

# A Wind Speed Prediction Model Based on Machine Learning in Guyuan Area

Shiyun Mu<sup>1</sup>, Yuming Zhai<sup>2\*</sup>, Hongxia Shi<sup>1</sup>, Shujie Yuan<sup>1</sup>, Lin Han<sup>1</sup>, Lixin Su<sup>3</sup>, Hailing Shi<sup>2</sup>, Juan Gu<sup>2</sup>

<sup>1</sup>School of Atmospheric Sciences, Chengdu University of Information and Technology, Chengdu, China

<sup>2</sup>Guyuan Meteorological Bureau, Guyuan, China

<sup>3</sup>Jilin Provincial Institute of Meteorological Science, Changchun, China

Email: \*Nxgyzym-001@163.com

**How to cite this paper:** Mu, S. Y., Zhai, Y. M., Shi, H. X., Yuan, S. J., Han, L., Su, L. X., Shi, H. L., & Gu, J. (2025). A Wind Speed Prediction Model Based on Machine Learning in Guyuan Area. *Journal of Geoscience and Environment Protection*, 13, 186-199. <https://doi.org/10.4236/gep.2025.1311010>

**Received:** October 20, 2025

**Accepted:** November 18, 2025

**Published:** November 21, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Under the context of global climate change, the frequent occurrence of strong winds in Guyuan has significantly hindered the development of local facility agriculture. Using hourly meteorological data from the Sanying National Station and the Guyuan Greenhouse Station between April 2024 and April 2025, this study employed machine learning methods to develop wind speed prediction models based on BP neural network, support vector machine, and random forest (referred to as BP, SVM, and RF models), aiming to provide references for local disaster prevention and mitigation. The results indicate that: 1) Wind speed at the Guyuan Greenhouse Station exhibits the strongest correlation with that at the National Station (0.489 - 0.595), followed by temperature and 24-hour precipitation (0.116 - 0.336). 2) The mean absolute error (MAE) of the BP, RF, and SVM models at all heights is below 1.5 m/s, the root mean square error (RMSE) is under 2.1 m/s, and the forecast accuracy (FA) exceeds 75%, indicating satisfactory model performance. Compared to 3 m, the MAE and RMSE of 0.5 m are larger, while the FA is smaller. This indicates that the wind speed of 0.5 m is close to the ground, and is more affected by surface roughness and turbulence effects, resulting in greater randomness and making the model more difficult. 3) Based on case analyses of May 10 and May 1, 2024, the overall simulation performance ranks as “RF model > SVM model > BP model”; however, the SVM model demonstrates higher accuracy in simulating strong wind events.

## Keywords

Guoyuan Strong Wind, BP Neural Network, Support Vector Machine, Random Forest, Wind Speed Prediction

## 1. Introduction

Anthropogenic climate change is intensifying extreme weather events, constitut-

ing a major threat to both social systems and the natural environment (Lyu et al., 2025; Landis et al., 2025). In China, the development of facility agriculture (e.g., solar and plastic greenhouses) has advanced rapidly since the 1980s. However, the capacity of these agricultural facilities to withstand natural disasters remains inadequate. Consequently, the increasing frequency of extreme weather events under climate change results in substantial production losses, presenting a critical constraint to the sustainable development of this sector (Tan et al., 2022; Romero-Gomez et al., 2008).

Guoyuan is located in the central hilly sector of the Loess Plateau in southern Ningxia and the northeastern zone of the Liupan Mountain region. The area is characterized by an intricate gully system and a fragile ecological environment, with scarce water resources and a high incidence of meteorological disasters including drought and strong winds. These factors have collectively posed constraints on regional socio-economic development and the advancement of new rural areas. In recent years, facility agriculture in Guoyuan has undergone rapid development, contributing to the restructuring of the agricultural supply system and emerging as a notable factor in raising farmers' incomes (Ma et al., 2020; Zeng et al., 2018). Nevertheless, in the context of global increases in extreme weather, Yuquan County within Guoyuan City frequently experiences strong winds (Chen et al., 2018a), which adversely affect greenhouse stability and crop production. Consequently, building a localized wind speed prediction model using machine learning is highly significant for enhancing disaster prevention and mitigation capabilities for greenhouse agriculture.

In recent years, significant progress has been made in the development of prediction models, with scholars adopting a variety of methodologies. Traditional statistical approaches, such as stepwise regression, have been widely applied across multiple fields (Gao et al., 2023; Deng, 2024; Kang et al., 2021). In the domain of machine learning, methods like support vector machines (Lin & Yang, 2019; Xu et al., 2009; Zhang et al., 2010; Fan et al., 2021; Kamal et al., 2021) and random forests (Yang et al., 2022; Wang et al., 2023) have demonstrated strong predictive performance. With the advancement of deep learning, models such as Long Short-Term Memory (LSTM) networks (Zhang et al., 2024; Ihsan & Mehmet, 2023) and BP neural networks (Chen et al., 2018b; Jin et al., 2011; Fan et al., 2024; Liu et al., 2022; Li et al., 2022) have also shown exceptional results in forecasting tasks. However, research on wind speed prediction in the Ningxia region has been limited and remains relatively scarce (Shen & Wei, 2016; Yan et al., 2014), particularly in localized areas such as Guoyuan City. As an important city in southern Ningxia, Guoyuan's unique geographical and climatic characteristics lend theoretical value and practical significance to wind speed prediction studies. Therefore, conducting wind speed prediction research in Guoyuan City can not only address a gap in this research area but also provide a scientific basis for the development of local wind energy resources and the prevention of meteorological disasters.

Accordingly, this paper presents the development of a wind speed prediction

model using machine learning methods, based on hourly meteorological data collected from the Sanying National Station and the Guoyuan Greenhouse Station between April 2024 and April 2025.

