



# A robust error correction method for numerical weather prediction wind speed based on Bayesian optimization, variational mode decomposition, principal component analysis, and random forest: VMD-PCA-RF (version 1.0.0)

Shaohui Zhou<sup>1</sup>, Chloe Yuchao Gao<sup>2,3</sup>, Zexia Duan<sup>1</sup>, Xingya Xi<sup>4</sup>, and Yubin Li<sup>1</sup>

<sup>1</sup>Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, 210044, China

<sup>2</sup>Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan University, Shanghai, 200438, China

<sup>3</sup>Shanghai Key Laboratory of Ocean-Land-Atmosphere Boundary Dynamics and Climate Change, Fudan University, Shanghai, 200438, China

<sup>4</sup>School of Atmospheric Sciences, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, 519082, China

**Correspondence:** Chloe Yuchao Gao (gyc@fudan.edu.cn)

Received: 9 May 2023 – Discussion started: 31 May 2023

Revised: 8 September 2023 – Accepted: 12 September 2023 – Published: 2 November 2023

**Abstract.** Accurate wind speed prediction is crucial for the safe and efficient utilization of wind resources. However, current single-value deterministic numerical weather prediction methods employed by wind farms do not adequately meet the actual needs of power grid dispatching. In this study, we propose a new hybrid forecasting method for correcting 10 m wind speed predictions made by the Weather Research and Forecasting (WRF) model. Our approach incorporates variational mode decomposition (VMD), principal component analysis (PCA), and five artificial intelligence algorithms: deep belief network (DBN), multilayer perceptron (MLP), random forest (RF), eXtreme gradient boosting (XGBoost), light gradient boosting machine (lightGBM), and the Bayesian optimization algorithm (BOA). We first predict wind speeds using the WRF model, with initial and lateral boundary conditions from the Global Forecast System (GFS). We then perform two sets of experiments with different input factors and apply BOA optimization to tune the four artificial intelligence models, ultimately building the final models. Furthermore, we compare the aforementioned five optimal artificial intelligence models suitable for five provinces in southern China in the wintertime: VMD-

PCA-RF in December 2021 and VMD-PCA-lightGBM in January 2022. We find that the VMD-PCA-RF evaluation indices exhibit relative stability over nearly a year: the correlation coefficient ( $R$ ) is above 0.6, forecasting accuracy (FA) is above 85 %, mean absolute error (MAE) is below  $0.6 \text{ m s}^{-1}$ , root mean square error (RMSE) is below  $0.8 \text{ m s}^{-1}$ , relative mean absolute error (rMAE) is below 60 %, and relative root mean square error (rRMSE) is below 75 %. Thus, for its promising performance and excellent year-round robustness, we recommend adopting the proposed VMD-PCA-RF method for improved wind speed prediction in models.

## 1 Introduction

Sustainable energy plays a vital role in reducing carbon footprint and increasing system reliability (Hanifi et al., 2020). As renewable energy sources have a negligible carbon footprint, they have become the preferred choice for many industries in the power sector (Dhiman and Deb, 2020). Among these sources, wind energy is a crucial low-carbon energy

technology with the potential to become a sustainable energy source (Tascikaraoglu and Uzunoglu, 2014). In 2022, the global wind power capacity reached 906 GW, with a 9 % year-on-year increase due to a newly installed capacity of 77.6 GW. The global onshore wind market increased by 68.8 GW, while facing a 5 % growth decline compared to the previous year. Such change is attributed to a slowdown in China and the US, the world's two largest wind markets that account for over two-thirds of the world's onshore wind farm installations (Joyce and Feng, 2023). The instability and unpredictability of wind power generation can lead to instability in the power system. In addition, the decline of the wind energy market also makes it more challenging to improve the accuracy of wind speed forecasts. An accurate wind speed prediction method is needed to reduce the instability risk of the power system and the economic loss of wind power enterprises (Huang et al., 2019). Therefore, accurate and stable wind speed prediction (WSP) is very important for the safe and stable operation of the power grid system and for improving the utilization rate of wind energy and economic development (Guo et al., 2021; Xiong et al., 2022; Tang et al., 2021).

Current WSP algorithms are primarily categorized into physical algorithms (Zhao et al., 2016), statistical algorithms (Wang and Hu, 2015; Barthelmie et al., 1993), machine learning (ML) algorithms (Huang et al., 2019; Salcedo-Sanz et al., 2011; Ma et al., 2020), and hybrid algorithms (Deng et al., 2020; Xu et al., 2021; Zhao et al., 2019; Xiong et al., 2022; Tang et al., 2021). Physical methods, such as numerical weather prediction (NWP), are commonly used in wind speed forecasting. NWP, which accounts for atmospheric processes and physical laws, solves discrete mass, momentum, and energy conservation equations along with other fundamental physical principles, establishing itself as a widely adopted and reliable physical method. Currently, the High-resolution Limited Area Model (HIRLAM) (Služenikina and Männik, 2016), the European Centre for Medium-Range Weather Forecasts (ECMWF) model, the Fifth-Generation Mesoscale Model (MM5) (Salcedo-Sanz et al., 2009), and the Weather Research and Forecasting (WRF) model (Skamarock et al., 2021) are extensively utilized for wind speed prediction. However, NWP modeling faces challenges due to the selection of parameterization schemes, such as model microphysics and systematic errors, which exhibit temporal and spatial differences and uncertainties. These uncertainties hinder the accuracy of NWP models in wind speed prediction, making it difficult to meet the rising demands of the grid system (Zhao et al., 2019; Xu et al., 2021).

Studies have demonstrated that enhancing the accuracy of numerical weather prediction (NWP) models and correcting prediction errors can effectively minimize the errors associated with wind speed prediction. These research endeavors have typically sought to optimize the physical and dynamic parameters of the NWP model (Cheng et al., 2013), refine the model structure (Jiménez and Dudhia, 2012), or improve

the accuracy of model inputs through preprocessing and denoising techniques (Xu et al., 2015). Additionally, improving initial field error through methods, such as target observation and data assimilation (Williams et al., 2013), can also minimize wind speed errors predicted by NWP models.

Physical methods are generally more appropriate for long-term wind speed prediction, such as those 48–72 h in advance, while their practical application in short-term forecasting is limited (Zhao et al., 2019; Deng et al., 2020; James et al., 2018). In contrast, statistical methods utilize historical data to establish a relationship between input and output variables and are therefore well suited for short-term wind speed prediction. They are usually time series models, such as autoregressive moving average (ARMA) (Erdem and Shi, 2011) and autoregressive integrated moving average (ARIMA) (Wang and Hu, 2015), whereas filtering models (Cassola and Burlando, 2012; Chen and Yu, 2014), machine learning models (Hu et al., 2013), and hybrid models (Huang et al., 2019) have been gradually developed to further improve wind speed prediction accuracy.

With purely statistical models becoming less suitable for wind speed predictions beyond 6 h, the use of a combination of physical and statistical methods has gained growing interest (Zjavka, 2015; Xu et al., 2021). The error correction model improves the accuracy of the NWP model by training on the relationship between the NWP predictor variables and the observed correlation variables (Sun et al., 2019). However, traditional error prediction models rely solely on historical wind speed sequences as input factors (Deng et al., 2020; Guo et al., 2021) and do not incorporate the characteristic meteorological factors forecasted by the WRF model. Studies have shown that considering all relevant historical meteorological factors can lead to more accurate predictions compared to only taking into account historical wind speed (Z. Zhang et al., 2019). Therefore, it is crucial to include meteorological characteristic factors as input in the prediction model.

For an error prediction model, wind speed is the most important input factor. Traditionally, the error prediction model uses historical wind speed data as input, without any feature selection. Feature selection methods, such as filtering methods, are commonly used in time series analysis. Currently, empirical mode decomposition (EMD) (Liu et al., 2018; Guo et al., 2012), ensemble empirical mode decomposition (EEMD) (Wang et al., 2017), wavelet decomposition (WD) (Y. Zhang et al., 2019), variational mode decomposition (VMD) (Hu et al., 2021; D. Zhang et al., 2019), and other filtering methods are used to select key features in the wind speed data. As mentioned above, studies have shown that these feature selection methods can effectively extract the hidden features in the wind speed series to improve wind speed prediction accuracy. However, despite the effectiveness of wind speed filtering methods in wind speed prediction, only a few studies have applied these methods to

the correction of wind speed errors in NWP forecasting (Xu et al., 2021; Li et al., 2022).

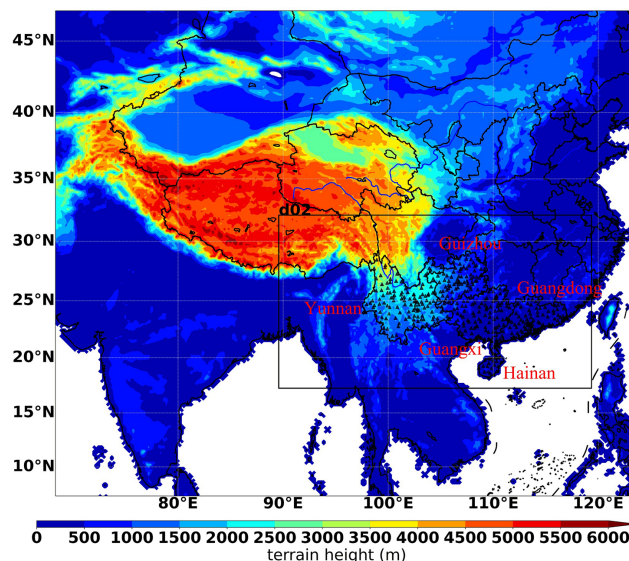
In addition, traditional error correction methods generally adopt linear regression (Dong et al., 2013), multiple linear regression (Liu et al., 2016), machine learning (Salcedo-Sanz et al., 2011), and deep learning algorithms (Z. Zhang et al., 2019). However, the efficacy of machine learning and deep learning algorithms is highly dependent on the selection of model parameters (Guo et al., 2021; Xiong et al., 2022). The Bayesian optimization algorithm (Li and Shi, 2010; Guo et al., 2021) is considered a relatively advanced algorithm for optimizing model parameters and has been widely used in MATLAB and Python packages.

In this study, we investigate a multi-step wind speed forecasting model that combines NWP simulation and an error correction strategy. We present two sets of experiments divided into three steps. (1) We use the first group of experiments to extract hidden features from various meteorological elements forecasted by NWP; the second group of experiments mainly focuses on the wind speed forecast of NWP, and the VMD-PCA algorithm is used to extract the hidden features in the forecasted wind speed. Each set of experimental input factors is matched with the actual 10 m wind speed data of 410 stations in time and space. (2) We employ four advanced machine learning algorithms optimized by the BOA algorithm and DBN deep learning algorithm to train the two groups of experiments and perform 5-fold cross-validation. (3) We analyze six distinct wind speed error indicators to compare and identify the most suitable wind speed error correction schemes for the five southern provinces (Yunnan, Guizhou, Guangxi, Guangdong, Hainan) in winter and throughout most of the year. The remainder of this paper is organized into sections discussing the effects of the BOA-VMD-PCA approach, the interpretability of RF feature importance, and the stability analysis of the proposed models.

