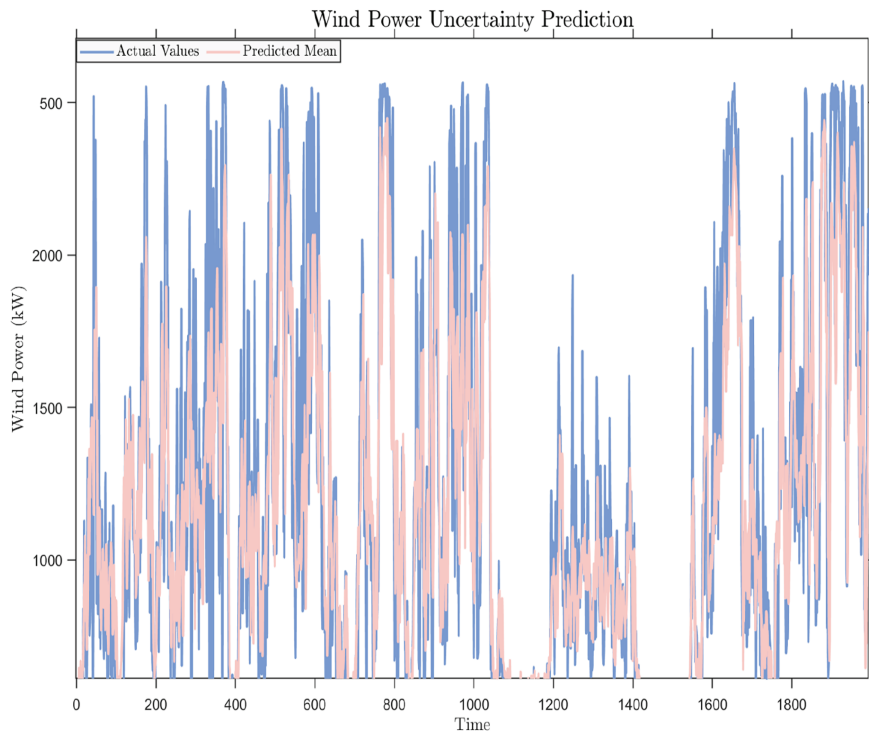


Figure 16: Uncertainty prediction 3.

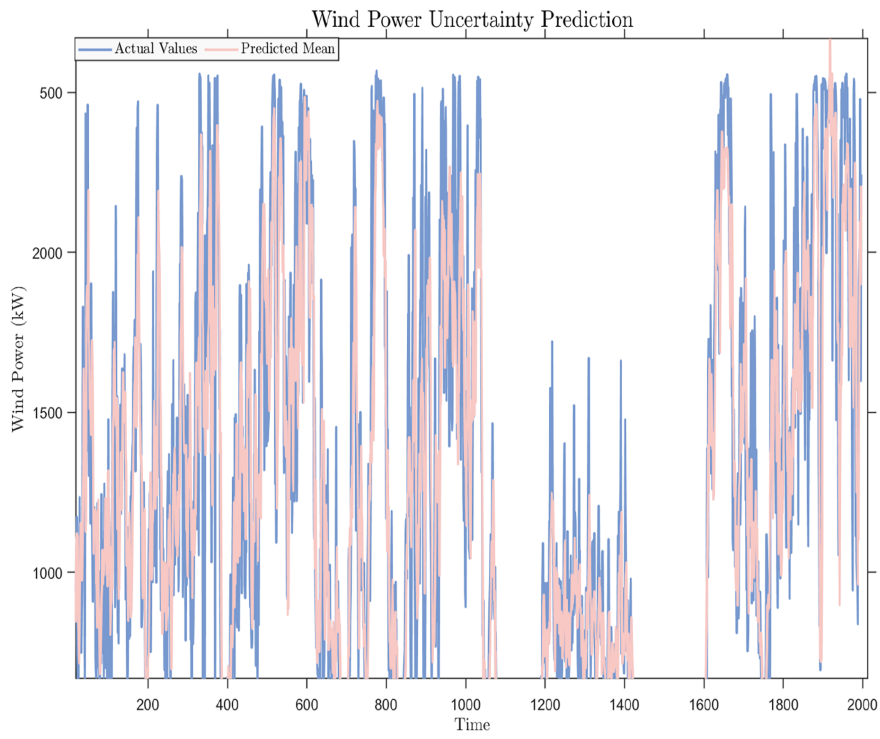


Figure 17: Uncertainty prediction 4.

