Assessment of the wind energy resource on the coast of China based on machine learning algorithms

Boming Liu[1], Xin Ma[1], Jianping Guo[2*], Hui Li[1], Shikuan Jin[1], Yingying Ma[1], and Wei Gong[1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430072, China

[2] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

Correspondence to: Dr./Prof. Jianping Guo (Email: jpguocams@gmail.com)

Abstract. Wind is one of the most essential clean and renewable energy sources in today's world. To achieve the goal of carbon emission peakpeaking carbon dioxide emissions and carbon neutrality in China, it is necessary to evaluate the wind energy resources on the coast of China. Nevertheless, the traditional power law method (PLM) relies on the constant coefficient to estimate the high-altitude wind speed at wind turbine hub height. The constant assumption may lead to significant uncertainties in wind energy assessment, given the large dependence on a variety of factors, such as terrain, time and heightaltitude etc. To minimize the uncertainties, we here use three machine learning (ML) algorithms to estimate the high-altitude wind speed at wind turbine hub heightfrom the radar wind profiler (RWP) measurementssurface wind. The radar wind profilerradar wind profilerRWP and surface synoptic observations at eight coastal stations from May 2018 to August 2020 are used as key inputs to investigate the wind energy resource. Afterwards, three ML models and the PLM are used to retrieve the wind speed at 120 m above ground level (WS_{120}). The comparison of results with the observations shows the random forest (RF) is the most suitable model for the estimation of WS_{120}. As suchBased on the WS_{120} from RF model, the diurnal variation of WS_{120} and wind power density (WPD) are then estimatedevaluated based on the WS_{120} from RF model. For land stations, the hourly mean WPD is larger at daytime from 0900 to 1600 local solar time (LST) and reach a peak at 1400 LST. This is mainly due to the influence of the prevailing sea-land breeze. On the contrary, the hourly mean WPD of island stations is relatively large at nighttime during 1800 to 2300 LST. This indicates that the wind energy peaks differ based onby the land surface types. In terms of the spatial distribution of the seasonal mean WS_{120} and WPD along the coastal region of China, the WPDs inat the Yangtze River Delta (YRD) reggion Qingdao, Dayang, and Dongtou are higher than 200 W/m^2 in most seasons, and the WPDs at the coastal regions of Shandong Peninsula and Yangtze River DeltaDongying, Penglai, Qingdao, and LianyungangYRD are much greater than over theat Pearl River Delta region Fuqing and Zhuhai. Thise result shows that the coastal regions of Bohai Sea and Yellow Sea have more abundant wind resources than those of East China Sea and the South China Sea. These findings obtained here provide insights for~~into~~ the development and utilization of wind energy industry on the coast of China in the future.

Key words: wind energy, radar wind profiler, remote sensing, machine learning

1. Introduction

With the rapid economic development of~~ the~~ world, the massive consumption of fossil fuels produces an increasing ~~amount ~~emission of carbon dioxide, sulfur dioxide and other pollutants~~emission~~ (Yuan, 2016; Magazzino et al., 2021~~; Shi et al., 2021; Pei et al., 2022~~). Large amounts of anthropogenic emissions of carbon dioxide and other greenhouse gases~~es~~ are a major driver for the global warming ~~emissions cause the greenhouse effect~~, leading to ever-rising air temperature (Shakun et al., 2012; Shi et al., 2021). To ~~address~~tackle this problem~~e climate change~~, it is increasingly becoming imperative to develop renewable clean energy (Hong et al., 2012). Among the myriad renewable energy resources, wind energy has gained more and more favors because of its abundant availability, good sustainability, and high cost-effectiveness (Li et al., 2018;~~), showing great promising prospect of commercial application (~~Leung et al., 2012). ~~By the end of 2020, the global cumulative installed capacity had reached 743 GW (Global Wind Report, 2021). It is estimated that wind power will account for approximately one-third of the increase in renewable power generation by 2035 (Khatib, 2012). Therefore, accurate estimation of wind energypowerprofile is of great importance.~~

~~In recent decades, wind energy has been extensively studied all over the world. The 2009 edition of Wind Energy Facts elaborated on all aspects of wind energy in Europe at full length (EWEA, 2009). Durisic et al. (2012) used wind data measured at four different heights to analyze the wind power in the South Banat region. Li et al. (2018) analyzed the wind speed and compared wind energy resources at offshore, nearshore, and onshore locations near Lake Erie. Their results showed that the offshore stations can offer more wind energy than onshore stations. Oh et al. (2012) applied the wind speed and wind direction data recorded in three meteorological masts to assess the wind energy and predict the annual energy production at the demonstration offshore wind farm in Korea. Based on 17 years of wind data on Deokjeok-do island, Ali et al. (2018) investigated the wind characteristics during different time scale. Band et al. (2021) estimated the wind energy in the Gulf of Oman by using the near-surface wind data from the Middle East and North Africa-COordinated Regional Downscaling Experiment.~~ As one of the largest energy consuming counties in the world, China is currently facing~~confronted with~~ an increasingly serious energy and climate situation (Khatib et al., 2012). The Chinese government proposes the peak carbon ~~dioxide ~~emissions and carbon neutrality strategy to deal with energy and environmental issues (~~Yuan, 2016~~Pei et al., 2022~~; Costoya, 2021~~~~Pei et al., 2022~~). With the stimulus of policies and the favor of investors, wind power industry in China is flourishing~~has been flourished~~. Therefore, scientific~~the~~ assessment of wind energy resources in China is of great~~ very~~ importan~~t~~ce for the healthy development of wind energy industry in the ~~years~~decades to come.~~ It is reported that the~~