## 2 Data and methods

### 2.1 Data

The observed data come from the China Meteorological Administration Land Data Assimilation System (CLDAS-V2.0) real-time product dataset. According to the description of the documents on the official website ([https://data.cma.cn/data/cdcdetail/dataCode/NAFP\\_CLDAS2.0\\_RT.html](https://data.cma.cn/data/cdcdetail/dataCode/NAFP_CLDAS2.0_RT.html), last access: 13 October 2023), the dataset is constructed through the integration of multiple sources, including ground and satellite data, and is refined using advanced techniques such as multi-grid variational assimilation, physical inversion, and terrain correction. This dataset exhibits superior quality in comparison to other products, offering higher spatial and temporal resolutions. The target observation data include 2 m air temperature, 2 m specific humidity, 10 m wind speed, surface



**Figure 1.** WRF model simulation area elevation diagram (d02 represents the nested area of the second layer of the WRF model, and the black triangles represent the meteorological sites).

pressure, and precipitation. These data are processed by the China Meteorological Public Service Center to an equivalent latitude and longitude grid scale, covering a geographical range of 15–32.97° N and 94–120.97° E. The spatial resolution of the grid is 0.03° × 0.03° (3 km by 3 km), and the temporal resolution is 1 h. The China Meteorological Public Service Center applied the nearest neighbor interpolation for precipitation and bilinear interpolation for the other four meteorological elements with downscaling from 3 km to 410 sites. We select the 10 m wind speed data of 410 sites, as illustrated in Fig. 1.

## 2.2 Methods

### 2.2.1 WRF simulation

The WRF 4.2 model (Skamarock et al., 2021), developed by the National Center for Atmospheric Research (NCAR), represents a new generation of mesoscale numerical models with numerous applications in research forecasting. We use the WRF model with forcing from the 0.25° × 0.25° Global Forecast System (GFS) model developed by the National Centers for Environmental Prediction (NCEP). We use the first 90 h of the daily GFS forecast initialized at 06:00 UTC (hereafter all time zones are provided as UTC), with 3 h output, to provide initial and boundary conditions for a daily 42 h WRF forecast, analyzing the 18–42 h forecast and discarding the first 18 h as spinup. Surface static data, such as terrain, soil data, and vegetation coverage, are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite with a resolution of 15 s (approximately 500 m). Incorporating a two-layer grid nesting configuration,

the forecast area is illustrated in Fig. 1. The WRF configuration process is detailed in Table 1. Given that the timescale of the meteorological station data in the study area is 1 h, the forecast data time interval of the WRF model is also set to 1 h. As a widely used numerical weather forecast model, the WRF model is suitable for weather studies from a few meters to several thousand kilometers. Therefore, this paper uses the WRF model to predict 10 m wind speed as the input factor for the error correction model.

**2.2.2 Variational mode decomposition**

As a new filtering method, VMD is robust in feature selection. The VMD algorithm decomposes a time series signal into several intrinsic mode functions (Isham et al., 2018). The sum of the modes equals the original signal, and the sum of the bandwidths is the smallest. The analysis signal is calculated using the Hilbert transform to estimate the modal bandwidth. The optimization model is described as

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

where  $K$  is the total number of modes;  $u_k$  is the decomposed  $k$ th mode;  $\omega_k$  is the corresponding center frequency; and  $v$  is the time series signal, representing the wind speed sequence predicted by the WRF model in this study.

The above-constrained problem can be transformed into an unconstrained problem using the Lagrangian function:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \omega_k^{n+1} \\ &= \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \sum_{k=1}^K \|\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right. \\ &\quad \left. \times e^{-j\omega_k t} \right\|_2^2 + \|v(t) - \sum_{k=1}^K u_k(t)\|_2^2 + \left\langle \lambda(t), v(t) - \sum_{k=1}^K u_k(t) \right\rangle, \quad (2) \end{aligned}$$

where  $\alpha$  is the penalty parameter and  $\lambda(t)$  is the Lagrange multiplier.

Then we update  $u_k$ ,  $\omega_k$ , and  $\lambda$  using the alternating direction method of the multiplier:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{v}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}, \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[ \hat{v}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega) \right], \quad (5)$$

where  $\tau$  is the update parameter.

When the accuracy (left side of the following expression) meets the following condition,  $u_k$ ,  $\omega_k$ , and  $\lambda$  would stop updating:

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon, \quad (6)$$

where  $\varepsilon$  is the tolerance of the convergence criterion.

The VMD algorithm is implemented to decompose the wind speed signal predicted by the WRF model. When using multiple sub-signals instead of the original signal, more features of the wind speed can be obtained. Therefore, it is beneficial to improve the prediction accuracy when using the sub-signal as input to the error correction model (Xu et al., 2021; Li et al., 2022).

**2.2.3 Principal component analysis**

Subsequences obtained by VMD usually have several illusory components. Using PCA to extract the principal components of subsequences increases the number of features input into the model and reduces the dimension of the data decomposed by VMD. When principal components (PCs) are used as the input of the error prediction algorithm, the PCs fully reflect the characteristics of the subsequence and reduce the model complexity. The PCs  $y_k, k = 1, 2, \dots, K$  of the subsequence matrix  $\mathbf{U}$  and the cumulative contribution rate  $\eta_n$  of the first  $n$  principal components are expressed as

$$y_k = c_k' \mathbf{U}, \quad (7)$$

$$\eta_n = \frac{\sum_{k=1}^n \lambda_k}{\sum_{k=1}^K \lambda_k}, \quad (8)$$

where  $c_k$  is the corresponding characteristic unit vector, with  $k = 1, 2, \dots, K$ ;  $\lambda_k$  is the characteristic root, with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$ .

**2.2.4 Evaluation indicators**

There are many commonly used predictive effect evaluation indicators. This article uses the following evaluation indicators: correlation coefficient ( $R$ ), root mean square error (RMSE), mean absolute error (MAE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and forecasting accuracy (FA). Six error indicators are used to evaluate the correction results of short-term wind speed forecasts of wind farms. The formula for calculating the error index is as follows:



Table 1. WRF configuration scheme.

Model (version)	WRF (V4.2)	
Domains	D1	D2
Horizontal grid points	600 × 500	967 × 535
Δx (km)	9	3
Vertical layers	58	
Longwave radiation	RRTMG (Iacono et al., 2008)	
Shortwave radiation	RRTMG (Iacono et al., 2008)	
Land surface	Noah LSM (Chen et al., 1997)	
Surface layer	MYJ (Janjić, 1994)	
Microphysics	Thompson (Thompson et al., 2008)	
Boundary layer	MYJ (Janjić, 1994)	
Cumulus	Tiedtke (Tiedtke, 1989; C. Zhang et al., 2011)	

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \tag{10}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \tag{11}$$

$$rRMSE = \left[ \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}{\left(\frac{1}{n} \sum_{i=1}^n y_i\right)}} \right] \times 100\%, \tag{12}$$

$$rMAE = \left( \frac{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|}{\left(\frac{1}{n} \sum_{i=1}^n y_i\right)} \right) \times 100\%, \tag{13}$$

$$FA = N_r / N_f. \tag{14}$$

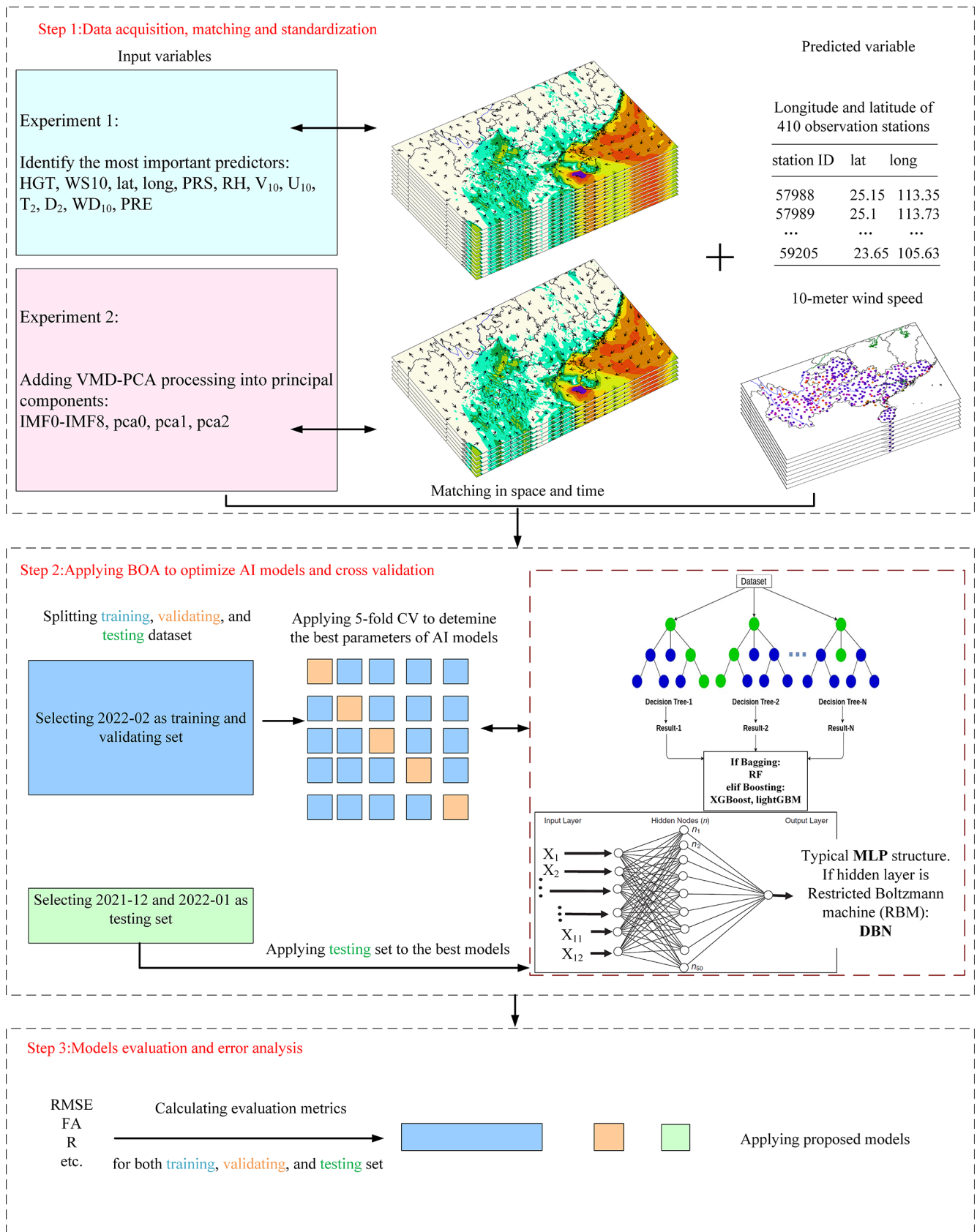
Among them,  $n$  represents the number of samples,  $\hat{y}_i$  represents the  $i$ th predicted value, and  $y_i$  represents the  $i$ th actual value;  $N_r$  represents the number of wind speed absolute errors not greater than  $1 \text{ m s}^{-1}$ , and  $N_f$  represents the number of research samples.