## 2. Sources of Data and Research Methods

### 2.1. Source of Data

The data is sourced from the Guoyuan Meteorological Bureau, including hourly maximum wind speed, maximum wind speed, 2-minute average wind speed, 10-minute average wind speed at the Sanying National Station from April 2024 to April 2025, and hourly temperature, humidity, air pressure, the latest 15-minute rainfall, the latest 1-hour rainfall, the latest 24-hour rainfall, 0.5 m wind speed, 2 m wind speed, and 3 m wind speed at the Guoyuan Greenhouse Station from April 2024 to April 2025.

### 2.2. BP Neural Network

The BP neural network is a common artificial neural network model used to solve various tasks, such as classification and regression. It consists of an input layer, a hidden layer, and an output layer. It is trained using the forward propagation and backpropagation algorithms. Forward propagation is used to calculate the network's output, while backpropagation is used to adjust the weights and biases to reduce errors. This process is iterated on the training data until the desired performance is achieved. The BP neural network has the characteristics of wide applicability, strong learning ability, high flexibility, and the ability to solve various complex problems. This is an excellent method for predicting the wind speed in the Guoyuan greenhouse.

Calculate the meteorological elements of each station at the Guoyuan greenhouse, the wind speed of each adjacent national station, and the correlation with the wind speed at the Guoyuan greenhouse station. Input the meteorological elements with better correlation to the Guoyuan greenhouse station into the BP neural network for training, and construct a wind speed prediction model based on the BP neural network (referred to as the BP model). The BP neural network consists of 3 layers, the input layer and the hidden layer.

To address the issue of inconsistent units of input variables, the Min-Max Normalization method was employed to normalize the meteorological elements, that is

$$x_i = \frac{X_{ij} - X_{i\min}}{X_{i\max} - X_{i\min}} \quad (1)$$

In the formula,  $x_i$  represents the normalized data of the  $i$ -th input variable (where  $i = 1, 2, 3, 4$ ),  $X_{ij}$  represents the  $j$ -th value of the  $i$ -th input variable, and  $X_{i\min}$  and  $X_{i\max}$  represent the maximum and minimum values of this input variable.

To ensure accuracy, 70% of the data was randomly selected for training and 30% for testing.

### 2.3. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm mainly used for classification and regression tasks. Its core idea is to find an optimal hyperplane that maximizes the distance between data points of different classes on this hyperplane. By using a kernel function, SVM can handle non-linearly separable data and map the original space to a high-dimensional feature space to achieve effective separation. SVM is particularly suitable for small sample and high-dimensional data, and has strong generalization ability, but it has high computational complexity and is sensitive to parameter and kernel function selection. This paper constructs a wind speed prediction model based on SVM (referred to as the SVM model). 70% of the data are randomly selected for training, and 30% for testing. The normalization method is consistent with the BP neural network model.

### 2.4. Random Forest

Random Forest is a machine learning algorithm based on ensemble learning. It improves the accuracy and robustness of the model by constructing multiple decision trees and combining their prediction results. Each tree is trained using a different data subset (Bootstrap sampling) and a randomly selected feature subset to reduce the risk of overfitting. The final classification result is determined by voting (for classification tasks) or averaging (for regression tasks). Random Forest has the advantages of fast training speed, ability to handle high-dimensional data, and insensitivity to noise and missing values, and is widely used in classification, regression, and feature selection tasks. This paper constructs a wind speed prediction model based on Random Forest (referred to as the RF model). 70% of the data is randomly selected for training, and 30% is used for testing. The normalization method is consistent with the BP neural network model.

### 2.5. Model Evaluation Methods

To clarify the forecasting performance of the BP neural network model, the support vector machine model and the random forest model, the following methods were used for evaluation:

- 1) Average Mean Squared Error MSE and Root Mean Squared Error RMSE:

$$\text{MSE} = \frac{\sum_{i=1}^n |y_{1i} - y_{2i}|}{n} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_{1i} - y_{2i})^2}{n}} \quad (3)$$

In the formula,  $n$  represents the number of samples,  $i$  denotes the  $i$ -th observed data,  $y_1$  represents the predicted wind speed, and  $y_2$  represents the measured wind speed at the Guoyuan greenhouse station.

- 2) Prediction Accuracy Rate FA: The prediction accuracy rate is the ratio of the number of correct predictions made by the model to the total number of predictions. According to the regulations in the "Quality Inspection Methods for

Weather Analysis and Forecast” of the China Meteorological Administration (Chen et al., 2012): If the error between the predicted value and the measured value is within 2 m/s, it indicates a correct prediction.

### 3. Research Results

#### 3.1. Correlation between Wind Speed and Various Meteorological Factors

A summary of the correlations between wind speeds at various heights and meteorological elements is provided in **Table 1**. Key findings include: 1) Wind speeds at 0.5 m, 2 m, and 3 m were significantly positively correlated with air temperature, with coefficients ranging from 0.149 to 0.336. The correlation was strongest at 0.5 m, whereas those at 2 m and 3 m were comparatively weaker. 2) All height levels demonstrated significant negative correlations with humidity, being most substantial at 2 m (−0.262) and least at 0.5 m (−0.126). 3) Weak negative correlations with air pressure were identified, which were most notable at 0.5 m and diminished at higher elevations. 4) Correlations with short-term precipitation intervals (15-minute, 1-hour) were minimal (0.052 - 0.073) and thus deemed practically insignificant, whereas a slight increase was noted for 24-hour precipitation, albeit remaining weak. 5) The strongest relationships were observed between all local wind speeds and national station wind metrics (including greatest, maximum, gust, 2-minute, and 10-minute averages), with coefficients from 0.489 to 0.630. Among these, the 2 m wind speed correlated most strongly, thereby highlighting its close linkage to synoptic-scale wind conditions.