~~top market of the world by the end of 2020 for cumulative wind power installations was China (Global Wind Report, 2021).~~

At present, there are three main methods for wind energy assessment. The first is based on the meteorological tower~~ or mast~~ data (Shu et al., 2016; Liu et al., 2018). The height of the meteorological tower is generally 100~~-~~–300 m above ground level (AGL), equipped with anemometer and other meteorological observation instruments. For instance, Durisic et al. (2012) analyzed the wind energy at four different heights in the South Banat region based on meteorological tower data. But due to the high construction and maintenance costs of meteorological tower~~ are high~~, ~~and~~ it is not suitable for large-scale networking observation. The second is based on ground meteorological station data, which can be used to evaluate the wind energy at the wind turbine hub height by empirical formula (Oh et al. 2012; Liu et al., 2019). Li et al. (2018) investigated the spatial and temporal variations of wind energy near Lake Erie shoreline based on the power law method (PLM). The PLM method generally assumes the wind speed below 150 m in the planetary boundary layer (PBL) varies exponentially with height (Hellman et al. 1914). But due to the influence of inhomogeneous underlying surface, land sea difference and ubiquitous atmospheric turbulence, wind varies constantly and greatly in the vertical (Tieleman 1992; Coleman et al., 2021), posing great challenges and uncertainties to wind energy assessment based on surface observation. The third is based on reanalysis data, such as the fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis system (ERA5). It can provide the hourly wind speed at a specific height (Hersbach et al., 2020; Liu et al., 2020). Compared to near-surface in-situ observations, it has better time continuity and spatial coverage, which can provide data support in the region with poor observational data. The hourly resolution of ERA5 reanalysis has been~~can be~~ used to assess the wind energy in the absence of observational data (Laurila et al., 2021; Gualtieri, 2021). But the spatial resolution of the ERA5 data is 0.25 * 0.25 degree, which is much lower than the high-resolution model output such as the weather research and forecasting (WRF) and the point-based observations. These methods are widely used in the field of wind energy assessment (Li et al., 2018; Band et al. 2021), but each method has certain limitations. Therefore, it is necessary to explore more new observation methods to support a comprehensive assessment of wind energy.

The radar wind profiler (RWP) network of China can measure the wind profiles from the ground surface to a height of 5-8 km AGL (Liu et al., 2019; Guo et al., 2021a), which provide a novel data source for wind energy assessment. Moreover, increasing wind turbine hub height reduces the impact of surface friction, enabling wind turbines to operate in high-quality wind resource environments

(Veers et al., 2019). The RWP can evaluate the wind energy at different heights, which is conducive to the selection of wind turbine hub height. Currently, wind turbine is generally installed at the top of wind mast with a height of 100-120 m AGL, which roughly corresponds to the surface layer (Stull 1988; Veers et al., 2019). This region is where obstructions such as trees, buildings, hills, and valleys cause turbulence and reduce the wind speed (Coleman et al., 2021; Solanki et al., 2022). It leads large uncertainties in the wind profile observations near the ground surface provided by the RWP, largely due to the influence of ground and intermittent clutter (May and Strauch 1998; Allabakash et al., 2019). Therefore, it is necessary to obtain accurate and continuous wind speed at the wind turbine hub height from RWP measurements, which will benefit the robust and scientific assessment of wind energy."

Given the abovementioned problems, we attempt to use machine learning (ML) algorithms to retrieve wind speed at 120 m AGL (WS_{120}) from RWP measurements. The surface in situ wind speed, high-altitude RWP wind speed and corresponding surface meteorological data from May 2018 to August 2020 are collected to develop the ML models. The performance of classical PLM method and three ML models were then compared. Next, the most effective RF model was used to assess the wind power on coast of China. The results of our study can provide useful information for the development of wind energy industry on the coast of China. The observational data is briefly introduced in section 2. The ML model construction and wind energy evaluation method are displayed in section 3. Section 4 discusses the accuracy of the ML models and the variation of wind energy resources. A summary of results is presented in section 5.