### 2.2.5 Proposed hybrid forecasting algorithms

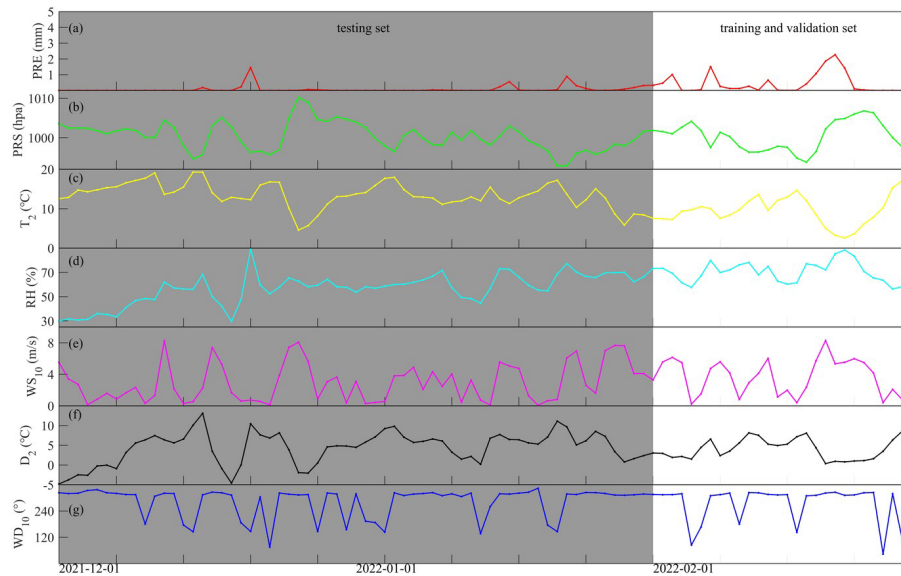
This study used five machine learning algorithms to conduct 10 experiments following two main paths. The first path involves increasing the meteorological variables possibly related to wind speed in the forecast field. The correlation between the WRF-predicted 10 m wind speed and the observed wind speed is the highest. The purpose of the second experimental path is using the VMD-PCA algorithm to dig out the hidden wind speed characteristics of the 10 m forecast wind speed, reduce the input of other meteorological factors such

as  $WD_{10}$  and  $D_2$ , and further prove that the VMD-PCA algorithm is effective before correcting the WRF-predicted wind speed. The overarching goal is to achieve accurate correction of the forecast field wind speed. The flowchart of the artificial intelligence models used to correct the WRF-predicted wind speed for the two main experimental paths is illustrated in Fig. 2 and comprises the following three steps:

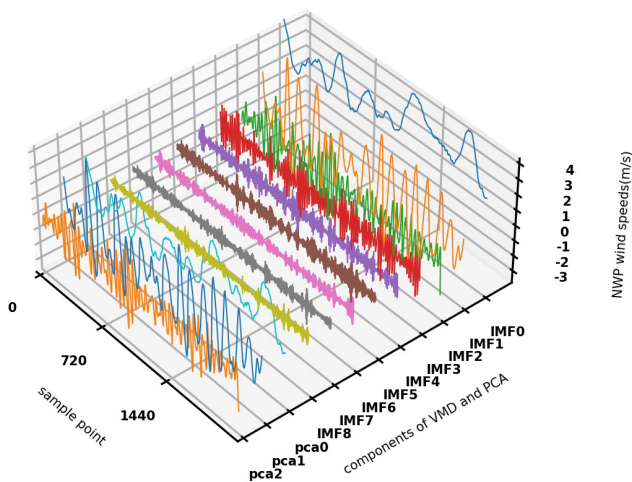
*Step 1 – data fusion, cleaning, and standardization.* As depicted in Fig. 2, this paper proposes two distinct experimental paths, with the primary difference being the selection of input variables. In experiment 1, as shown in Fig. 2, 12 sets of data are selected from the WRF forecast field, including altitude (HGT), 10 m wind speed ( $WS_{10}$ ), latitude (lat), longitude (long), surface pressure (PRS), relative humidity (RH), 10 m meridional wind ( $V_{10}$ ), 10 m zonal wind ( $U_{10}$ ), 2 m temperature ( $T_2$ ), 2 m dew point temperature ( $D_2$ ), 10 m wind direction ( $WD_{10}$ ), and hourly precipitation (PRE). Experiment 2 derives eight sets of data by reducing the selected WRF field forecast data to include only altitude, 10 m wind speed, latitude, longitude, surface pressure, relative humidity, 2 m temperature, and hourly precipitation. The focus is on unearthing hidden characteristic information of forecast wind speed. In this experiment, the wind speed is decomposed into nine intrinsic mode functions (IMFks;  $k = 0, 1, 2, \dots, 8$ ) using VMD. Subsequently, a low-dimensional wind speed vector is extracted from the nine IMF components via PCA dimensionality reduction (pca0, pca1, pca2), and all data are concatenated to construct the input factors for the model in experiment 2. The time points in the dataset where missing values are located are eliminated. Experiment 1 (experiment 2) standardizes 12 sets of meteorological elements (8 sets of meteorological elements in Fig. 3, 9 IMF components, and 3 PCA vectors in Fig. 4) and wind speed observation data, respectively. Standardization addresses the issue of varying meteorological factor values during training, which may result in different



**Figure 2.** Flowchart of the artificial intelligence (AI) model used to correct WRF-predicted wind speeds in the two main experimental pathways.



**Figure 3.** Daily average hourly rainfall (a), surface pressure (b), 2 m temperature (c), 2 m relative humidity (d), 10 m wind speed (e), 2 m dew point temperature (f), and 10 m wind direction (g) which are located at Guangdong Lechang Station from 1 December 2021, to 28 February 2022. (February 2022 represents the training and verification sets, and December 2021 to January 2022 represents the testing set).



**Figure 4.** Three-dimensional view of 12 wind speed components after VMD and PCA processing of the 10 m forecast wind speed at Lechang Station in Guangdong from 1 December 2021 to 28 February 2022.

contributions. In this paper, the 24 h forecast data correspond to the observation data of the subsequent 24 h. The dataset spans from 00:00 on 1 December 2021 to 23:00 on 28 February 2022, totaling 2160 h and encompassing 410 weather stations. Consequently, the original dataset comprises  $2160 \times 410$  samples, with each sample containing 12 meteorological features in experiment 1 and 20 input features in experiment 2. While similar

past studies for wind speed correction from NWP models usually use several years for training and at least 1 year for testing, our periods are shorter, but the size of our dataset is sufficient. For example, Sun et al. (2019) used a dataset that contained 1827 d, from January 2012 to December 2016, using 143 grid points with a resolution of  $0.5^\circ \times 0.5^\circ$  predicted by ECMWF, followed by 24 features for each sample, with a training set size of  $1827 \times 143 \times 24$  for each prediction time. Meanwhile, the size of our training set is about  $2160 \times 410 \times 12$ . Therefore, even though it only took us a month to train, for this project, we trained millions of data. Second, the training data we used here were obtained through daily operational runs of numerical weather forecasting, so it would have taken several years to get an equal amount of training data.

*Step 2 – BOA optimization of AI models and cross-validation.* In this study, the dataset is partitioned into training, validation, and test sets in accordance with the time series. February 2022 serves as the training and validation sets, while December 2021 and January 2022 constitute the test set. The training and validation sets are divided based on 5-fold cross-validation. Both experiments employ five machine learning algorithms (DBN, MLP, RF, XGBoost, and lightGBM) to construct distinct machine learning models. Concurrently, this paper utilizes the BOA algorithm to tune the parameters of all models, except for DBN, resulting in the optimal hyperparameters for each model.

*Step 3 – model evaluation and error analysis.* The trained machine learning models are applied to the test set to obtain the revised wind speed data, and ultimately, the accuracy of all models is assessed through the wind speed evaluation index. The ultimate goal here is to identify the best wind speed correction model suitable for the entire year. Accordingly, the generalization of all models is evaluated across other seasonal months of the year, culminating in the selection of the best model.

### 3 Results

#### 3.1 Experiment 1 evaluation

In experiment 1, the BOA optimization algorithm was applied to five AI models to correct the 10 m wind speed forecasted by WRF. There were 12 meteorological element features to establish five different AI models (see Table S1 for the hyperparameters of the five AI models). The training, validation, and testing results for 10 m wind speed are shown in Figs. S1–S5 in the Supplement. It is clear from Table 2 that all models, except the DBN model, can fit the training set data well. The DBN model exhibits the weakest performance on both the training and the validation sets. Alternatively, the lightGBM and XGBoost models demonstrate superior prediction performance on the training set compared to the validation set. The scatters of the training sets of these two models accumulate on the 1 : 1 diagonal, indicating slight overfitting. As shown in Figs. S1–S5d, f, considering different evaluation indices, the revision effects of the five models in 2 months demonstrate that RMSE in January 2022 is generally lower than in December 2021, FA in January 2022 is generally higher than in December 2021, and  $R$  in January 2022 is generally lower than in December 2021. Overall, the prediction performance of the five models in January 2022 surpassed that in December 2021. Furthermore, the lightGBM and RF models exhibited the best performance among the five models in the 2-month test sets, while the DBN model had the least effective correction effect.

As illustrated in Fig. 5a, b,  $WS_{10}$  showed the strongest positive correlation with  $WS_{obs}$ , with the highest  $R$  of 0.51, which was consistent with the highest variable importance value of 31 % (23 %) in experiment 1 (experiment 2). In addition to  $WS_{10}$ , experiment 1 (experiment 2) also had another three dominant variables, namely lat, HGT, and long, with importance values of 16 % (14 %), 15 % (15 %), and 15 % (13 %), respectively. Meanwhile, in experiment 2, IMF0 and pca0 generated by the VMD-PCA algorithm have a good importance value of 9 % and 4 %, and the  $R$  values of them with  $WS_{obs}$  are as high as 0.47 and 0.45.

Concerning the importance of RF characteristics (Fig. 5a, c), it is indisputable that the 10 m wind speed predicted by WRF plays a dominant role in correcting the actual wind speed. The ones following are latitude, longitude, and to-

pographic height, which represent spatial geographic information, and the actual wind speed is closely related to geographic information. Subsequently, relative humidity is of lesser importance. The distribution of the humidity field typically correlates with the movement of the atmosphere, which is also closely related to wind speed. Certain meteorological elements, such as rainfall, 2 m dew point temperature, and 2 m temperature, contribute less importance.