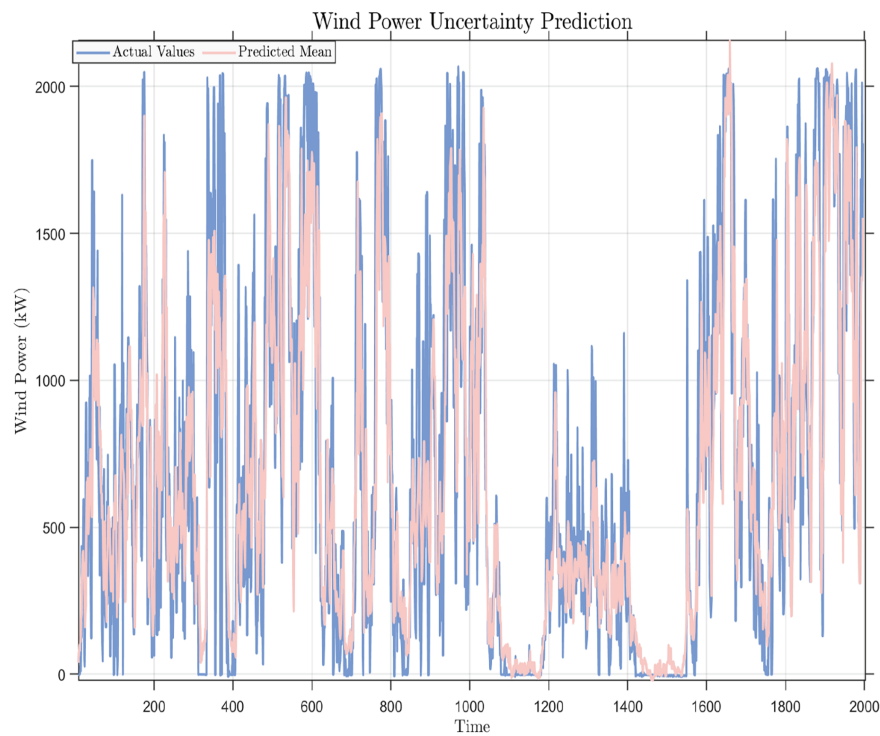


Figure 18: Uncertainty prediction 5.

It is worth emphasizing that BNNs' uncertainty modeling capability specifically addresses the prediction challenges brought about by the intermittency of wind energy and the individual differences of wind turbines, providing a more comprehensive decision-making basis for the practical application of wind power plants. The visualization and quantification results of 1800 time steps jointly indicate that BNNs not only converges to stable prediction performance for each wind turbine characteristic during the training process, but also maintains a robust uncertainty quantification effect throughout the prediction period even during extreme gusts (such as the 1000th to 1200th time steps), the power of all five wind turbines experiences sudden increases and decreases. BNNs can still respond quickly. By dynamically adjusting the width of the confidence interval, it ensures that the actual values do not systematically overflow, thereby enhancing the model's applicability in the prediction of wind power plant clusters for scenarios with dual demands for accuracy and reliability.

5 Conclusion

In summary, the proposed Bi-GRU-TA and Bayesian Neural Networks (BNNs) offer robust and complementary approaches to deterministic and probabilistic wind power forecasting, respectively, thereby enhancing both predictive accuracy and uncertainty quantification in renewable energy modeling. The framework explicitly specifies activation functions for each core module to ensure stable model performance: the Bi-GRU layer employs the tanh function, which is particularly suitable for processing temporal sequence data and helps mitigate gradient vanishing during training; the fully connected layers utilize the ReLU function to enhance feature representation capability and reduce the risk of neuron saturation. This well-considered configuration of activation functions establishes a solid foundation for efficient model training and reliable prediction outcomes. The experimental results demonstrate the advantages of integrating attention mechanisms with Bayesian methods in wind power forecasting, as evidenced by 15.73% and 10.94%

reductions in nRMSE and nMAE, respectively, along with a high coefficient of determination (R^2) of 0.9278, offering valuable insights for the practical deployment of renewable energy within power systems.