**Table 1.** Correlation between wind speed at various heights and various meteorological elements.

| Wind speed | Temperature                   | Humidity                                    | Atmospheric pressure                       | The rainfall in the last 15 minutes             | The rainfall in the last hour                            |
|------------|-------------------------------|---|--|---|--|
| 0.5 m      | 0.336**                       | −0.126**                                    | −0.200**                                   | 0.068**   | 0.073**  |
| 2 m        | 0.149**                       | −0.262**                                    | −0.081**                                   | 0.059**   | 0.065**  |
| 3 m        | 0.162**                       | −0.207**                                    | −0.099**                                   | 0.052**   | 0.062**  |
| Wind speed | Rainfall in the last 24 hours | Greatest wind speed at the national station | Maximum wind speed at the national station | 2-minute average wind speed at National Station | The 10-minute average wind speed at the national station |
| 0.5 m      | 0.157**                       | 0.560**                                     | 0.531**                                    | 0.496**   | 0.514**  |
| 2 m        | 0.140**                       | 0.630**                                     | 0.602**                                    | 0.575**   | 0.595**  |
| 3 m        | 0.116**                       | 0.539**                                     | 0.515**                                    | 0.489**   | 0.503**  |

Note: \*\* 0.01 significance.

#### 3.2. Wind Speed Prediction Model Based on BP Neural Network

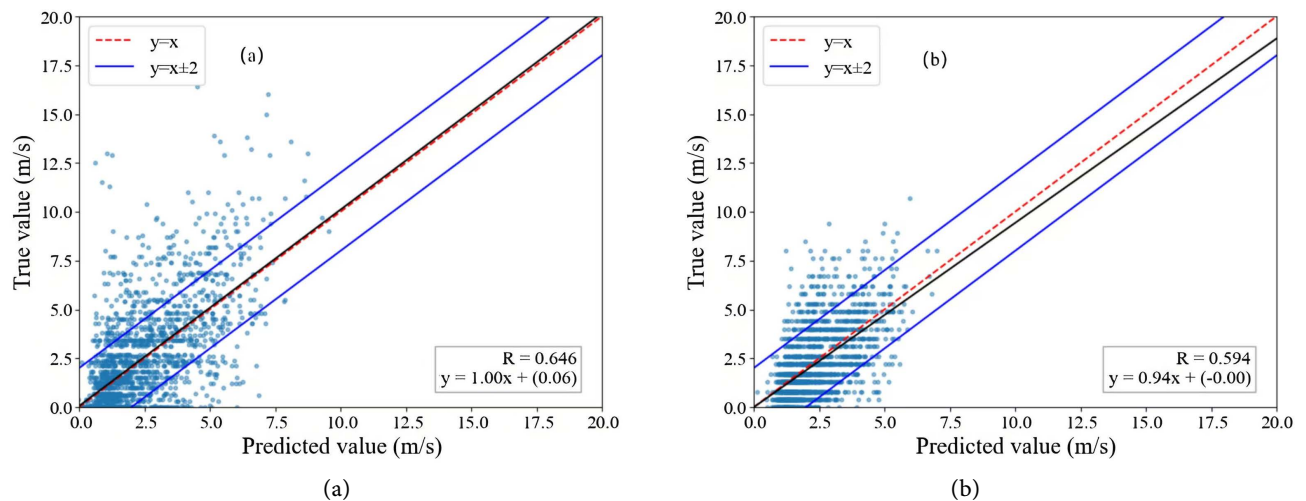
A BP neural network is a multi-layer feedforward architecture trained according to the error backpropagation algorithm. It comprises an input layer, at least one hidden layer, and an output layer. The network operates by computing an output

value via forward propagation; the resulting error is then propagated backward to adjust synaptic weights and neuronal biases across all layers. Utilizing the gradient descent method, the algorithm minimizes a predefined loss function, thereby enabling the modeling of complex nonlinear relationships. Its defining feature is the use of the chain rule to propagate errors layer by layer and update parameters accordingly, rendering it applicable to classification and regression problems—though it is susceptible to issues like vanishing gradients and overfitting. To construct the optimal BP neural network model, this study systematically evaluated the impact of different configurations on the model's performance. The model will use ReLU, Sigmoid, and Tanh as activation functions, and select Adam and Stochastic Gradient Descent (SGD) as optimizers. We set the number of hidden layer nodes to 10, 14, 18, and 22, and the learning rate to be chosen from 0.01, 0.001, and 0.0001. All models have a maximum iteration of 500 times and the early stopping method is set to ensure the model converges early. Finally, we obtained the best BP model configuration: ReLU activation function, Adam optimizer, learning rate of 0.001, and 18 layers of hidden layers.

Based on the metrics of the BP model presented in **Table 2** and the comparison between predicted and actual values in **Figure 1**, the following conclusions can be drawn: 1) At each height, the mean absolute error (MAE) and root mean square error (RMSE) are both below 2 m/s, while the forecast accuracy (FA) exceeds 75%. Moreover, most scatter points are concentrated within the range of  $y = x \pm 2$ , indicating a satisfactory simulation performance of the BP neural network model. 2) The trend equations for predicted versus actual values at 0.5 m and 3 m heights are  $y = x + 0.06$  and  $y = 0.94x$ , respectively, both of which are close to the ideal line  $y = x$ . The coefficients of determination (R) are 0.646 and 0.594, further supporting the good simulation capability of the BP neural network model. 3) As the height increases, the MAE decreases from 1.42 to 1.06, the RMSE decreases from 2.00 to 1.40, and the FA increases from 77.84% to 87.13%. This suggests that wind speed at 0.5 m, being closer to the ground, experiences more strong wind events and is more susceptible to terrain and other influencing factors, thereby posing greater challenges for accurate modeling. The accuracy of wind speed prediction at a height of 0.5 meters above the ground is relatively low. The main reason for this is that this height is located at the bottom of the atmospheric boundary layer, where the influence of surface friction, terrain disturbance, and small-scale turbulence is extremely significant. This results in not only large instantaneous changes in wind speed but also stronger spatial randomness. These complex near-ground layer physical characteristics make it difficult for models based on historical data to accurately depict and predict its fine variations. Overall, the simulation performance improves with increasing height.