2. Materials and Data

2.1 RWP network of China

The RWP is a remote sensing device that can observe the atmospheric wind profiles (Liu et al., 2019). The RWP network of China began to develop as of 2008, and the number of RWP stations increased~~developed~~ to 134 by the end of 2020 (Liu et al., 2020). The time resolution of RWP data can reach minute level. The RWP has high and low detection modes in the vertical direction, and the corresponding vertical resolutions are 120 and 60 m, respectively (Liu et al., 2020). Here, eight RWP stations on the coast from north to south in eastern China are selected, including Dongying, Penglai, Qingdao, Lianyungang, Dayang, Dongtou, Fuqing, and Zhuhai. The spatial distribution of these stations is shown in Fig. 1, marked by red points. Most stations are located on land along the coast, only Dayang and Dongtou are located on island ~~along the coast~~ (Table 1). G~~In terms of~~ geographically location, Dongying, Penglai, Qingdao and Lianyungang are located on Shandong Peninsula ofin the northern Chinanorth of China's coastline, and the other four stations are located on Yangtze River Delta to Pearl River Delta in the south of China's coastline. The hourly wind speed profiles over the eight stations are obtained from 1 May 2018 to 31 August 2020. The RWP data has not been released temporarily, but it can request to Dr. Jianping Guo by (eEmail: (jpguocams@gmail.com).

2.2 Anemometer

The China Meteorological Administration has established more than 2500 weather stations instrumented with wind cup anemometers (Mo et al., 2015). The 10-m wind is measured by this wind cup anemometer can measure the instantaneous wind speed, which is installed 10 m AGL at the weather station, and is installed at 10 m AGLcan measure the instantaneous wind speed (Mo et al., 2015). The sensing part of wind cup anemometer is composed of three or four conical or hemispherical empty cups. It can provide surface wind data with an error of less than 10% (Zhang et al., 2020). This device is also installed aAt eight RWP stations. The 10 m wind speed data can be downloaded in http://www.nmic.cn/data/cdcdetail/dataCode/A.0012.0001.html (, last access: 15 November 2022)., Here, the 10 -m wind speed data at the eight RWP stations were also obtained from 1 May 2018 to 31 August 2020. The 10 -m wind speed data was processed into hourly average value to match the RWP data.

2.3 Radiosonde measurementdata

The RS measurements provides the profiles of wind speed and wind direction twice a day at 0800 and 2000 local solar time (LST) LST (Guo et al., 2020; 2021b; Li et al., 2021; Liu et al., 2022). The accuracy of RS wind speed is within 0.1 m/s in the PBL (Guo et al., 2021b). One noteworthy drawback is that the operational RS can only provide observations of wind profiles only twice per day: 0800 and 2000 local solar time (LST). Note that only the station of Qingdao is equipped with RS and RWP at the same time. The RS data also collectedduring the study period from 1 May 2018 to 31 August 2020, which. The RS10 m wind speed data can be downloaded fromin http://www.nmic.cn/data/cdcdetail/dataCode/B.0011.0001C.html , (last access: 15 November 2022).

2.4 ERA5 data

The fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis system (ERA5) is the reanalysis data combining model data and observations, which provides global, hourly estimates of atmospheric variables (Hoffmann et al., 2019)., The horizontal resolution can reach 0.25 * 0.25 degree31 km, at a horizontal resolution of 31 km andand there areis 137 vertical levels in vertical direction.the reanalysis combines model data with observations from across the world into a

~~globally complete and consistent dataset using the laws of physics (Hoffmann et al., 2019).~~ "ERA5 hourly data on single levels from 1959 to present" is a dataset of ERA5, which ~~contains a series of surface parameters. It~~ can provide a series of surface parameters such as temperature, humidity, pressure and radiation~~the atmospheric, ocean-wave and land-surface quantities~~~~surface parameters~~ etc. ~~on a 0.25 degree * 0.25 degree~~~~0.25 x 0.25-degree grid~~ (Hersbach et al., 2020). It~~This data~~ can be downloaded from the website of~~in~~ https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview ~~,~~ (last accessed on~~:~~ 15 November 2022). ~~Due to~~It is known that the generation of –wind is closely associated with~~generally caused by~~ uneven heating of the Eearth's surface by solar radiation and ~~gradient difference of~~ atmospheric pressure gradient force (Solanki et al., 2022). Therefore~~Here~~, nine parameters that may affect the variation of wind speed have been collected, including charnock coefficient (Char), forecast surface roughness (FSR), friction velocity (FV), dew point (DP), temperature (Temp), pressure (Pres), net solar radiation (Rn), latent heat flux (LHF), and sensible heat flux (SHF). Char, FSR and FV are related to surface roughness and friction, and thus can evaluate the influence of different surface types on the wind speed in the surface layer. DP, Temp and Press are the meteorological parameters associated with wind speed.

Rn, LHF and SHF indicate the solar radiation level, which is directly related to the generation of wind.

According to the longitude and latitude information of the RWP station, the grid where the RWP station is located is selected and those parameters in the corresponding grid are obtained accordingly~~data closest to the station is collected as the input data~~~~surface parameters of the station.~~ These data were ~~also~~ obtained from 1 May 2018 to 31 August 2020 at eight stations.

In addition, the hourly wind data can also be provided by ERA5. The u and v component of wind data at 100 m AGL were also downloaded for wind energy assessment.

3. Methods

In this section~~, the classical PLM method was used to retrieve the WS_{120} based on the surface 10-m wind speed. Three ML algorithms were then attempted to retrieve the WS_{120}. Finally, the method of wind energy evaluation is introduced.~~, we introduce firstly the classical PLM method to retrieve the WS_{120} based on 10 -m wind speed measurement. Then, we describe the three ML algorithms used to retrieve WS_{120}. We finally present the method for evaluating wind energy.