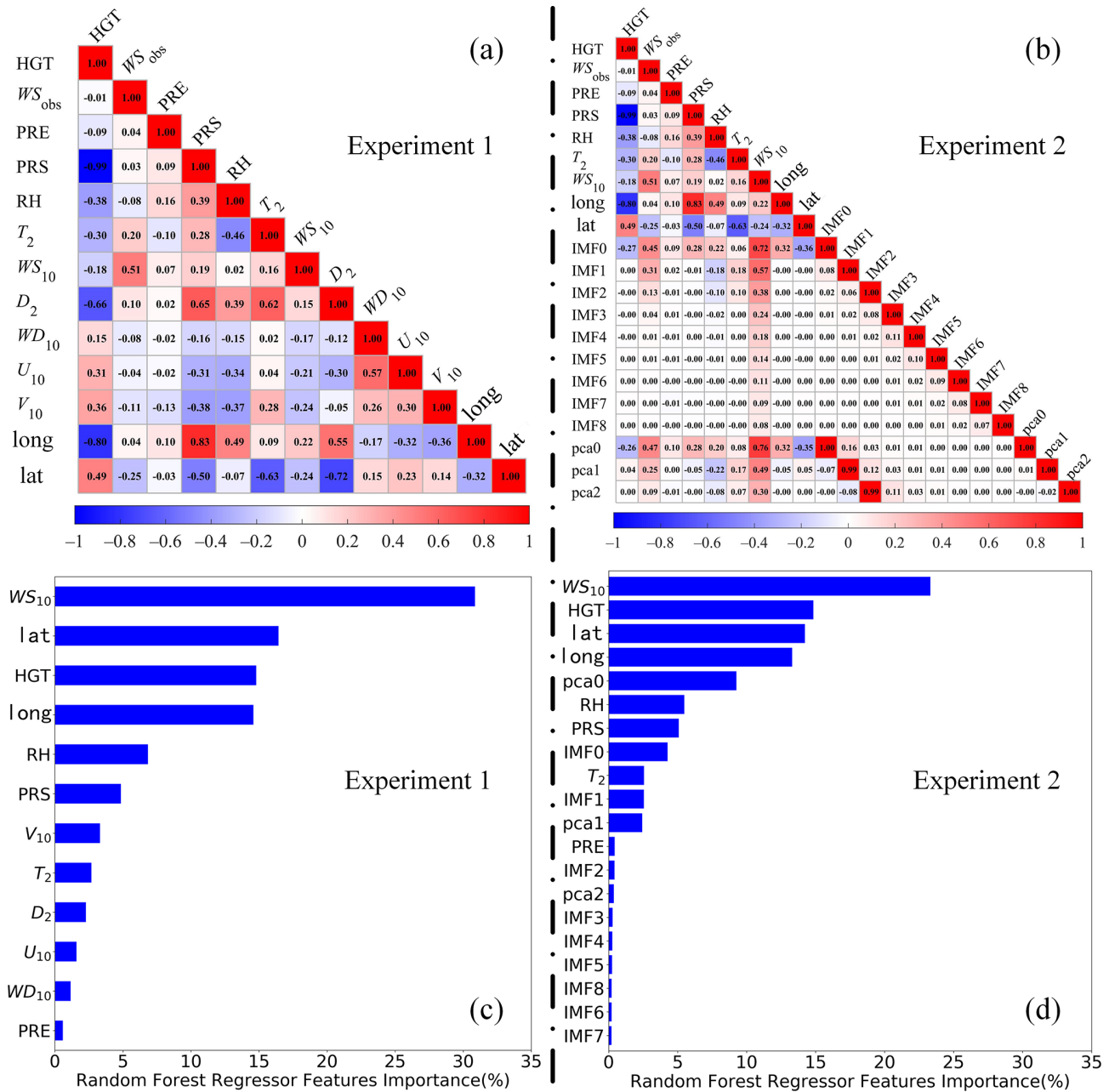
#### 3.2 Experiment 2 evaluation

Experiment 2 builds upon experiment 1, concentrating on the predicted 10 m wind speed by the WRF model. We use the VMD algorithm to decompose the predicted wind speed into nine components and use the PCA algorithm to extract the main three principal components. In the RF feature importance analysis (Fig. 5b, d), it is evident that the VMD algorithm can decompose IMF0 and IMF1, with contributions surpassing those of 2 m temperature and precipitation, respectively. The importance of the pca0 component, after PCA extraction, reaches up to 8 %. What is particularly interesting is that in the correlation analysis, the correlation values between the IMF0 and pca0 components and the actual wind speed are 0.50 and 0.51, which are second only to the forecasted wind speed.

From the indices (RMSE, FA,  $R$ ) of the training and validation sets shown in Table 2, in comparison to the above five artificial intelligence methods, the training results of VMD-PCA-DBN are relatively inferior. VMD-PCA-lightGBM and VMD-PCA-XGBoost models still train the processed data effectively. According to the scatter density figure (Figs. 6a, 7a), the scatters are relatively concentrated on the 1 : 1 line. From the indicators (RMSE, FA,  $R$ ) of the testing set shown in Figs. S6–S8d, f and 6–7d, f, the test results of the five models in experiment 2 in December 2021 and January 2022 show that the error indices of RMSE and FA of each model exhibit a minimal difference in 2 months. Nonetheless, disregarding the  $R$  results, the performance of the five models in December 2021 is inferior to that in January 2022. The diurnal variation scatterplot of 2 months is tested. As is shown in Figs. S6–S8d, f and 6–7d, f, the red scatters represent the nighttime wind speed, which is more concentrated on the 1 : 1 line. In contrast, the blue scatters represent the afternoon wind speed, which is slightly away from the 1 : 1 line. This suggests that the correction effect of the five models (VMD-PCA-lightGBM, VMD-PCA-XGBoost, VMD-PCA-RF, VMD-PCA-DBN, and VMD-PCA-MLP) exhibits a noticeable diurnal variation.

#### 3.3 Comparison of the two experiments

Firstly, all 10 models effectively corrected the 10 m wind speed forecasted by WRF. Tables S2 and S3 represent the evaluation indices of wind speed errors predicted by the 10 models in December 2021 and January 2022. From



**Figure 5.** Schematic diagram of correlation coefficients (represented by the  $WS_{10}$  and input variables) and feature importance (calculated by the scikit-learn Python package) for two sets of experiments. Panels (a) and (c) represent experiment 1, and (b) and (d) represent experiment 2.

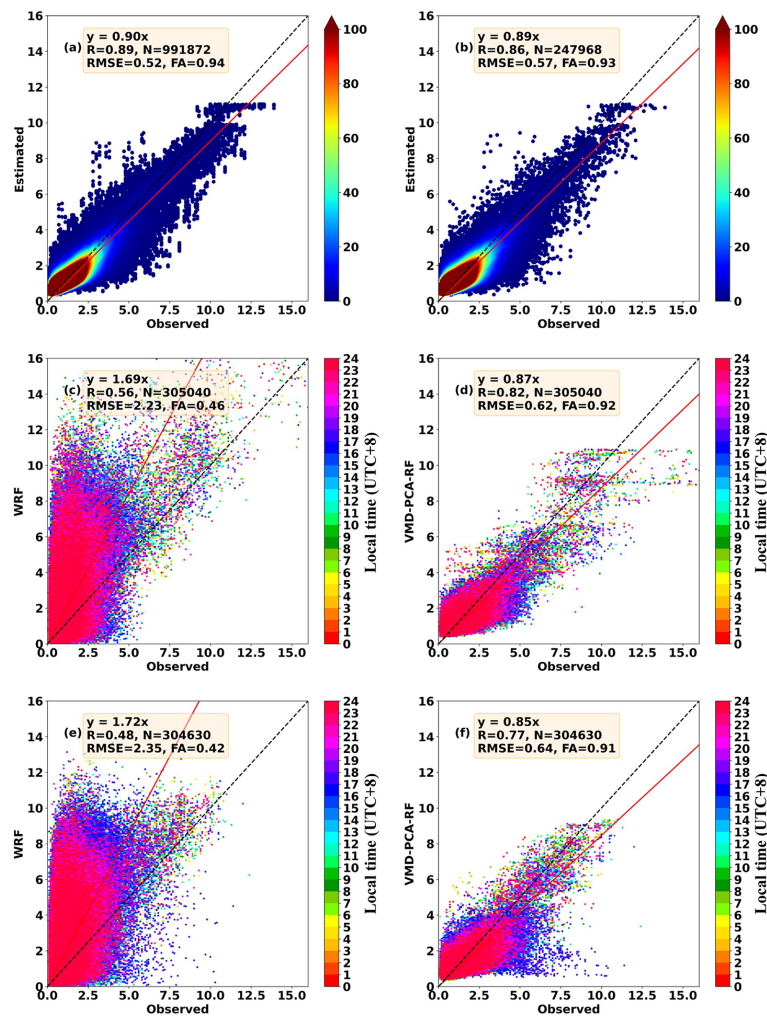
the two tables, it is evident that the VMD-PCA-RF and VMD-PCA-lightGBM models have the best performance in December 2021 and January 2022, respectively, with the most comprehensive performance of the forecast indicators. The MAE, RMSE, rMAE, rMAE, and FA for the two models VMD-PCA-RF (VMD-PCA-lightGBM) were  $0.46 \text{ m s}^{-1}$  ( $0.45 \text{ m s}^{-1}$ ),  $0.62 \text{ m s}^{-1}$  ( $0.63 \text{ m s}^{-1}$ ), 37.36 % (34.75 %), 50.39 % (48.65 %), and 91.79 % (91.49 %) in December 2021 (January 2022). Additionally, based on the

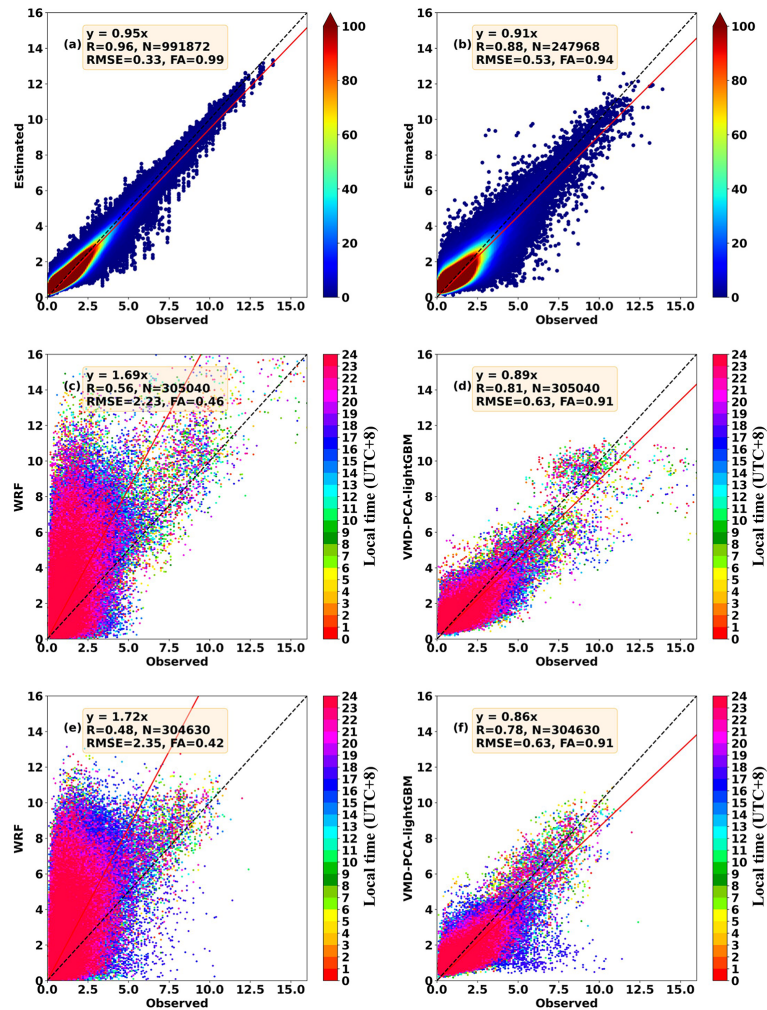
analysis of the Taylor chart (Fig. 8e, f) of the 10 models in Fig. 8, it can also be seen that the scatter distance of VMD-PCA-RF and VMD-PCA-lightGBM models is the closest to the observed dotted black line and the black triangle position. The two models show that the standard deviation is close to the observed wind speed, with the lowest RMSE and the highest  $R$ . Secondly, in the comparison of cumulative probability distributions, all models passed Kolmogorov's 5 % confidence interval test when the interval of wind speed is



**Table 2.** Table of the evaluation indices of wind speed error trained and verified by the 10 models in February 2022.

Model	training set			validation set		
	<i>R</i>	RMSE ( $\text{m s}^{-1}$ )	FA	<i>R</i>	RMSE ( $\text{m s}^{-1}$ )	FA
VMD-PCA-lightGBM	0.96	0.33	0.99	0.88	0.53	0.94
VMD-PCA-XGBoost	0.96	0.31	1.00	0.87	0.54	0.94
VMD-PCA-RF	0.89	0.52	0.94	0.86	0.57	0.93
VMD-PCA-DBN	0.74	0.75	0.87	0.74	0.75	0.87
VMD-PCA-MLP	0.84	0.60	0.91	0.81	0.66	0.90
lightGBM	0.93	0.41	0.98	0.88	0.54	0.94
XGBoost	0.96	0.31	0.99	0.87	0.56	0.93
RF	0.89	0.52	0.94	0.86	0.57	0.93
DBN	0.76	0.73	0.88	0.76	0.73	0.88
MLP	0.85	0.59	0.92	0.83	0.62	0.91

**Figure 6.** The scatter density map compared with the actual 10 m wind speed: (a) 10-fold cross-validation training set of the VMD-PCA-RF model in February 2022 and (b) 10-fold cross-validation validation set of the VMD-PCA-RF model in February 2022. The 24 h scatter map compared with the actual 10 m wind speed: (c) WRF forecasts in December 2021, (d) the VMD-PCA-RF model forecasts in December 2021, (e) WRF forecasts in January 2022, and (f) the VMD-PCA-RF model forecasts in January 2022.