**Table 2.** Indicators of the BP Model.

| Height (m) | MAE (m/s) | RMAE (m/s) | FA (%) |
|------------|-----------|------------|--------|
| 0.5        | 1.42      | 2.00       | 77.84  |
| 3          | 1.07      | 1.40       | 87.13  |



**Figure 1.** Comparison between predicted values and actual values of the BP model at (a) 0.5 m; (b) 3 m.

### 3.3. Wind Speed Prediction Model Based on Support Vector Machine

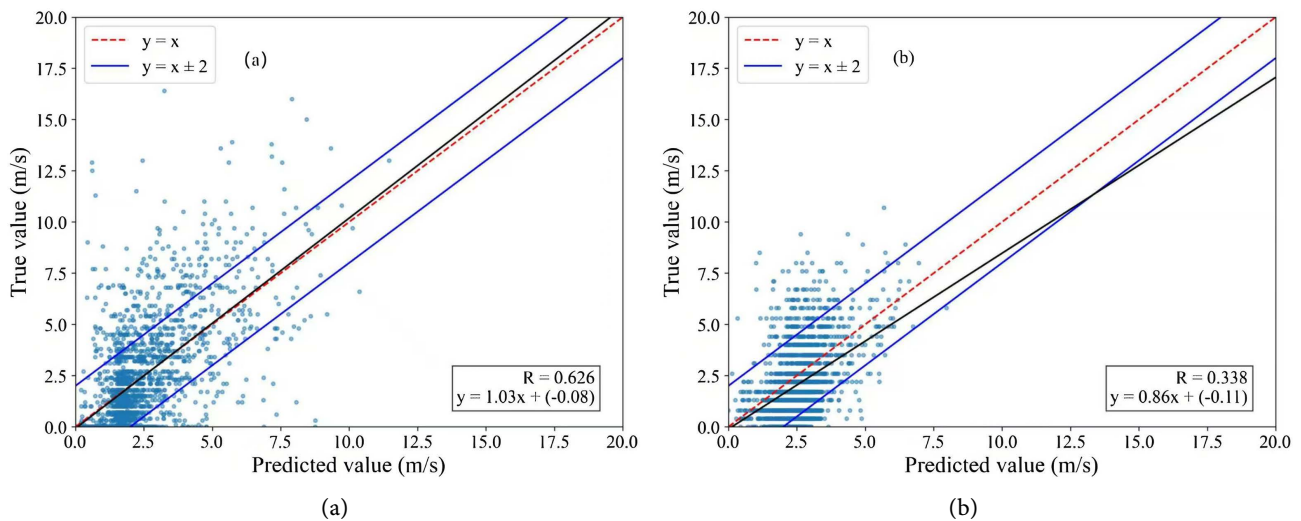
Support Vector Machine (SVM) is a supervised learning algorithm designed for both classification and regression tasks. The core principle of SVM is to identify the optimal hyperplane that maximizes the margin between different classes of data. The model defines the decision boundary using support vectors—the samples closest to the hyperplane—and employs kernel functions to handle nonlinear problems. SVM performs well with high-dimensional data and small sample sizes, and generally exhibits strong generalization capability. However, its computational efficiency decreases with large-scale datasets, and its performance is sensitive to parameter selection. To determine the optimal configuration of the Support Vector Machine (SVM), the system evaluated the performance of different combinations of hyperparameters. The kernel function of the model was selected from “Linear”, “Poly”, “Sigmoid” and “RBF”. The regularization parameters were set to 0.01, 0.1 and 1, the width of the insensitive region was set to 0.01, 0.1 and 1, and the core coefficients of all kernel functions were calculated using the “scale” method. After a comprehensive comparison, the optimal configuration of the SVM model was finally determined to be: radial basis function kernel (RBF), regularization parameter of 1, and insensitive region width of 0.1.

Based on the performance metrics of the SVM model provided in **Table 3** and the predicted-versus-actual comparison in **Figure 2**, the following conclusions can be drawn: 1) For wind speeds at both 0.5 m and 3 m heights, the mean absolute error (MAE) ranges from 1.33 to 1.48 m/s, and the root mean square error (RMSE) ranges from 1.63 to 2.04 m/s, all of which remain below 2 m/s. The forecast accuracy (FA) varies between 78.93% and 79.24%, consistently exceeding 75%. In addition, most scatter points at these heights fall within the range of  $y = x \pm 2$ , indicating a generally good performance of the support vector machine model. 2) The trend equations relating predicted to actual values are  $y = 1.03x - 0.08$  at 0.5 m height and  $y = 0.86x - 0.11$  at 3 m height, with coefficients of determination (R) of 0.626 and 0.338, respectively. The trend line at 0.5 m height is closer to the ideal  $y$

= x line and exhibits a higher  $R^2$  value compared to that at 3 m. 3) In comparison to the 3 m height, the MAE and RMSE are larger at 0.5 m, while the FA is lower. This once again demonstrates the influence of surface roughness and turbulence on the wind speed of 0.5 meters.