The proposed hybrid framework exhibits several notable strengths. First, it achieves synergistic improvement through multi-module integration: Empirical Mode Decomposition (EMD) effectively decomposes non-stationary wind power signals into intrinsic mode functions; the Bi-GRU-TA architecture captures bidirectional temporal dependencies and emphasizes salient features via the attention mechanism; and the BNNs enables robust uncertainty quantification. Second, the framework strikes a balance between predictive accuracy and interpretability—its high R^2 value indicates strong agreement between predicted and observed values, while its probabilistic outputs support risk-aware decision-making in power grid dispatching. Third, it demonstrates robust adaptability to complex wind power dynamics, maintaining performance under meteorological disturbances and fluctuating operational conditions.

Nevertheless, the framework presents certain inherent limitations. Notably, increasing the number of neurons in the Bi-GRU layer leads to a substantial increase in model complexity and computational overhead, thereby compromising real-time inference capabilities and limiting suitability for operational grids requiring rapid response times. Moreover, the purely data-driven design lacks full integration of physical characteristics of wind turbines—such as aerodynamic behavior and mechanical transmission efficiency—which may impair predictive performance under extreme weather events. The use of a single-turbine dataset further constrains the model's generalizability across different turbine types and geographical regions. Additionally, the BNNs' 95% confidence interval does not consistently encompass prediction errors during transient states, such as turbine start-up and shutdown, indicating room for improvement in dynamic uncertainty estimation.

Despite these limitations, this study offers a meaningful contribution to wind power forecasting research. To address the identified challenges, four directions for future work are recommended. First, expanding the dataset to include multi-turbine and multi-site scenarios across diverse wind farms, combined with transfer learning techniques, can enhance cross-context model adaptability. Second, developing a hybrid modeling approach that integrates data-driven architectures (e.g., Bi-GRU-TA) with physics-informed components (e.g., power curve models and wake effect simulations) would improve predictive fidelity while embedding domain-specific knowledge. Third, structural optimization strategies—such as lightweight network designs or parameter pruning methods—should be explored to balance model capacity and computational efficiency, thereby improving real-time applicability. Fourth, refining the BNNs-based uncertainty quantification module to better capture prediction variability during transient operations will strengthen the reliability of confidence estimates. Advancing along these pathways will facilitate the development of more accurate, robust, and generalizable forecasting models, ultimately supporting the integration of high-penetration wind energy into future smart grid systems.

Acknowledgement: None.

Funding Statement: This research was funded by Name of Funder, grant number Yunnan Fundamental Research Projects (202401CF070073).

Author Contributions: Conceptualization, Yanting Wang and Xiaolan Li; methodology, Xiaolan Li and Yanting Wang; software, Yanting Wang; validation, Yanting Wang and Xiaolan Li; formal analysis, Yanting Wang; investigation, Xiaolan Li and Yanting Wang; resources, Yanting Wang; data curation, Yanting Wang; writing—original draft preparation, Xiaolan Li and Yanting Wang; writing—review and editing, Xiaolan Li; visualization, Xiaolan Li and Yanting Wang; supervision, Yanting Wang; project administration, Xiaolan Li; funding acquisition, Xiaolan Li. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data openly available in a public repository.

Ethics Approval: Not applicable.