**Figure 7.** The scatter density map compared with the actual 10 m wind speed: (a) 10-fold cross-validation training set of the VMD-PCA-lightGBM model in February 2022 and (b) 10-fold cross-validation validation set of the VMD-PCA-lightGBM model in February 2022. The 24 h scatter map compared with the actual 10 m wind speed: (c) WRF forecasts in December 2021, (d) the VMD-PCA-lightGBM model forecasts in December 2021, (e) WRF forecasts in January 2022, and (f) the VMD-PCA-lightGBM model forecasts in January 2022.

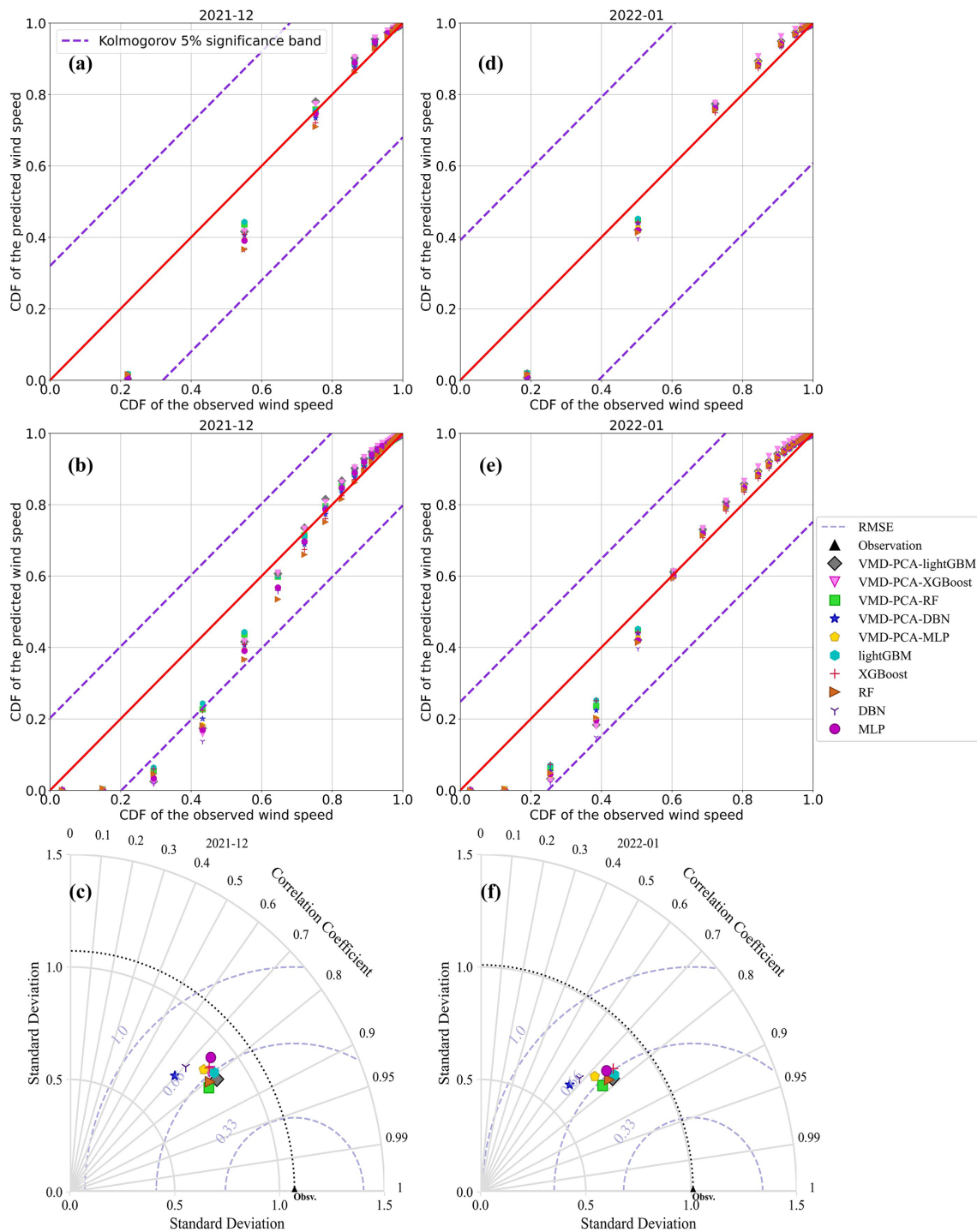
$0.5 \text{ m s}^{-1}$  (Fig. 8a, d). However, when the interval of wind speed is  $0.2 \text{ m s}^{-1}$  (Fig. 8b, e), the VMD-PCA-lightGBM model deviated from Kolmogorov's 5% confidence interval detection in December 2021. This indicates that the VMD-PCA-RF model has a better predictive effect than the VMD-PCA-lightGBM model in December 2021 when the actual wind speed is within the range of  $0.4\text{--}0.8 \text{ m s}^{-1}$ .

### 3.4 Spatial–temporal variations in the best models

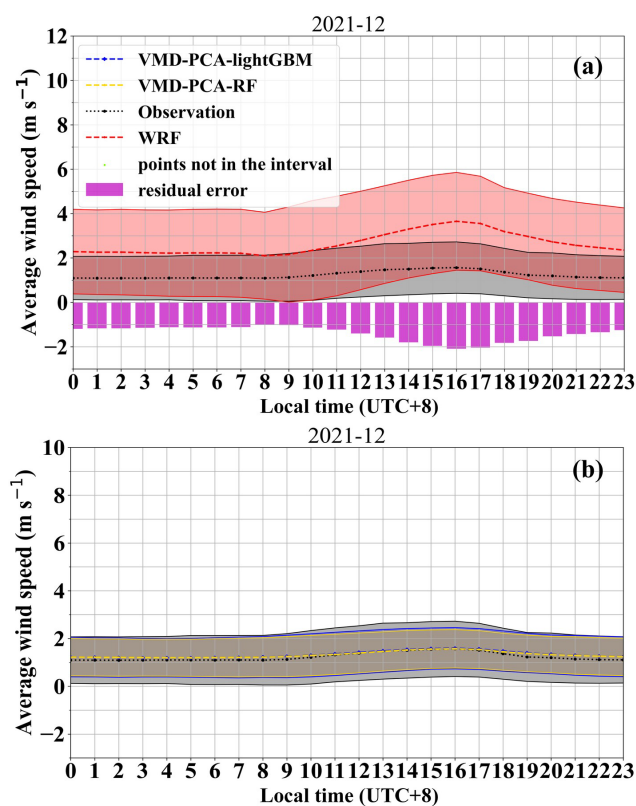
Based on our comparative analysis results, we conclude that the best-performing combination models in December 2021 and January 2022 are VMD-PCA-RF and VMD-PCA-lightGBM, respectively. Figures 9 and S9 show the diurnal variation corrections of the two best models for a given month, as well as the diurnal variation of wind speed in the original WRF forecast. The wind speed of the original WRF

numerical weather forecast shows a noticeable overestimation, which is confirmed in Fig. 7c and e. The scatters of the WRF forecast predominantly deviate towards the upper-left corner, with relatively low correlation coefficients, 0.56 and 0.23, respectively. Furthermore, the wind speed forecast by WRF displays obvious diurnal variation traits, characterized by large errors between the afternoon and evening, specifically between 11:00 and 20:00 (Figs. 9a, S9a). Moreover, the actual average wind speed in January 2022 deviates from the range of 1 standard deviation of the WRF forecast wind speed at 17:00 and 18:00 (Fig. S9a). This demonstrates that the wind speed forecast by WRF is inaccurate and exhibits substantial diurnal variation errors.

After the best model was corrected, the error in diurnal variation was significantly reduced (Figs. 9b, S9b). First, the average wind speed corrected by the best model is essentially consistent with the actual average wind speed curve, with



**Figure 8.** The cumulative distribution probability scatterplots of the actual wind speed and the predicted wind speed of the 10 models in wind speed intervals of  $0.5 \text{ m s}^{-1}$  (a represents December 2021, d represents January 2022) and  $0.2 \text{ m s}^{-1}$  (b represents December 2021, e represents January 2022), respectively; Taylor distribution map (c represents December 2021, f represents January 2022).



**Figure 9.** VMD-PCA-lightGBM, VMD-PCA-RF, and WRF daily variation in predicted and actual wind speeds in December 2021. (The shaded areas represent an interval of 1 standard deviation, which is a 68 % confidence interval).

minimal error and no diurnal variation. Second, the 1 standard deviation range of the corrected and actual wind speeds is also well matched, indicating that the corrected and actual wind speed distributions are consistent. The correction effect at 16:00 and 17:00 on January 2022 is suboptimal, which may be due to the insufficient generalization of the training model and the excessive fluctuation of the actual wind speed at these two time points.

The FA (Figs. 10a, S10a) and RMSE (Figs. 10b, S10b) distribution of the WRF forecast 10 m wind speed at 410 stations in the five southern provinces shows that the 10 m wind speed prediction effect of the WRF model in Yunnan is superior to that in the other four provinces. In the regions of Hainan, Guangxi, and Guangdong, the number of sites with an RMSE for 10 m wind speed forecast in December 2021 ranging from 5.6 to 6.0  $\text{m s}^{-1}$  was significantly higher than in January 2022, especially in coastal areas (Figs. 10b, S10b). In the Yunnan area, the FA of most WRF forecast station 10 m wind speeds exceeds 40 %, and the RMSE value is mostly below 2.4  $\text{m s}^{-1}$ . Conversely, in other regions, such as Guangxi, Guangdong, and Hainan, the terrain is relatively flat. The FA of the 10 m wind speed forecast by WRF is as low as 30 % at some stations, and the RMSE reaches up

to 5.4  $\text{m s}^{-1}$ . However, after the VMD-PCA-RF and VMD-PCA-lightGBM models are corrected, the FA of most stations in the five southern provinces is as high as 90 %, and the RMSE is as low as 0.6  $\text{m s}^{-1}$ . Moreover, in Guangxi, Guangdong, and Hainan, where the WRF forecast effect is subpar, the accuracy of the corrected 10 m wind speed by VMD-PCA-RF (VMD-PCA-lightGBM) is significantly improved.