**Table 3.** Indicators of the SVM Model.

| Height (m) | MAE (m/s) | RMAE (m/s) | FA (%) |
|------------|-----------|------------|--------|
| 0.5        | 1.48      | 2.04       | 78.93  |
| 3          | 1.33      | 1.63       | 79.24  |



**Figure 2.** Comparison of predicted values and actual values of the SVM model at (a) 0.5 m; (b) 3 m.

### 3.4. Wind Speed Prediction Model Based on Random Forest

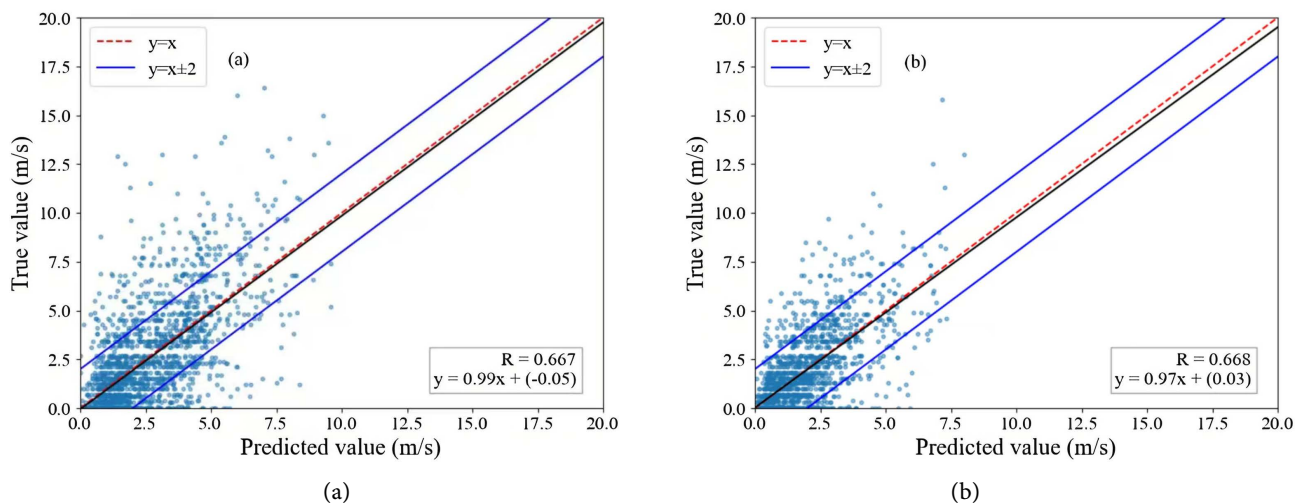
Random Forest is a supervised learning method based on ensemble learning. Its core mechanism involves building a multitude of decision trees and integrating their outputs to improve predictive accuracy and model robustness. By introducing randomness via feature subset selection and bootstrap sampling, it effectively reduces the risk of overfitting. The algorithm is applicable to classification and regression problems and is recognized for its training efficiency, interpretability, and capability to process high-dimensional data. A notable limitation, however, is its potential suboptimal performance on noisy datasets. To determine the optimal parameters for the random forest model, this study systematically examined the performance of different configurations: the number of trees was selected between 60, 80, and 100, while the minimum number of samples per leaf node was adjusted between 3, 5, and 7. After comprehensive comparison, the final optimal configuration for the model was determined to be 100 trees and 5 minimum leaf node samples, and this combination achieved the best performance on the validation set.

From **Table 4** and **Figure 3**, which present the performance metrics of the RF model and the comparison between its predicted and actual values, the following conclusions can be drawn: 1) The random forest model demonstrates a satisfac-

tory simulation performance for wind speed at both heights. Specifically, the mean absolute error (MAE) and root mean square error (RMSE) are both below 2 m/s, while the forecast accuracy (FA) generally exceeds 75%. In addition, the scatter points for wind speed at each height are predominantly distributed within the reasonable error margin of  $y = x \pm 2$ , further confirming the model's high prediction accuracy and reliability. 2) The trend equations between predicted and actual values are  $y = 0.99x - 0.05$  at 0.5 m and  $y = 0.92x + 0.09$  at 3 m, both of which are close to the ideal line  $y = x$ . The coefficients of determination (R) are 0.667 and 0.619, respectively, indicating a good fit of the random forest model. 3) As the height increases, the simulation accuracy improves notably: the MAE decreases from 1.39 m/s to 1.95 m/s, and the RMSE decreases from 1.95 m/s to 1.36 m/s. Meanwhile, the FA rises significantly from 77.70% to 87.72%.

**Table 4.** Indicators of the RF Model.

| Height (m) | MAE (m/s) | RMAE (m/s) | FA (%) |
|------------|-----------|------------|--------|
| 0.5        | 1.39      | 1.95       | 77.70  |
| 3          | 1.03      | 1.36       | 87.72  |



**Figure 3.** Comparison of predicted values and actual values of the RF model at (a) 0.5 m; (b) 3 m.

### 3.5. Comparison of Forecasting Effects of Each Model

To evaluate the prediction performance of each model, we compared their mean absolute error (MAE), root mean square error (RMSE), forecast accuracy (FA), and coefficient of determination (R), as summarized in **Table 5**. The results indicate that: The RF model achieved the best performance in wind speed prediction, exhibiting the lowest MAE (ranging from 1.03 to 1.39 m/s) and RMSE (ranging from 1.36 to 1.95 m/s), the highest FA (ranging from 77.70% to 87.72%), and the largest R (ranging from 0.338 to 0.446). This suggests that the RF model offers superior prediction accuracy and stability compared to the other models. The BP model ranked second, with error metrics similar to those of the RF model (MAE: 1.07 - 1.42 m/s; RMSE:

1.40 - 2.00 m/s; FA: 77.84% - 87.13%), though it had a slightly lower R. The SVM model showed slightly higher errors, with MAE ranging from 1.22 to 1.48 m/s, RMSE from 1.63 to 2.04 m/s, and FA from 78.93% to 82.86%. Notably, the SVM model performed anomalously at the 3m height (R = 0.338), which may indicate overfitting or an inappropriate parameter configuration. Overall, the RF model performs the best, while the BP model and the SVM model have similar simulation effects.