3.1 Power law method

The PLM method i~~wa~~s proposed by Hellman et al. (1914). It assumes~~d~~ that the wind speed below 150 m in the PBL varies exponentially with height. As a result, the wind speed at a certain height is~~has been~~ typically estimated using the following formula~~e~~ (Abbes et al., 2012):

$$v_2 = \quad v_1 \times \left(\frac{h_2}{h_1}\right)^{\alpha} \tag{1}$$

where v_1 and v_2 are the wind speed at height h_1 and h_2, respectively. The α is the wind shear coefficient, which varies with time, altitude, and location (Durisic et al., 2012).

In engineering application, the value of α is determined by the terrain type, and~~ the variation range~~ generally~~is~~ is estimated to range~~d~~ from 0.1 to 0.4 (Li et al., 2018). Here, the general value of α for coastal topography was set to 0.15 based on former studies (Patel et al., 2005; Banuelos et al., 2010).

3.2 Machine learning algorithms

Three ML algorithms, including the k nearest neighbor (KNN), support vector machine (SVM) and random forest (RF), are applied to retrieve the WS_{120}. For the ML algorithms, one of the most important things is to prepare appropriate characteristic values and accurate reference values as input. Here, the input data include 10 m~~surface~~ wind speed (WS_{10}) and direction (WD_{10}) from wind cup anemometer ~~at 10 m AGL~~, wind speed (WS_{300}) and direction (WD_{300}) at 300 m AGL measured by RWP, and nine surface parameters in ERA5. The reference value is the WS_{120} measured by RS. These values are listed in Table S1~~2~~. At Qingdao station, a total of 746 sample data are obtained after data matching. We use 5-fold crossover to train ML models. The specific training process of each model is presented as follows.

3.2.1 KNN~~ nearest neighbor~~

KNN is one of the simplest ML algorithms, which can be used for regression (~~Altman, 1992; ~~Coomans et al., 1982). Its basic idea is to find k nearest neighbors of a ~~sample, and~~sample and assign the average value of these neighbors' attributes to the sample. In this way, the value of the attribute corresponding to the sample can be obtained (Altman, 1992). The schematic diagram of KNN~~ model~~ is shown in Fig. S1a. For a given test sample (orange square),

it need to~~As shown in Fig. 2a.I its basic idea is to~~ find the nearest K training samples (inside the gray circle) in the training dataset based on the distance measurement~~ of a given test sample (orange square)~~, and then assign the average attribute value of the K samples to the test sample~~make predictions~~. ~~As shown in Fig. 2a.~~

Therefore, the setting of K value is important to the accuracy of the KNN~~ model~~.

Here, the KNN algorithm in MATLAB R2020b was used for regression. Figs. 23a and 23d show the tuning parameter process for K value. The K value varies from 1-20 with an interval of 1. Correlation coefficient (R) and root mean square error (RMSE) were used to evaluate the accuracy of the model. We need to set an appropriate K value to maximize R and minimize RMSE. According to the curve of R and RMSE changing with K value, the R reach to 0.77 and RMSE is 2.44 m/s when the K was set to 3. Therefore, the K value was set to 3 for KNN model. The code and usage of KNN model are referred to the MATLAB help centrecenter (https://ww2.mathworks.cn/help/stats/fitcknn.html, last access: 15 November 2022).

3.2.2 SVMupport vector machine

SVM is SVM is a kind of supervised classification algorithm (Cortes et al., 1995), which can also be used in regressiona linear classifier with separation hyperplane with maximal interval (Cortes et al., 1995). InWhen it is used for regression analysis, SVM modelit is to obtain the optimal distance fitting curve. The schematic diagram of SVM model is shown in Fig. S1b.

As shown in Fig. 2b, tThe red line and Δ represent the fitting curve and slack variable, respectively. The penalty parameter (C) is used to measure the loss caused by outliers. For SVM model, iIt needs to obtain the optimal fitting curve with acceptable loss. The principle of SVM model is obtained a hyperplane with maximum geometric interval to divide the training data set correctly.

For SVM model, Tthe penalty parameter (C) is used to measure the loss caused by outliers. The loss of objective function is increased with C value when the sum of relaxation variables of all outliers is certain. Therefore, it needs to take an appropriate C to ensure the performance of SVM model. is a value that must be specified in advance. C value determines the loss caused by outliers. The loss of objective function is increased with C value when the sum of relaxation variables of all outliers is certain. Therefore, it needs to take an appropriate C to ensure the performance SVM model.

Here, we used the SVM algorithm for regression in MATLAB R2020b was used for regression. The tuning parameters process Aiss seen in Figs. 23b and 32e., Tthe value of R increases first and then decreases with the increase of C. On the contrary, the RMSE decreases first and then increases with the increase of C. When C equals 0.75, R reaches the maximum value (0.79) and RMSE reaches the minimum value (1.74 m/s). Therefore, the C value was set to 0.75 for SVM model. In additionaddition,. the code and usage of SVM model are referred to the MATLAB help centre (https://ww2.mathworks.cn/help/stats/fitrsvm.html, last access: 15 November 2022).