## 4 Discussion

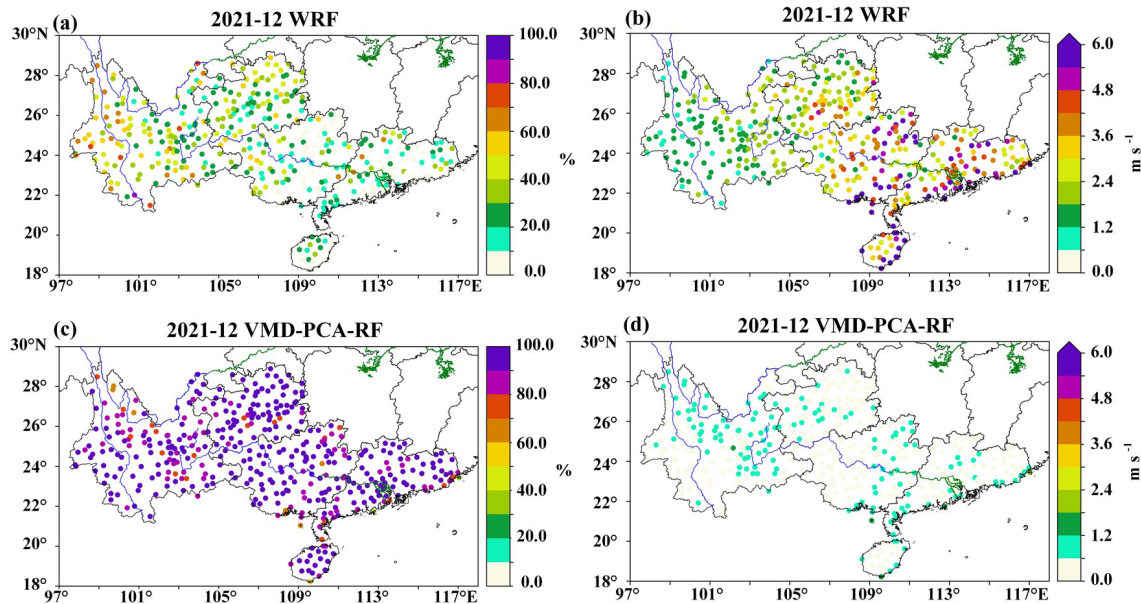
### 4.1 The effects of BOA-VMD-PCA

It is shown in Table S1 that the hyperparameters of the 10 models in the two experiments are different. Since the DBN model is not added to the scikit-learn Python learning package, it is challenging to use the BOA algorithm for tuning parameters. Apart from the DBN model, all the other models are optimized using the BOA algorithm. From the various evaluation indicators in Tables S2 and S3, the DBN model, which does not use the BOA algorithm to adjust the model parameters to obtain an optimal parameter configuration, yields the worst prediction results in December 2021 and January 2022. Moreover, studies (Xiong et al., 2022) have also shown that BOA can further improve the model's prediction accuracy by configuring optimal hyperparameters. The hyperparameters, such as the number of neurons and learning rate in the hidden layer, significantly impact the model's performance. When the same model is applied to different datasets of two experiments, the BOA adaptively obtains the optimal combination of hyperparameters, overcoming the limitations of manual parameter adjustment (Guo et al., 2021). This suggests that the selection of model hyperparameters introduces considerable uncertainty in our prediction results. Therefore, the choice of optimization model parameters represents one source of uncertainty in the correction results, which entails the complexity of parameter selection. However, a more advanced parameter tuning method, such as the BOA tuning algorithm, is essential.

The VMD is used to obtain unknown but meaningful features hidden in the 10 m wind speed sequences predicted using WRF models (Li et al., 2022). In addition, the PCA can extract important components of anemometer subsequences. When the stationary subsequence serves as an input to the error correction model, it contains more valuable information than the previous non-stationary wind speed sequences (Xu et al., 2021).

The complexity of the input factors in this study is one of the sources of uncertainty in the process of correcting WRF prediction results. The input factors of the two experiments are not identical. In the second set of experiments, the input of meteorological factors is reduced based on the first set of experiments, while component information of the 10 m wind speed predicted by WRF is increased. Multiple wind speed





**Figure 10.** FA (a, c) and RMSE (b, d) distribution maps of the VMD-PCA-RF and WRF models at 410 sites in five southern provinces in December 2021.

components processed by VMD-PCA and noise reduction are introduced. Among them, the importance of introduced *pca0* and *IMF0* is approximately 5%. In the 13-month test sets, the correction accuracy of experiment 2 is no less than the results of experiment 1 (Figs. S11, 12), indicating that the 10 m wind speed components introduced by VMD-PCA contribute positively to the correction results.

#### 4.2 RF feature importance

To further understand the feature importance ranking of the RF models, we divided the model prediction results and actual wind speeds of the 410 stations into 20 equal parts according to terrain height above sea level (Fig. 11). First of all, the actual wind speed in December 2021 and January 2022 varies with the height of the station, showing that the lower the height of the station, the more significant the change in wind speed. This relationship is associated with the wind speed profile of the atmosphere, where wind speed increases as height decreases. Secondly, the wind speed during the day is generally greater than the wind speed at night, which is related to the turbulent motion of the atmosphere during the day. Solar radiation causes the atmosphere to mix, resulting in convective movement. The 10 m wind speed at night is affected by the cooling radiation of the surface, and the atmosphere is relatively stable.

The 10 m wind speed predicted by WRF has the highest feature importance in the correction process of the RF models. Input factors with distinct geographic information, such as latitude, longitude, and height, rank highly in feature importance. Similarly, when Sun et al. (2019) used machine

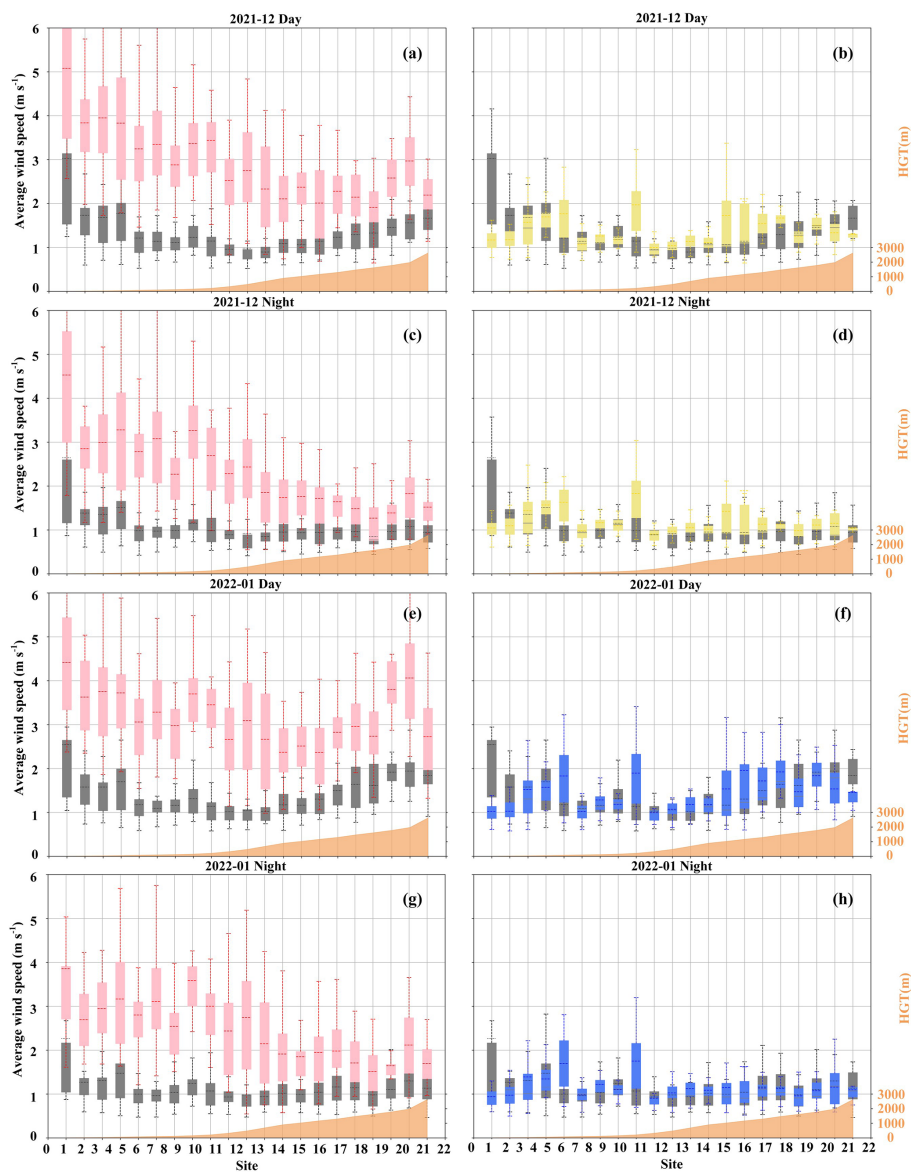
learning to correct the 10 m wind speed predicted by the numerical weather prediction model ECMWF, the characteristic weight of the 10 m wind speed predicted by the model was the highest, followed by the sea–land factor. Moreover, as the 10 m wind speed forecast by WRF increased, the instability of the 10 m wind speed corrected by the 10 machine learning models gradually increased, and the correction accuracy gradually decreased (Fig. 12). This partly explains the higher importance of the 10 m wind speed forecast by WRF.

With 1 km as the center, the measured 10 m wind speed is more variable in areas where the station terrain height increases or decreases. However, the pink box of the 10 m wind speed predicted by WRF becomes wider as the station terrain height decreases (Fig. 11). The distance between the gray box and the pink box is greater as the station terrain height decreases. It shows that the 10 m wind speed predicted by WRF has less accuracy with the station terrain height decreases. The VMD-PCA-RF and VMD-PCA-lightGBM models significantly reduce the variability in the 10 m wind speed predicted by WRF. When the height of the station increases or decreases at 1 km, the correction intensity tends to increase gradually. This further explains the higher importance of the height factor in the RF model training.

#### 4.3 Stability analysis of the proposed models

In order to identify the best model of the five southern provinces and assess the model's stability, we evaluated all 10 models over 13 different months. Figure 13 shows the evaluation histogram of the 10 m wind speed predicted by the 10 models in experiment 1 and experiment 2, as well

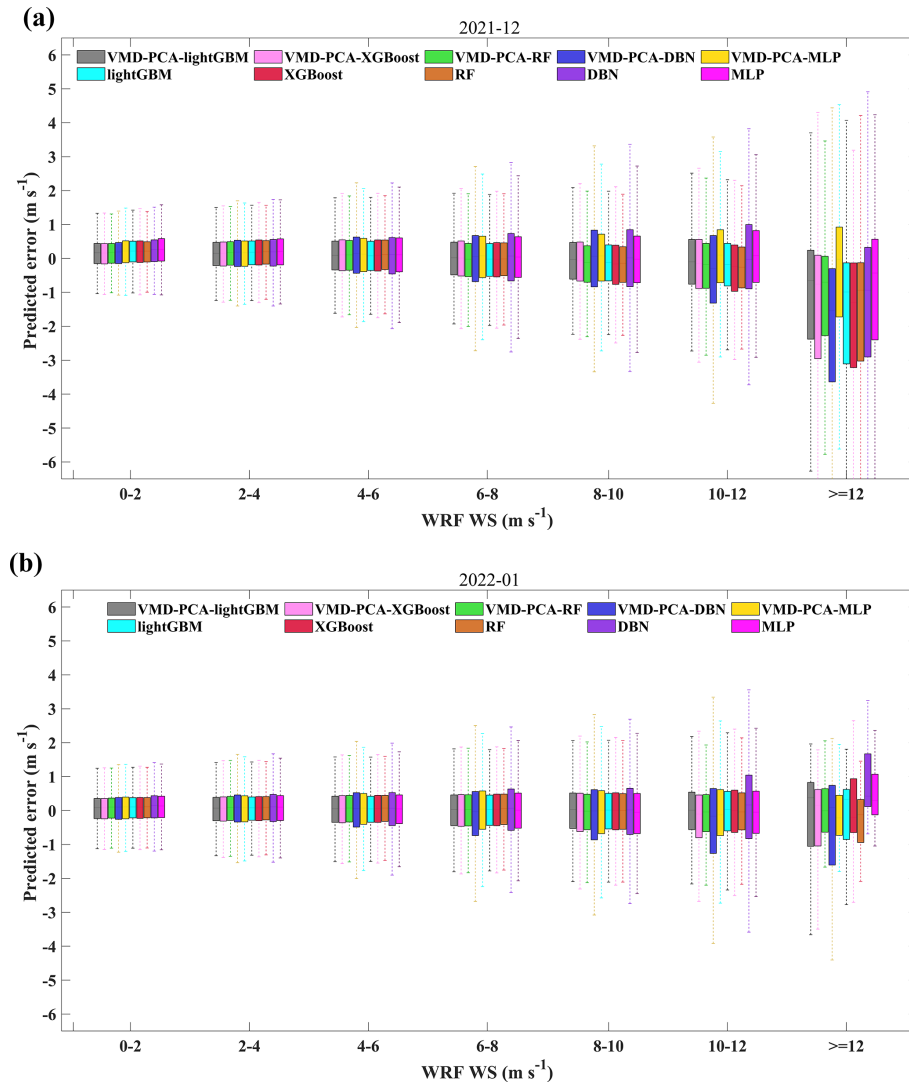




**Figure 11.** The boxplots of the predicted wind speeds of the VMD-PCA-RF (yellow), VMD-PCA-lightGBM (blue), and WRF (pink) models at 20 stations at different height intervals and the boxplots of the actual wind speeds (gray).