**Table 5.** Model indicators.

| Model | Height (m) | MAE (m/s) | RMAE (m/s) | FA (%) | R     |
|-------|------------|-----------|------------|--------|-------|
| BP    | 0.5        | 1.42      | 2.00       | 77.84  | 0.418 |
|       | 3          | 1.07      | 1.40       | 87.13  | 0.353 |
| SVM   | 0.5        | 1.48      | 2.04       | 78.93  | 0.392 |
|       | 3          | 1.33      | 1.63       | 79.24  | 0.114 |
| RF    | 0.5        | 1.39      | 1.95       | 77.70  | 0.445 |
|       | 3          | 1.03      | 1.36       | 87.72  | 0.383 |

To evaluate the actual prediction performance of the models, wind speed on May 10, 2024, was forecasted using each model, as illustrated in **Figure 4**. The following observations can be made: 1) At a height of 0.5 m, the measured wind speed and the predictions from the BP, SVM, and RF models all exhibited overestimation during the 10 - 16 hour period. The measured values ranged from 11.5 to 13.6 m/s, while the BP, SVM, and RF models predicted ranges of 7.4 - 9.7 m/s, 9.3 - 12.3 m/s, and 8.9 - 10.5 m/s, respectively. Among these, the SVM model's predictions were closest to the observed values. Similarly, at 3 m height, all models also overestimated wind speed from 10 to 16 hours, with measured values ranging from 7.6 to 10.3 m/s, and BP, SVM, and RF predictions ranging from 5.8 - 7.6 m/s, 6.2 - 8.5 m/s, and 6.7 - 7.1 m/s, respectively. Again, the SVM model showed the closest agreement with measurements. Overall, the predicted wind speed trends from the BP, SVM, and RF models at both heights were consistent with the observed trends. 2) On that day, at 0.5 m height, the mean absolute errors (MAE) of the BP, SVM, and RF models were 2.67 m/s, 2.18 m/s, and 1.94 m/s, respectively. At 3 m height, the MAE values were 1.79 m/s, 1.46 m/s, and 1.35 m/s, respectively. These results indicate that the RF model generally provided more accurate predictions on this windy day. However, between 10 and 18 hours, the SVM model's predictions were closer to the measured values than those of the RF model, suggesting that the SVM model performs better in forecasting strong wind events. The SVM model provides more accurate simulations of strong winds. The main reason for this is its "maximum margin" classification concept, its powerful ability to depict high-dimensional nonlinear relationships, and its effective alleviation of the "sample imbalance" problem through parameter adjustment. This enables it to exhibit superior discriminative ability compared to RF and BP when capturing the unique characteristics of rare but crucial events such as strong winds. The overall high score of RF is driven by its extremely

high accuracy rate when classifying a large number of non-strong wind samples.

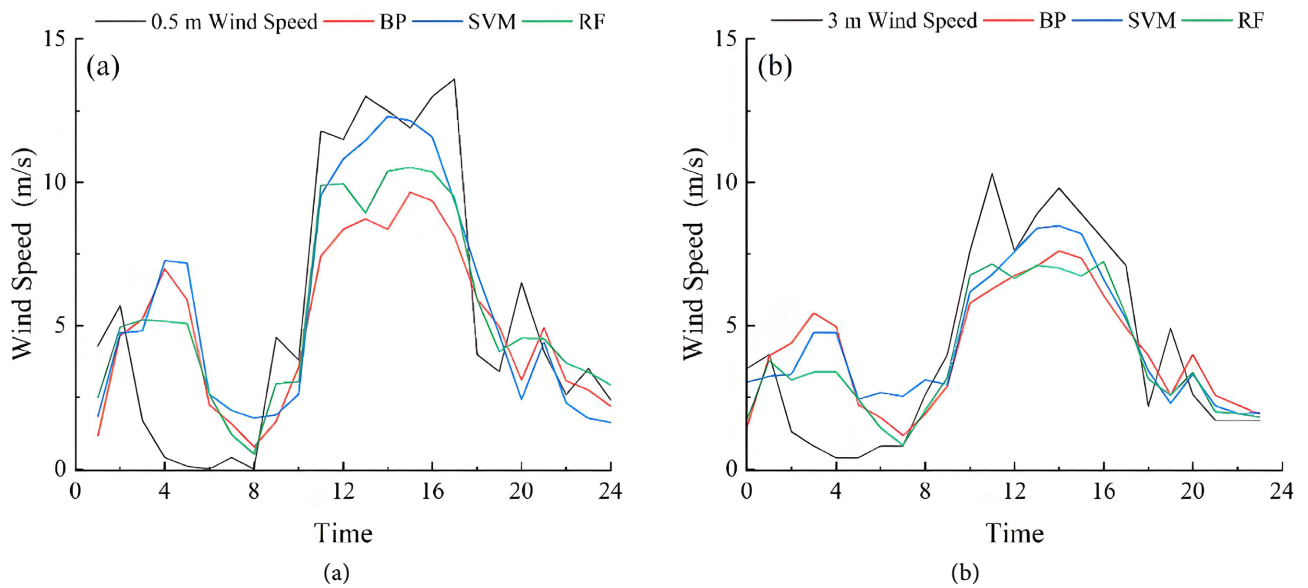


Figure 4. Wind speed forecast for May 10, 2024 at (a) 0.5 m; (b) 3 m.