3.2.3 RFandom forest

RF model is an ensemble ML methodis one of the cluster classification models (Breiman, 2001), which has been widely used in regressive calculation. It is a method to integrate many decision trees into forests and predict the final results. Schematic diagram of RF model isAs shown in Fig. S12c., It is a method to integrate many decision trees into forests and predict the final results. Tthe RF model is composed of many decision trees, and each decision tree is irrelevant. It is a method to integrate many decision trees into forests and predict the final results. The performance of RF model is determined by the aggregation of the results of all the trees (Ma et al., 2021). For RF model, the number of trees is an important parameter to achieve the optimal performance of the model. –

The further detailed information can be referred to Breiman (2001). Here, we used the RF –algorithm for regression in MATLAB R2020b. –

Figures 32c and 32f show the tuning parameters process for number of tree (N). The N value varies from 1-500 with an interval of 20. It can find that the R increased with N value increased, while the R was almost unchanged when N value is greater than 100. When N equals 300, R reaches the maximum value (0.81) and RMSE reaches the minimum value (1.64 m/s). Therefore, the N value is set to 300 for RF model. In additionaddition,. the code and usage of RF modelalgorithm are referred to the MATLAB help centre for further details (https://ww2.mathworks.cn/help/stats/treebagger.html, last access: 15 November 2022).

3.2.4 Importance of variables

Figure 3 shows the importance analysis of input variables for three ML models. The importance of the variable indicates the dependence of the model on this parameter. The input variables with importance lager than 0.1 wereas marked by red bar. For KNN model, the importance values of WS_{10}, FV and Char are 0.3, 0.3, and 0.15, which areis much larger than that of other inputs. For SVM model, the importance values of WS_{10} and FV are larger than 0.1, while the importance values of other inputs are less than 0.1. For RF model, the importance values of WS_{10}, FV and Char are 0.23, 0.14, and 0.13, respectively. Combined with these results, it found that WS_{10} and FV are mainly input features for these three models. WS_{10} was the surface 10 m wind speed. FV is a theoretical wind speed at the Earth's surface which increases with the roughness of the surface. This result confirms that the WS_{120} is mainly affected by the surface wind speed and terrain typefriction. In addition, the importance values of WS_{10} and FV for KNN model is obviously larger than that of other inputs. By contrary, for RF model, although the importance values of WS_{10} and FV areis large, the importance values of some input variables areis also relatively large with varies from 0.105-0.15. It indicated that the factors such as heat transfer and high-altitude wind speed constraints will also be considered in the inversion process of RF model.

3.32.4 Sensitivity analysisSImportance of characteristicsSensitivity analysisensitivity analysis

To further discuss the generalization of the different methods, we investigated the difference between estimated WS_{120} and observed WS_{120}-, which as a function ofwas found to varyied with WS_{10} and FV (Fig. 4). Since the model is expected to be applicable to various input values, the variation of the deviation with the input features can reflect the generalization of the model (Ma et al., 2021). It was found that the deviation of the PLM and KNN is changed with the increase of WS_{10} and FV. It indicated that the generalization of the PLM and KNN models needed to be improved. The generalization of SVM model iwas better than that of PLM and KNN model, but most of the SVM results tended to beare still overestimated when FV is larger than 0.4 m/s. As for RF model, the deviation wais relatively stable and didoes not change with the increase of WS_{10} and FV. This suggestedIt indicated that the generalization of RF wais better than other three methods. This could be likelyis due to the fact that RF model tends to increasedadds random disturbance in the sample space, parameter space and model space, therebyus reducing the impact of "cases" and improving the generalization ability (Breiman, 2001). MoreoverIn addition, Figure S2 shows the distribution of main input variables of RF model (WS_{10}, FV, Char, SHF, and WS_{300}) at eight RWP stations. The red dashed lines represent the maximum and minimum values of each variable at Qingdao station. In the range of the red line, the RF can provide stable output due to its good generalization ability. It can be found that almost all the input values of other seven stations have appeared in Qingdao station. Therefore, the RF model has sufficient generalization and can be used in other coastal stations. In addition, it is noteworthy that the ML model needs to be reconstructed when most of the inputs at a research site are not within the range of the red line.

it can be seen from the importance analysis that besides WS_{10} and FV, the RF model also depends on Char, SHF and WS_{300}. Fig. S2 shows the difference between estimated WS_{120} and observed WS_{120}

~~varied with these three inputs. The deviation is also stable and does not change with the increase of Char, SHF and WS_{300}. The results show that the RF model has a good generalization to the value changes of all input features. For the other stations, similar coastal environment will not significantly change the size of input features. Therefore, the RF model has sufficient generalization and can be used in other coastal stations. In addition, it notes that the ML model needs to be retrained and set new parameters when using in other environments, such as desert area.~~

3.4~~3~~ Assessment methods of wind energy

For the obtained WS_{120}, a series of indicators need to be used to evaluate wind energy, such as Weibull distribution and wind power density (WPD) (Pishgar et al., 2015). These parameters are commonly used to evaluate the wind energy at a certain station (Fagbenle et al., 2011; Liu et al., 2018).