as the actual wind speed in various months. Meanwhile, Figs. S11 and S12 can more effectively illustrate the daily changes in the revised results of the 10 models in 13 different months. As shown in Fig. 13, the evaluation indices of the model trained in experiment 2, after VMD-PCA processing, outperform those of the model trained in experiment 1. The RF model demonstrates exceptional robustness, while the MLP model exhibits the poorest performance. VMD-PCA-RF evaluation indices are relatively stable across the 13 months, with a correlation coefficient  $R$  above 0.6, FA above 85 %, MAE below  $0.6 \text{ m s}^{-1}$ , RMSE below  $0.8 \text{ m s}^{-1}$ , rMAE below 60 %, and rRMSE below 75 %. However, the robustness of the VMD-PCA-lightGBM and VMD-PCA-

XGBoost models is inferior to that of the VMD-PCA-RF model, with all six evaluation indices performing worse than the VMD-PCA-RF model as the seasons and months change. In general, VMD-PCA-lightGBM is the superior wind speed correction model for the winter, and VMD-PCA-RF performs the best throughout the entire year in the five southern provinces. In cases where ample machine CPU and other hardware resources, as well as training time, are available, we recommend using VMD-PCA-lightGBM for modeling each season. However, when dealing with limited resources such as a laptop and constrained training time, we recommend using VMD-PCA-RF to train data for a single month, as this yields the most robust correction results.



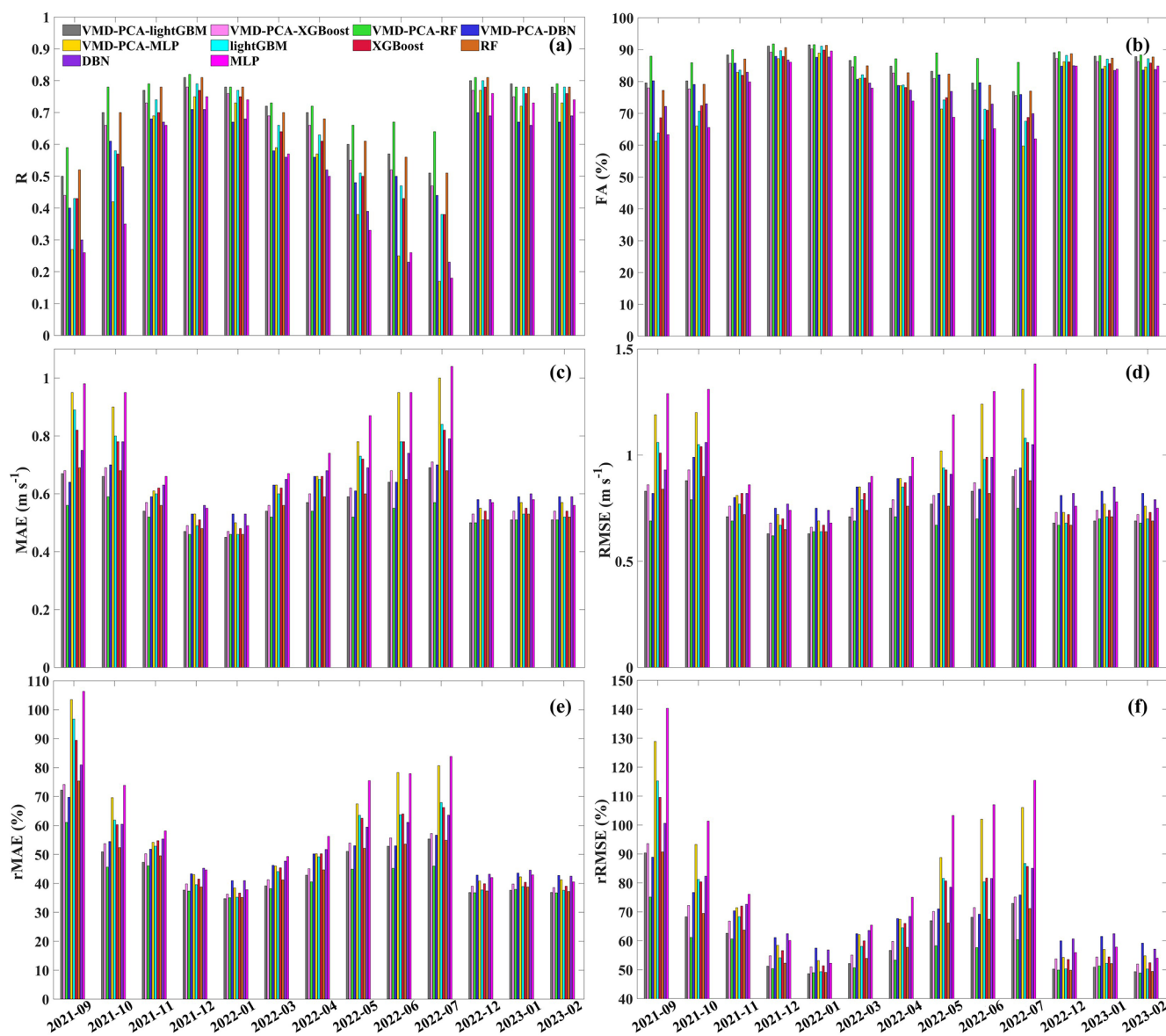
**Figure 12.** The prediction error boxplots of 10 models in different WRF prediction intervals.

## 5 Conclusions

In an effort to enhance the wind speed prediction performance for wind farms, this study developed a WRF-based multi-step wind speed prediction model. A hybrid error correction strategy combining BOA, VMD, PCA, and RF (lightGBM) is proposed to increase the accuracy of WRF simulations. The first group of experiments used various meteorological elements as input factors in a control experiment. In the second group of experiments, the wind speed sequence predicted by the WRF model was decomposed into multiple IMFs using the VMD algorithm for feature extraction. A principal component analysis method is used to extract meaningful principal components from these subsequence IMFs to improve computational efficiency. In the error correction model, RF (lightGBM) and other algorithms are used

to train the relationship between different input factors and the actual wind speed error, respectively.

Through a case analysis of 410 stations in five southern provinces in China, the following conclusions can be drawn: (1) the machine learning models tuned by the BOA-VMD-PCA algorithm exhibit a positive impact on wind speed error correction; (2) feature importance analysis revealed that the top eight contributing factors for correcting WRF-forecasted wind speed include WRF forecast 10 m wind speed ( $WS_{10}$ ), latitude, longitude, altitude,  $pca_0$  ( $pca_0$  physically represents the lowest-frequency wind speed series after PCA treatment of all IMF $_k$  ( $k = 0, 1, 2, \dots, 8$ ) sub-series with reduced dimension), humidity, pressure, IMF $_0$  (IMF $_0$  physically represents the wind speed stationary series with a specific lowest center frequency after the original wind speed series has been processed by VMD); (3) VMD-PCA-RF and VMD-PCA-lightGBM are the most suitable



**Figure 13.** Evaluation histograms of 10 m wind speed predicted by the 10 models in different months in experiment 1 and experiment 2 (a, b, c, d, e, and f represent  $R$ , FA (%), MAE ( $\text{m s}^{-1}$ ), RMSE ( $\text{m s}^{-1}$ ), rMAE (%), and rRMSE (%), respectively).

wind speed correction algorithms for December 2021 and January 2022, respectively. The MAE, RMSE, FA, rMAE, rRMSE, and  $R$  of the corrected wind speed and the actual wind speed are  $0.46$  ( $0.45$ ),  $0.62 \text{ m s}^{-1}$  ( $0.63 \text{ m s}^{-1}$ ),  $37.36 \%$  ( $34.75 \%$ ),  $50.39 \%$  ( $48.65 \%$ ),  $91.79 \%$  ( $91.49 \%$ ), and  $0.82$  ( $0.78$ ); and (4) the proposed wind speed correction model (VMD-PCA-RF) demonstrates the highest prediction accuracy and stability in the five southern provinces in nearly a year and at different heights. VMD-PCA-RF evaluation indices for 13 months remain relatively stable:  $R$  is above  $0.6$ , FA is above  $85 \%$ , MAE is below  $0.6 \text{ m s}^{-1}$ , RMSE is below  $0.8 \text{ m s}^{-1}$ , rMAE is below  $60 \%$ , and rRMSE is below  $75 \%$ . In future research, the proposed VMD-PCA-RF algo-

rithm can be extrapolated to the  $3 \text{ km}$  grid points of the five southern provinces to generate a  $3 \text{ km}$  grid-corrected wind speed product.

*Code and data availability.* The code and model are available as a free-access repository on Zenodo at <https://doi.org/10.5281/zenodo.8108889> (Zhou, 2023). The data are available as a free-access repository on Zenodo at <https://doi.org/10.5281/zenodo.8108889> (Zhou, 2023).

*Supplement.* The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-16-6247-2023-supplement>.

*Author contributions.* SZ developed the software, visualized the data, and prepared the original draft. SZ and CYG developed the methodology and carried out the formal analysis. XX and SZ validated data. SZ, CYG, XX, ZD, and YL reviewed and edited the text. All authors have read and agreed to the published version of the paper.

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

*Acknowledgements.* We thank the Second Tibetan Plateau Scientific Expedition and Research program and the National Natural Science Foundation of China for supporting this work. We are also very grateful to the editor and anonymous reviewers for their insightful comments and suggestions.

*Financial support.* This research has been supported by the Second Tibetan Plateau Scientific Expedition and Research program (grant no. 2019QZKK0102) and the National Natural Science Foundation of China (grant no. 42175082).

*Review statement.* This paper was edited by Mohamed Salim and reviewed by four anonymous referees.