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

References

1. Abhyankar N, Lin J, Kahrl F, Yin S, Paliwal U, Liu X, et al. Achieving an 80% carbon-free electricity system in China by 2035. *iScience*. 2022;25(10):105180. doi:10.1016/j.isci.2022.105180.
2. Allal Z, Noura HN, Salman O, Chahine K. Machine learning solutions for renewable energy systems: applications, challenges, limitations, and future directions. *J Environ Manage*. 2024;354:120392. doi:10.1016/j.jenvman.2024.120392.
3. Prasad SR, Naik MG. A review on wind power forecasting models for improved renewable energy integration. *Int Conf Inform Sci Technol Innov (ICoSTEC)*. 2022;1(1):35–40. doi:10.35842/icostec.v1i1.20.
4. Loza B, Minchala LI, Ochoa-Correa D, Martinez S. Grid-friendly integration of wind energy: a review of power forecasting and frequency control techniques. *Sustainability*. 2024;16(21):9535. doi:10.3390/su16219535.
5. Xie Y, Li C, Li M, Liu F, Taukenova M. An overview of deterministic and probabilistic forecasting methods of wind energy. *iScience*. 2023;26(1):105804. doi:10.1016/j.isci.2022.105804.
6. Canché-Cab L, San-Pedro L, Ali B, Rivero M, Escalante M. The atmospheric boundary layer: a review of current challenges and a new generation of machine learning techniques. *Artif Intell Rev*. 2024;57(12):339. doi:10.1007/s10462-024-10962-5.
7. Donadio L, Fang J, Porté-Agel F. Numerical weather prediction and artificial neural network coupling for wind energy forecast. *Energies*. 2021;14(2):338. doi:10.3390/en14020338.
8. Yakoub G, Mathew S, Leal J. Power production forecast for distributed wind energy systems using support vector regression. *Energy Sci Eng*. 2022;10(12):4662–73. doi:10.1002/ese3.1295.
9. Shireen T, Shao C, Wang H, Li J, Zhang X, Li M. Iterative multi-task learning for time-series modeling of solar panel PV outputs. *Appl Energy*. 2018;212:654–62. doi:10.1016/j.apenergy.2017.12.058.
10. Li Y, Ye Y, Xu Y, Li L, Chen X, Huang J. Two-stage forecasting of TCN-GRU short-term load considering error compensation and real-time decomposition. *Earth Sci Inform*. 2024;17(6):5347–57. doi:10.1007/s12145-024-01456-7.
11. Liu ZF, Liu YY, Chen XR, Zhang SR, Luo XF, Li LL, et al. A novel deep learning-based evolutionary model with potential attention and memory decay-enhancement strategy for short-term wind power point-interval forecasting. *Appl Energy*. 2024;360:122785. doi:10.1016/j.apenergy.2024.122785.
12. De Giorgi M, Campilongo S, Ficarella A, Congedo P. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). *Energies*. 2014;7(8):5251–72. doi:10.3390/en7085251.
13. Su Y, Wang Z, Dong Z, Hua X, Ye T, Song Z, et al. Frequency-aware ultra-short-term wind power forecasting using CEEMDAN-VMD-SE and Transformer-GRU networks. *Energy*. 2025;338:138715. doi:10.1016/j.energy.2025.138715.
14. Zhao Z, Yun S, Jia L, Guo J, Meng Y, He N, et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng Appl Artif Intell*. 2023;121:105982. doi:10.1016/j.engappai.2023.105982.
15. Dhaka P, Sreejeth M, Tripathi MM. Empirical mode decomposition for improved wind power forecasting with boosted GRU model. In: *Proceedings 2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*; 2024 Apr 26–28; Delhi, India. p. 722–6.
16. Li D, Ye Y. Ultra-short-term wind power forecasting based on ICEEMDAN-Informed BiGRU network with multi-head attention. *Adv Comput Signals Syst*. 2025;9(3):79–87. doi:10.23977/acss.2025.090310.
17. Ruiz-Aguilar JJ, Turias I, González-Enrique J, Urda D, Elizondo D. A permutation entropy-based EMD-ANN forecasting ensemble approach for wind speed prediction. *Neural Comput Appl*. 2020;33(7):2369–91. doi:10.1007/s00521-020-05141-w.