3.4~~3~~4.1 Weibull distribution

The Weibull distribution can calculate the cumulative probability F(v) and probability density f(v) function of WS_{120} in a certain period of time, which are expressed as follows (Chang et al., 2011):

$$F(v) = 1 - exp\left[-\left(\frac{v}{c}\right)^k\right] \quad (2)$$

$$f(v) = \frac{dF(v)}{dv} = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} exp\left[-\left(\frac{v}{c}\right)^k\right] \quad (3)$$

where *v* is the WS_{120}; k and c are the shape parameter of Weibull distribution~~, and represent the intensity and stability of wind speed, respectively~~. ~~The~~ H~~h~~igher c indicates~~ that~~ larger ~~the~~ wind speed ~~is higher~~, while the k indicates~~ the~~ wind stability~~ (Saleh et al., 2012).~~ Saleh et al. (2012) compared different methods to estimate k and c and pointed out that the moments method is recommended in estimating the Weibull shape parameter. Therefore, we use the moments method to calculate the k and c, which shows as follows (Rocha et al., 2012):

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \quad (4)$$

$$c = \frac{\bar{v}}{\mathcal{T}\left(1+\frac{1}{k}\right)} \quad (5)$$

where $\bar{v}$ and σ are the mean and square deviation of WS_{120}, respectively, and Γ is the gamma function, which has a standard form as follows:

$$\mathcal{T}(x) = \int_0^\infty e^{-u} u^{x-1}\, du \quad (6)$$

3.4~~3~~4.2 Wind power density

The WPD is the wind energy per unit area that the airflow passes vertically in unit time, and generally takes the form like (Akpinar et al., 2005):

$$WPD = \frac{1}{2}\rho c^3 \Gamma\left(\frac{k+3}{k}\right) \quad (7)$$

where ρ is the air density, k and c are the shape parameter of Weibull (equ.4 and 5), and Γ is the gamma function (equ.6).

4. Results and discussion

The accuracy of four methods is firstly evaluated by comparing with RS measurements. The characteristics of WS_{120} were then analyzed based on the results from RF model. Finally, the variation of wind resource was analyzed.

4.1 Intercomparison of WS_{120} using different methods

To evaluate the performance of four methods, the estimated WS_{120} of PLM, KNN, SVM and RF were compared with observation. Given that only Qingdao has RS data, the comparison of different methods was conducted based on the data at Qingdao. Figure 545 shows the comparisons between the observed WS_{120} and the estimated WS_{120} for four methods under different time. RMSE is also displayed on the panel. Overall, tThe R (RMSE) resulting fromof PLM, KNN, SVM and RF models underfor all timeall times were 0.79 (2.33 m/s), 0.81 (1.97 m/s), 0.85 (1.52 m/s), and 0.94 (1.00 m/s), respectively.

No matter from the R or RMSE results, It can be seen from the metrics of R and RMSE it shows that the accuracy of ML models is better than that of PLM, in terms of. Moreover, for each method, the comparison results under 0800 and 2000 LST were similar, irrespective of 0800 and 2000 LST. to that under all time. It indicates that the performance of the four methods does is not varyied with hour of the day. For the PLM method, most of estimated results are underestimated when the observed WS_{120} is high. Meanwwhile, This is due to the PLM methods depends on the exponential relationship between WS_{120} and WS_{10}.

However, This is due to the WS_{120}wind speed in the PBL is affected by turbulence, surface friction and other factors (Tieleman 1992; Solanki et al., 2022). The turbulence caused by inhomogeneous underlying surface can change the wind direction and reduce the horizontal wind speed (Coleman et al., 2021). Especially in coastal areas, the sea land interaction and complex surface types make the variations of near surface wind profiles more complex. Simple exponential relationship between target wind speed and WS_{10} is unable to obtain the WS_{120} with high accuracy, especially at high wind speed condition. Similarly, the most of results from KNN model are underestimated under the high observed WS_{120}. As for KNN, the estimated WS_{120} is obtained by averaging the nearest k points in the training set.

Combine with the result in Fig. 34a, the KNN model is mainly based on WS_{10} and FV to build the model. It is the mapping relationship between target wind speed and surface wind speed. This is essentially similar to the principle of PLM. Therefore, tThe performance R and RMSE of KNN model is has been slightly improved compared with PLM method. It is due to, because the estimated WS_{120} of KNN is obtained by averaging the nearest k points in the training set (Altman, 1992). Essentially, KNN model is to establish the relationship between main characteristics (WS_{10} and FV) and WS_{120}. Therefore, the performance of KNN is similar to PLM.

On the contrary, for SVM and RF models, although the SVM and RF models tend to slightly overestimate small values and underestimate high values, the R and RMSE between the observed WS_{120} and the estimated WS_{120} are significantly improved improvement. Especially for the RF model, the highest R (0.94) and the smallest RMSE (1.00 m/s) show that the RF model is the best model to retrieve WS_{120}. This ismay be dueeo to the fact that it considers more environmental factors, such as SHF, Char, WS_{300}, and WD_{300}. These results indicated that considering heat transfer and high-altitude wind speed constraints in inversion process can improve the accuracy of the model.