The wind speed on May 1, 2024, characterized by low wind conditions, was predicted using each model, as shown in Figure 5. The results indicate that at the 0.5 m height, the mean absolute errors (MAE) of the BP, SVM, and RF models were 1.51 m/s, 1.59 m/s, and 1.12 m/s, respectively. At the 3 m height, the corresponding MAE values were 1.06 m/s, 0.84 m/s, and 0.83 m/s. Compared to the BP and SVM models, the wind speed predicted by the RF model was closer to the actual observations. Overall, the prediction performance of the models can be ranked as: RF model > SVM model > BP model.

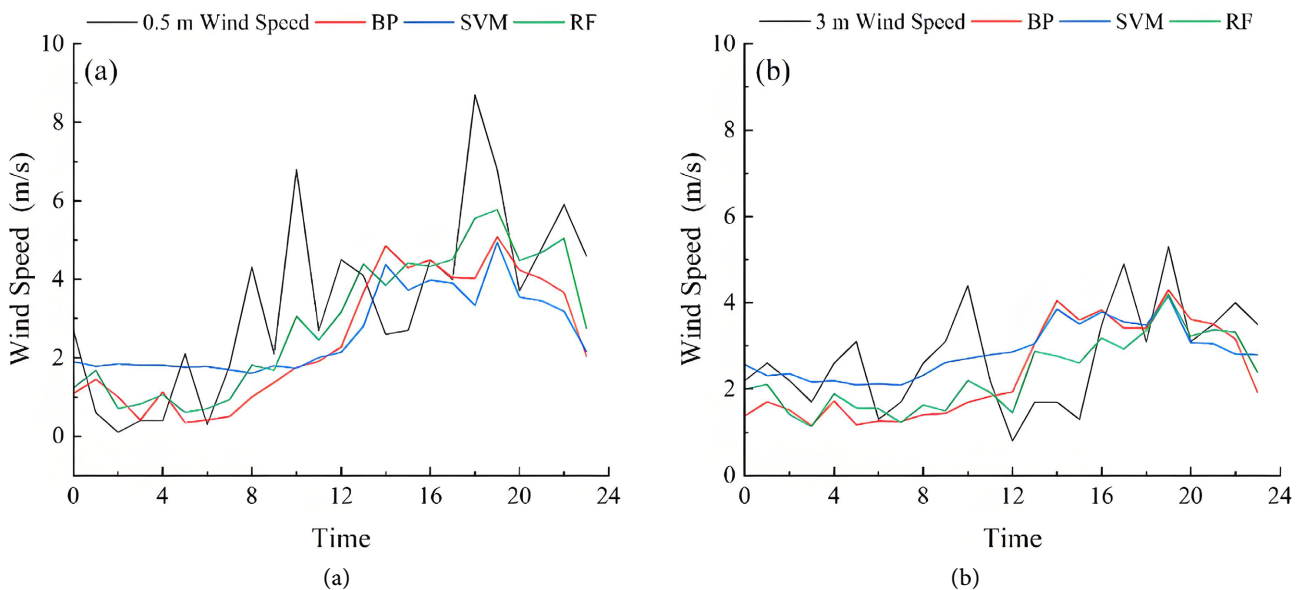


Figure 5. Wind speed forecast for May 1, 2024 at (a) 0.5 m; (b) 3 m.

In conclusion, based on MAE, RMAE, FA, and R, the RF model has the best simulation accuracy and stability, while the SVM model and the BP model are similar. However, case studies reveal that both the SVM and RF models exhibit smaller mean absolute errors in predicted versus actual wind speeds compared to the BP model, demonstrating the overall inferior performance of the BP model. Therefore, the comprehensive simulation performance can be ranked as RF model > SVM model > BP model. Nonetheless, the SVM model shows particular strength in simulating strong wind events.

#### 4. Conclusion

Based on hourly meteorological data from the Sanying National Station and the Guyuan Greenhouse Station between April 2024 and April 2025, this study developed wind speed prediction models using machine learning methods. The results are summarized as follows: 1) Wind speed at the Guoyuan Greenhouse Station exhibited the strongest correlation with that at the national station (0.489 - 0.560), followed by temperature and 24-hour accumulated rainfall (0.116 - 0.336). In contrast, negative correlations were observed with humidity and air pressure (-0.207 - -0.099), while correlations with shorter-term rainfall (15-minute and 1-hour accumulations) were negligible. 2) All three models (BP, SVM, and RF) showed MAE and RMSE values below 1.5 m/s and FA values above 75% across all heights. Scatter plots revealed a concentration of points along the  $y = x$  line, indicating satisfactory simulation accuracy and stability. The average absolute error (MAE) and root mean square error (RMSE) at an altitude of 0.5 meters are relatively large, while the prediction accuracy (FA) is relatively small. This indicates that the wind speed at 0.5 meters is close to the ground level, with more strong winds. It is also more affected by surface roughness and turbulent effects, resulting in greater randomness and making the model more difficult. 3) Case analysis demonstrated that the overall simulation performance ranked as RF model > SVM model > BP model. However, the SVM model exhibited superior accuracy in predicting strong wind events.

#### Funding

This work was jointly supported by Ningxia Natural Science Foundation Project (2023AAC02088), Liangshan Prefecture 2023 Annual Science and Technology Planning Project (23ZDYF0182).

#### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

#### References

Chen, C., Wang, D., Li, G. et al. (2012). Validation and Evaluation of GRAPESMeso Element Forecasts in Winter at High Altitudes in Complex Terrain. *Meteorology*, 38, 657-668.