18. Shen W, Jiang N, Li N. An EMD-RF based short-term wind power forecasting method. In: Proceedings 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS); 2018 May 25–27; Enshi, China.
19. Wang J, Zhang L, Liu Z, Niu X. A novel decomposition-ensemble forecasting system for dynamic dispatching of smart grid with sub-model selection and intelligent optimization. *Expert Syst Appl.* 2022;201:117201. doi:10.1016/j.eswa.2022.117201.
20. Pholsena K, Pan L, Zheng Z. Mode decomposition based deep learning model for multi-section traffic prediction. *World Wide Web.* 2020;23(4):2513–27. doi:10.1007/s11280-020-00791-1.
21. Bashir T, Wang H, Tahir M, Zhang Y. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models. *Renew Energy.* 2025;239:122055. doi:10.1016/j.renene.2024.122055.
22. Qin C, Xie J, Cao Y, Zhu B. Forecasting short-term wind power with multi-view attention mechanism and dual recurrent neural networks. *Expert Syst Appl.* 2026;297:129472. doi:10.1016/j.eswa.2025.129472.
23. Yuan J, Wu F, Wu H. Multivariate time-series classification using memory and attention for long and short-term dependence. *Appl Intell.* 2023;53(24):29677–92. doi:10.1007/s10489-023-05079-1.
24. He M, Qian Q, Liu X, Zhang J, Curry J. Recent progress on surface water quality models utilizing machine learning techniques. *Water.* 2024;16(24):3616. doi:10.3390/w16243616.
25. Xue T, Qu L, Chen G, Wang D, Yi Z, Xu Y, et al. Short-term wind energy forecasting using attention-based encoder decoder GRU framework. In: Proceedings 2024 6th International Conference on Energy, Power and Grid (ICEPG); 2024 Sep 27–29; Guangzhou, China. p. 566–71.
26. Xu W, Liu Y, Fan X, Shen Z, Wu Q. Short-term wind power forecasting based on dual attention mechanism and gated recurrent unit neural network. *Front Energy Res.* 2024;12:111. doi:10.3389/fenrg.2024.1346000.
27. Akkilic AN, Sabir Z, Bhat SA, Bulut H. A radial basis deep neural network process using the Bayesian regularization optimization for the monkeypox transmission model. *Expert Syst Appl.* 2024;235:121257. doi:10.1016/j.eswa.2023.121257.
28. Lafuente-Cacho M, Izquierdo-Monge O, Peña-Carro P, Hernández-Jiménez Á., Callejo LH, Losada AMP, et al. State of the Art for solar and wind energy-forecasting methods for sustainable grid integration. *Curr Sustain Renewable Energy Rep.* 2025;12(1):2541. doi:10.1007/s40518-025-00262-z.
29. Qureshi S, Shaikh F, Kumar L, Ali F, Awais M, Gürel AE. Short-term forecasting of wind power generation using artificial intelligence. *Environ Challenges.* 2023;11:100722. doi:10.1016/j.envc.2023.100722.
30. Dhaka P, Sreejeth M, Tripathi MM. A survey of artificial intelligence applications in wind energy forecasting. *Arch Comput Methods Eng.* 2024;31(8):4853–78. doi:10.1007/s11831-024-10182-8.
31. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renew Energy.* 2012;37(1):1–8. doi:10.1016/j.renene.2011.05.033.
32. Cevik HH, Çunkaş M. Day-ahead wind power forecasting using numerical weather prediction. *Proc Int Conf Acad Res Sci Technol Eng.* 2023;1(1):32–8. doi:10.33422/icarste.v1i1.42.
33. Huang CL, Wu YK, Phan QT, Tsai CC, Hong JS. Enhancing wind power forecasts via bias correction technologies for numerical weather prediction model. *IEEE Trans Ind Appl.* 2025;61(4):5406–19. doi:10.1109/tia.2025.3546589.
34. Alam MTU, Mozomder AS. Comparative machine learning and time series forecasting of wind power output using SCADA data. *J Comput Sci Technol Studies.* 2025;7(7):14–30. doi:10.32996/jcsts.2025.7.7.2.
35. Wang H, Hu Z, Chen Z, Zhang M, He J, Li C. A hybrid model for wind power forecasting based on ensemble empirical mode decomposition and wavelet neural networks. *Diangong Jishu Xuebao or Trans China Electrotech Soc.* 2013;31:2099. (In Chinese). doi:10.1109/ciced.2018.8592083.
36. Uzair M, Shah I, Ali S. An adaptive strategy for wind speed forecasting under functional data horizon: a way toward enhancing clean energy. *IEEE Access.* 2024;12:68730–46. doi:10.1109/access.2024.3401038.
37. Li Y, Tang F, Gao X, Zhang T, Qi J, Xie J, et al. Numerical weather prediction correction strategy for short-term wind power forecasting based on bidirectional gated recurrent unit and XGBoost. *Front Energy Res.* 2022;9:150. doi:10.3389/fenrg.2021.836144.

38. Alkhatat G, Hasan SH, Mehmood R. A hybrid model of variational mode decomposition and long short-term memory for next-hour wind speed forecasting in a hot desert climate. *Sustainability*. 2023;15(24):16759. doi:10.3390/su152416759.
39. Araujo MLS, Kitagawa YKL, Weyll ALC, Lima FJLD, Santos TSD, Jacondino WD, et al. Wind power forecasting in a semi-arid region based on machine learning error correction. *Wind*. 2023;3(4):496–512. doi:10.3390/wind3040028.