Figure 6 shows the comparisons between the observed WS_{120} and the estimated WS_{120} for four methods under different season. The red, green, blue and black represent the spring, summer, autumn and winter, respectively. The PLM performs best in autumn (R=0.83, RMSE=1.95 m/s), with and worst in summer (R=0.72, RMSE=2.37 m/s). The slopes of fitting line at spring, summer, autumn and winter were 0.58, 0.47, 0.72 and 0.8, respectively. It shows that the performance of PLM is affected by seasonal factors, which . This is likely due to the wind shear coefficient varying dramaticallyies with height, time and season (Banuelos-Ruedas et al., 2010). In contrast, the comparison results of ML models areis less affected by seasonal factor. The fitting result of KNN at different season is similar except for winter. Similarly, the performance of SVM at spring (winter) is similar to summer (autumn). The R(slopes) of fitting line forfor SVM at spring, summer, autumn and winter weare 0.8 (0.66), 0.87 (0.67), 0.88 (0.8) and 0.87 (0.82), respectively.

As for RF, the fitting result in spring is slightly lower than that in other seasons. The slopes of fitting line at four seasonseasons were ranged from 0.75 to 0.85. This indicates that RF is least affected by seasons.

Overall, in terms of stability and accuracy, the RF is the best model to retrieve WS_{120}.

the error analysis of the four methods is conducted based on the WS_{10} and FV. The difference between estimated WS_{120} and observed WS_{120} is shown in Figure 6. The mean difference between PLM-observed, KNN-observed, SVM-observed, and RF-observed are -1.47, -1.00, 0.01 and 0.01 m/s, respectively. The inversion results of PLM and KNN models are underestimated relative to the RS observations. ConverselyBy contrast, the mean difference of SVM and RF models is obvious smaller than that of PLM and KNN models. Moreover, it found that the deviation of the PLM and KNN is change with the increase of WS_{10} and FV. It indicated that the stability of the PLM and KNN models need to be improved. The stability of SVM model is better than that of PLM and KNN model, but most of the SVM results are still overestimated when FV is larger than 0.4 m/s. As for RF model, the deviation is relatively stable and does not change with the increase of WS_{10} and FV. It indicated that the performance of RF is better than other three models. Overall, in terms of stability and accuracy, the RF is the best model to retrieve WS_{120}.

4.2 Characteristics of wind speed

Figure 7 shows the monthly and diurnal cycles of WS_{120} at eight stations. For all stations, the seasonal variation of wind speed is obvious. At Dongying, Penglai, Qingdao, and Lianyungang, wind speed is lager in spring (June to September) and lower in autumn (June to September). Especially at Penglai, there is an obvious low wind speed belt in July and August. By contrast, wind speed is higher in winter (December to February) at Dayang, Dongtou, and Fuqing. As for the Zhuhai stations, wind speed is relatively small throughout the year. These results indicate that the monthly variations of wind speed are significantly different in different regions. It is because of the differences in monsoon and geographical environment (Durisic et al., 2012). Also shown in Fig. 7 is the diurnal variation. At the land stations like Dongying, Penglai, Qingdao, Lianyungang, Fuqing, and Zhuhai, the wind speed is larger at daytime from 0900 to 1600 LST. The daily cycle of wind speed is mainly affected by the changes of sea-land breeze (Liu et al., 2018). The surface is heated by solar radiation at daytime, causing turbulence to intensify. Strong turbulence leads to large downward transmission of high-level wind, resulting in high wind speed during the day. After sunset, the surface radiation cools and the air layer tends to stabilize, resulting in a gradual decrease in wind speed. Similar diurnal variations in 10 m wind speed were also observed at three other stations in China (Liu et al., 2013). On the contrary, the wind speed at the Dayang and Dongtou (island stations) is higher at nighttime from 1800 to 2300 LST. This is largely due to the much higher specific heat capacity over ocean compared with over land (Li et al., 2018). The land-ocean thermal condition tends to result in a low wind speed at daytime and a high wind speed at nighttime, particularly in the absence of synoptic-scale forcing.

The histograms of WS_{120}~~wind speed~~ with corresponding Weibull distributions at eight coastal stations are plotted in Fig. 7~~8~~. The blue bar and pink lines represent occurrence probability and Weibull distributions, respectively. ~~The Weibull distribution matches well with the frequency of wind speed at all observational stations. From the probability density function, the Weibull distribution generally has long tail effect, which also indicates right skewed distribution. It makes Weibull distribution closer to reality than normal distribution (Pishgar-Komleh et al., 2015).~~ Moreover, the mean WS_{120}~~k~~ and ~~c values~~ Weibull distribution parameters for~~at~~ all eight stations are listed in Table 2. ~~The higher c indicates that the wind speed is higher, while the k indicates the wind stability (Saleh et al., 2012).~~