- Chen, J., Ma, J., & Nie, J. (2018a). Analysis of the Release of Main Meteorological Disaster Warning Signals in Ningxia. *Ningxia Engineering Technology*, 17, 300-305+310.
- Chen, L., Zhu, X., Hu, B. et al. (2018b). Research on the Typhoon Disaster Prediction Model of Ningbo City Based on BP Neural Network. *Acta Meteorologica Sinica*, 41, 668-675.
- Deng, Q. (2024). Research on Non-Stationary Temperature Change Prediction Model Based on Improved Autoregressive Theory. *Science & Technology Information*, 22, 241-244.
- Fan, L., Ji, Y., & Wu, G. (2021). Research on Temperature Prediction Model in Greenhouse Based on Improved SVR. *Journal of Physics: Conference Series*, 1802, Article 042001. <https://doi.org/10.1088/1742-6596/1802/4/042001>
- Fan, Y., Song, Z., Yang, X. et al. (2024). A Method for Fitting Temperature Changes during Föhn Processes in Southern Mountainous Areas of Sichuan Using BP Neural Network. *Research on Plateau and Mountainous Climate*, 44, 95-101.
- Gao, Q., Xue, X., & Duan, A. (2003). Research on Temperature Characteristics and Variation Patterns in Solar Greenhouses. *Journal of Irrigation and Drainage*, No. 6, 50-53.
- Ihsan, U., & Mehmet, B. (2023). Daily Air Temperature Forecasting Using LSTM-CNN and GRU-CNN Models. *Acta Geophysica*, 72, 2107-2126. <https://doi.org/10.1007/s11600-023-01241-y>
- Jin, Z., Fu, G., Huang, H. et al. (2011). Research on Temperature Prediction Model for Pineapple Greenhouses Based on BP Neural Network. *Chinese Journal of Agricultural Meteorology*, 32, 362-367.
- Kamal, P., Bhaskar, B., & Sandipan, K. (2021). An Efficient Decision-Making Approach for Short Term Indoor Room Temperature Forecasting in Smart Environment: Evidence from India. *International Journal of Information Technology & Decision Making*, 20, 733-774. <https://doi.org/10.1142/s0219622021500164>
- Kang, Y., Deng, R., Chen, C. et al. (2021). Research on the Impact Degree of Flood Release Atomization at Xiangjiaba Hydropower Station on Temperature. *Research on Plateau Mountainous Climate*, 41, 108-112.
- Landis, F. (2025). Consumption-Based Responsibility for Global Warming and the Distribution of Impacts of Swiss Climate Policy. *Swiss Journal of Economics and Statistics*, 161, Article No. 8. <https://doi.org/10.1186/s41937-025-00140-x>
- Li, W., Li, Y., Du, X. et al. (2022). Research on Field Fire Prediction Based on Improved Grey Correlation Prediction Based on BP Neural Network. *Academic Journal of Computing & Information Science*, 5, No. 6.
- Lin, J., & Yang, A. (2019). Economic Loss Assessment of Typhoon Disasters Based on BP Neural Network and VSM. *Disasters*, 34, 22-26.
- Liu, R., Yuan, S., & Han, L. (2022). Evaluation and Analysis on the Temperature Prediction Model for Bailing Mushroom in Jizhou, Tianjin. *Agriculture*, 12, Article 2044. <https://doi.org/10.3390/agriculture12122044>
- Lyu, H., Wang, W., Xiao, W., Zhang, K., & Lee, X. (2025). Global Land-Lake Thermal Contrast in a Warming World. *Journal of Climate*, 38, 3861-3876. <https://doi.org/10.1175/jcli-d-24-0559.1>
- Ma, J. (2020). Initial Exploration on Ways to Improve the Economic Benefits of Greenhouse Farms in Ningxia. *Business Insights*, No. 34, 1-2.
- Romero-Gomez, P., Lopez-Cruz, L. I., & Choi, Y. C. (2008). Analysis of Greenhouse Natural Ventilation under the Environmental Conditions of Central Mexico. *Transactions of the ASABE*, 51, 1753-1761. <https://doi.org/10.13031/2013.25309>
- Shen, Y., & Wei, H. (2016). Wind Speed Interval Prediction for Wind Farms Based on RBF

- Neural Network. *Industrial Control Computer*, 29, 55-57.
- Tan, F., He, L., Zhao, X. et al. (2022). Construction of Wind Disaster Indicators for Solar Greenhouses and Plastic Greenhouses. *Meteorology*, 48, 1186-1194.
- Wang, B., Wang, J., Cao, P. et al. (2023). Hourly Heavy Precipitation Correction Analysis Based on Random Forest Model. *Research on Plateau Mountainous Climate*, 43, 91-98.
- Xu, L., Feng, H., & Chen, J. (2009). Research on the Rainstorm Warning Index of Sichuan Based on Ensemble Forecast Products. *Research on Plateau Mountainous Climate*, 29, 44-48.
- Yan, X., Li, J., Yang, K. et al. (2014). Comparison of Two Modes in Wind Speed Prediction Application of Wind Farms. *Arid Meteorology*, 32, 1015-1024.
- Yang, X., Zhang, L., & Wang, Z. (2022). Risk Assessment of Economic Losses Caused by Typhoon Disasters in County Areas Based on Machine Learning Algorithms. *Journal of Tropical Meteorology*, 38, 651-661.
- Zeng, W., Zhang, J., Wang, H., & Zhou, H. (2018). Supplier Development and Its Incentives in Infrastructure Mega-Projects: A Case Study on Hong Kong-Zhuhai-Macao Bridge Project. *Frontiers of Engineering Management*, 5, 88-97.  
<https://doi.org/10.15302/j-fem-2018077>
- Zhang, G., Ding, X., He, F. et al. (2024). Construction of a Greenhouse Air Temperature Prediction Model Based on LSTM-AT. *Transactions of the Chinese Society of Agricultural Engineering*, 40, 194-201.
- Zhang, X., & Lou, W. (2010). Evaluation of the Number of House Collapses Caused by Typhoon Disasters Based on Principal Component Support Vector Machine. *Anhui Agricultural Sciences*, 38, 6339-6341.