## References

- Barthelmie, R. J., Palutikof, J. P., and Davies, T. D.: Estimation of sector roughness lengths and the effect on prediction of the vertical wind speed profile, *Bound.-Lay. Meteorol.*, 66, 19–47, <https://doi.org/10.1007/BF00705458>, 1993.
- Cassola, F. and Burlando, M.: Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, *Appl. Energ.*, 99, 154–166, <https://doi.org/10.1016/j.apenergy.2012.03.054>, 2012.
- Chen, F., Janjić, Z., and Mitchell, K.: Impact of Atmospheric Surface-layer Parameterizations in the new Land-surface Scheme of the NCEP Mesoscale Eta Model, *Bound.-Lay. Meteorol.*, 85, 391–421, <https://doi.org/10.1023/A:1000531001463>, 1997.
- Chen, K. and Yu, J.: Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach, *Appl. Energ.*, 113, 690–705, <https://doi.org/10.1016/j.apenergy.2013.08.025>, 2014.
- Cheng, W. Y. Y., Liu, Y., Liu, Y., Zhang, Y., Mahoney, W. P., and Warner, T. T.: The impact of model physics on numerical wind forecasts, *Renew. Energ.*, 55, 347–356, <https://doi.org/10.1016/j.renene.2012.12.041>, 2013.
- Deng, Y., Wang, B., and Lu, Z.: A hybrid model based on data preprocessing strategy and error correction system for wind speed forecasting, *Energ. Convers. Manage.*, 212, 112779, <https://doi.org/10.1016/j.enconman.2020.112779>, 2020.
- Dhiman, H. S. and Deb, D.: A Review of Wind Speed and Wind Power Forecasting Techniques, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2009.02279>, 2 September 2020.
- Dong, L., Ren, L., Gao, S., Gao, Y., and Liao, X.: Studies on wind farms ultra-short term NWP wind speed correction methods, in: 2013 25th Chinese Control and Decision Conference (CCDC), 2013 25th Chinese Control and Decision Conference (CCDC), Guiyang, China, 1576–1579, <https://doi.org/10.1109/CCDC.2013.6561180>, 25–27 May 2013.
- Erdem, E. and Shi, J.: ARMA based approaches for forecasting the tuple of wind speed and direction, *Appl. Energ.*, 88, 1405–1414, <https://doi.org/10.1016/j.apenergy.2010.10.031>, 2011.
- Guo, X., Zhu, C., Hao, J., Zhang, S., and Zhu, L.: A hybrid method for short-term wind speed forecasting based on Bayesian optimization and error correction, *J. Renew. Sustain. Ener.*, 13, 036101, <https://doi.org/10.1063/5.0048686>, 2021.
- Guo, Z., Zhao, W., Lu, H., and Wang, J.: Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model, *Renew. Energ.*, 37, 241–249, <https://doi.org/10.1016/j.renene.2011.06.023>, 2012.
- Hanifi, S., Liu, X., Lin, Z., and Lotfian, S.: A Critical Review of Wind Power Forecasting Methods – Past, Present and Future, *Energies*, 13, 3764, <https://doi.org/10.3390/en13153764>, 2020.
- Hu, H., Wang, L., and Tao, R.: Wind speed forecasting based on variational mode decomposition and improved echo state network, *Renew. Energ.*, 164, 729–751, <https://doi.org/10.1016/j.renene.2020.09.109>, 2021.
- Hu, J., Wang, J., and Zeng, G.: A hybrid forecasting approach applied to wind speed time series, *Renew. Energ.*, 60, 185–194, <https://doi.org/10.1016/j.renene.2013.05.012>, 2013.
- Huang, Y., Yang, L., Liu, S., and Wang, G.: Multi-Step Wind Speed Forecasting Based On Ensemble Empirical Mode Decomposition, Long Short Term Memory Network and Error Correction Strategy, *Energies*, 12, 1822, <https://doi.org/10.3390/en12101822>, 2019.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, 113, D13103, <https://doi.org/10.1029/2008JD009944>, 2008.
- Isham, M. F., Leong, M. S., Lim, M. H., and Ahmad, Z. A.: Variational mode decomposition: mode determination method for rotating machinery diagnosis, *J. Vibroeng.*, 20, 2604–2621, <https://doi.org/10.21595/jve.2018.19479>, 2018.
- James, E. P., Benjamin, S. G., and Marquis, M.: Offshore wind speed estimates from a high-resolution rapidly updating numerical weather prediction model forecast dataset, *Wind Energy*, 21, 264–284, <https://doi.org/10.1002/we.2161>, 2018.
- Janjić, Z. I.: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sub-layer, and Turbulence Closure Schemes, *Mon. Weather Rev.*, 122, 927–945, [https://doi.org/10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2), 1994.

- Jiménez, P. A. and Dudhia, J.: Improving the Representation of Resolved and Unresolved Topographic Effects on Surface Wind in the WRF Model, *J. Appl. Meteorol. Clim.*, 51, 300–316, <https://doi.org/10.1175/JAMC-D-11-084.1>, 2012.
- Joyce, L. and Feng Z.: Global Wind Report 2023, Global Wind Energy Council, <https://gwec.net/globalwindreport2023>, last access: 9 May 2023.
- Li, G. and Shi, J.: Application of Bayesian model averaging in modeling long-term wind speed distributions, *Renew. Energ.*, 35, 1192–1202, <https://doi.org/10.1016/j.renene.2009.09.003>, 2010.
- Li, Y., Tang, F., Gao, X., Zhang, T., Qi, J., Xie, J., Li, X., and Guo, Y.: Numerical Weather Prediction Correction Strategy for Short-Term Wind Power Forecasting Based on Bidirectional Gated Recurrent Unit and XGBoost, *Front. Energy Res.*, 9, 836144, <https://doi.org/10.3389/fenrg.2021.836144>, 2022.
- Liu, H., Mi, X., and Li, Y.: An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm, *Renew. Energ.*, 123, 694–705, <https://doi.org/10.1016/j.renene.2018.02.092>, 2018.
- Liu, Y., Wang, Y., Li, L., Han, S., and Infield, D.: Numerical weather prediction wind correction methods and its impact on computational fluid dynamics based wind power forecasting, *J. Renew. Sustain. Ener.*, 8, 033302, <https://doi.org/10.1063/1.4950972>, 2016.
- Ma, Z., Chen, H., Wang, J., Yang, X., Yan, R., Jia, J., and Xu, W.: Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction, *Energ. Convers. Manage.*, 205, 112345, <https://doi.org/10.1016/j.enconman.2019.112345>, 2020.
- Salcedo-Sanz, S., Ángel M. Pérez-Bellido, Ortiz-García, E. G., Portilla-Figueras, A., Prieto, L., and Paredes, D.: Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction, *Renew. Energ.*, 34, 1451–1457, <https://doi.org/10.1016/j.renene.2008.10.017>, 2009.
- Salcedo-Sanz, S., Ortiz-García, E., Pérez-Bellido, Á., Portilla-Figueras, A., and Prieto, L.: Short term wind speed prediction based on evolutionary support vector regression algorithms, *Expert Syst. Appl.*, 38, 4052–4057, <https://doi.org/10.1016/j.eswa.2010.09.067>, 2011.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., and Huang, X.-Y.: A Description of the Advanced Research WRF Model Version 4, NCAR Tech. Note, 145, 1–30, <https://doi.org/10.5065/1dfh-6p97>, 2021.
- Služenikina, J. and Männik, A.: Impact of the ASCAT scatterometer winds on the quality of HIRLAM analysis in case of severe storms, *Proc. Estonian Acad. Sci.*, 65, 177–194, <https://doi.org/10.3176/proc.2016.3.03>, 2016.
- Sun, Q., Jiao, R., Xia, J., Yan, Z., Li, H., Sun, J., Wang, L., and Liang, Z.: Adjusting Wind Speed Prediction of Numerical Weather Forecast Model Based on Machine Learning Methods, *Meteorological Monthly*, 45, 426–436, <http://qxqk.nmc.cn/html/2019/3/20190312.html> (last access: 13 October 2023), 2019.
- Tang, R., Ning, Y., Li, C., Feng, W., Chen, Y., and Xie, X.: Numerical Forecast Correction of Temperature and Wind Using a Single-Station Single-Time Spatial LightGBM Method, *Sensors*, 22, 193, <https://doi.org/10.3390/s22010193>, 2021.
- Tascikaraoglu, A. and Uzunoglu, M.: A review of combined approaches for prediction of short-term wind speed and power, *Renewable and Sustainable Energy Reviews*, 34, 243–254, <https://doi.org/10.1016/j.rser.2014.03.033>, 2014.
- Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, *Mon. Weather Rev.*, 136, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>, 2008.
- Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, *Mon. Weather Rev.*, 117, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2), 1989.
- Wang, C., Zhang, H., Fan, W., and Ma, P.: A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction, *Energy*, 138, 977–990, <https://doi.org/10.1016/j.energy.2017.07.112>, 2017.
- Wang, J. and Hu, J.: A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model, *Energy*, 93, 41–56, <https://doi.org/10.1016/j.energy.2015.08.045>, 2015.
- Williams, J. L., Maxwell, R. M., and Monache, L. D.: Development and verification of a new wind speed forecasting system using an ensemble Kalman filter data assimilation technique in a fully coupled hydrologic and atmospheric model: Data Assimilation in a Coupled Forecasting System, *J. Adv. Model. Earth Syst.*, 5, 785–800, <https://doi.org/10.1002/jame.20051>, 2013.
- Xiong, X., Guo, X., Zeng, P., Zou, R., and Wang, X.: A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization, *Front. Energy Res.*, 10, 905155, <https://doi.org/10.3389/fenrg.2022.905155>, 2022.
- Xu, Q., He, D., Zhang, N., Kang, C., Xia, Q., Bai, J., and Huang, J.: A Short-Term Wind Power Forecasting Approach With Adjustment of Numerical Weather Prediction Input by Data Mining, *IEEE Trans. Sustain. Energy*, 6, 1283–1291, <https://doi.org/10.1109/TSTE.2015.2429586>, 2015.
- Xu, W., Liu, P., Cheng, L., Zhou, Y., Xia, Q., Gong, Y., and Liu, Y.: Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy, *Renew. Energ.*, 163, 772–782, <https://doi.org/10.1016/j.renene.2020.09.032>, 2021.
- Zhang, C., Wang, Y., and Hamilton, K.: Improved Representation of Boundary Layer Clouds over the Southeast Pacific in ARW-WRF Using a Modified Tiedtke Cumulus Parameterization Scheme\*, *Mon. Weather Rev.*, 139, 3489–3513, <https://doi.org/10.1175/MWR-D-10-05091.1>, 2011.
- Zhang, D., Peng, X., Pan, K., and Liu, Y.: A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine, *Energ. Convers. Manage.*, 180, 338–357, <https://doi.org/10.1016/j.enconman.2018.10.089>, 2019.
- Zhang, Y., Chen, B., Pan, G., and Zhao, Y.: A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting, *Energ. Convers. Manage.*, 195, 180–197, <https://doi.org/10.1016/j.enconman.2019.05.005>, 2019.
- Zhang, Z., Ye, L., Qin, H., Liu, Y., Wang, C., Yu, X., Yin, X., and Li, J.: Wind speed prediction method using Shared Weight Long Short-Term Memory Network and



- Gaussian Process Regression, *Appl. Energ.*, 247, 270–284, <https://doi.org/10.1016/j.apenergy.2019.04.047>, 2019.
- Zhao, J., Guo, Z.-H., Su, Z.-Y., Zhao, Z.-Y., Xiao, X., and Liu, F.: An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed, *Appl. Energ.*, 162, 808–826, <https://doi.org/10.1016/j.apenergy.2015.10.145>, 2016.
- Zhao, J., Wang, J., Guo, Z., Guo, Y., Lin, W., and Lin, Y.: Multi-step wind speed forecasting based on numerical simulations and an optimized stochastic ensemble method, *Appl. Energ.*, 255, 113833, <https://doi.org/10.1016/j.apenergy.2019.113833>, 2019.
- Zhou, S.: A robust error correction method for numerical weather prediction wind speed based on Bayesian optimization, Variational Mode Decomposition, Principal Component Analysis, and Random Forest: VMD-PCA-RF (version 1.0.0): Second release of my code, Zenodo [code and data set], <https://doi.org/10.5281/zenodo.8108889>, 2023.
- Zjavka, L.: Wind speed forecast correction models using polynomial neural networks, *Renew. Energ.*, 83, 998–1006, <https://doi.org/10.1016/j.renene.2015.04.054>, 2015.