~~Moreover, t~~The occurrence probability~~shape~~ of WS_{120}~~the Weibull distributions~~ over these stations can be divided into two types. One type is the unimodal~~Weibull~~ distribution~~s~~ at land sites, such as Dongying, Penglai, Qingdao, Lianyungang, ~~Dayang~~Fuqing, and Zhuhai~~Dongtou~~, with a peak probability in medium wind speed (about 5~~6~~ m/s) and a low probability in high and low wind speed. The other type is~~ the~~ bimodal~~Weibull distributions~~ at island sites, such as Dayang~~Fuqing~~ and Dongtou~~Zhuhai~~ stations, with a ~~particularly~~ maximum~~high~~ ~~probability~~ peak in ~~low wind speed (about~~ 4 m/s~~)~~ and a~~ –~~ local peak at 12 m/s~~decreasing probability as the wind speed increases~~. The mean wind speed at island stations is slightly higher than that at coastal land stations. This is due to the influence~~effect~~ of ~~the~~ underlying surface roughness and ~~the~~ atmospheric stability, resulting in the difference between sea ~~breeze~~ and land breeze (Li et al., 2018; Li et al., 2020). ~~Moreover, the k and c values at all stations are listed in Table 2. The higher c indicates that the wind speed is higher, while the k indicates the wind stability (Saleh et al., 2012).~~

~~The Weibull distribution matches well with the frequency of wind speed at all stations. From the probability density function,~~

In addition, ~~it notes that~~ there is a deviation between the probability density function and the frequency of occurrence at some stations, which~~. It~~ is due to the fact that Weibull distribution generally has a long tail effect or ~~, which also indicates~~a right skewed distribution~~. It makes Weibull distribution closer to reality~~ (Pishgar-Komleh et al., 2015; Ali et al., 2018). Overall, the Weibull distribution matches ~~well~~ with the frequency of wind speed at all stations. Therefore, t~~T~~he Weibull distribution parameters can be applied for the wind energy assessment. ~~Moreover, the k and c values at all eight stations are listed in Table 3~~2~~. The higher c indicates that the wind speed is higher, while the k indicates the wind stability (Saleh et al., 2012). The wind resources at Dongying, Penglai, Qingdao, Lianyungang, Dayang, and Dongtou are richer than those at Fuqing and Zhuhai.~~

4.3 Variation of wind resource

Figure 8~~9~~ shows the diurnal variation of WS_{120}~~mean wind speed~~ and WPD at~~ all~~ eight stations. The blue and red lines are the WS_{120}~~mean wind speed~~ and WPD, respectively. At the land stations like Dongying, Penglai, Qingdao, Lianyungang, Fuqing, and Zhuhai, the WS_{120} is larger at daytime from 0900 to 1600 LST. Thise daily cycle of WS_{120} is mainly affected by the solar radiation and sea-land breeze. On the one hand, the surface is heated by solar radiation at daytime, warming the low-level air. The convecti~~ve turbulence~~on~~turbulence~~ formed by rising warm air mass ~~,~~resultsi~~ng~~ in high wind speed during the daytime. After sunset, the surface radiation cools and the air layer tends to stabilize, resulting in a gradual decrease in wind speed (Liu et al., 2018). On the other hand,~~ due to~~ the difference of specific heat capacity between~~ the~~ sea and land can form~~s,~~ the difference of thermal properties between sea and land~~ is formed~~. The difference of air pressure is obvious, which is easy to form sea land breeze (Li et al., 2020). Similar diurnal variations in 10 m wind speed were also observed at three other stations in China (Liu et al., 2013). On the contrary, the WS_{120} at the Dayang and Dongtou (island stations) is higher at nighttime from 1800 to 2300 LST. This is largely due to the much higher specific heat capacity over ocean compared with over land. The land-ocean thermal condition tends to result in a low wind speed at daytime and a high wind speed at nighttime, particularly in the absence of synoptic-scale forcing (Li et al., 2018).

~~For each station, the diurnal variation of WPD follows the same pattern of mean wind speed.~~ Overall~~On the whole~~, there are two diurnal variation patterns of wind energy at these stations~~two diurnal variation patterns can be found~~. One is for land stations,~~ such as Dongying, Penglai, Qingdao, Lianyungang, and Fuqing.~~ t~~T~~he hourly mean WPD is larger at daytime from 0900 to 1600 LST with a peak at 1400 LST. ~~This is mainly due to the influence of the sea-land breeze (Liu et al., 2018).~~ The other is for island stations,~~ such as Dayang, and Dongtou.~~ t~~T~~he hourly mean WPD of these stations remains at a high level at all day and is relatively large at nighttime from 1800 to 2300 LST. The urban electricity demand usually reaches peaks at around noon in the daytime and in the evening (Hong et al., 2012). This means that the wind energy at the land and island stations can support the power demand during the noon and midnight, respectively. When the demand and the supply achieve a balance, wind energy will be used more effectively. In addition, it is worth noting that the mean wind speed and WPD at island stations are generally higher than that at land stations, which may be due to the difference in specific heats between land and sea. Li et al. (2018) also pointed out that the offshore stations offer more wind energy than onshore stations.

Figure 910 shows the monthly variation of WS_{120}mean wind speed and WPD at eight stations. For all sitesAt northern Chinanorth of Chinathe coastline, such as Dongying, Penglai, Qingdao, and Lianyungang, wind speedthe seasonal distribution of WS_{120} is lagestlargestr at springMayin spring (June to September) and winter, and is lowestr in summer and autumnSeptemberautumn (June to September). This is due to the influence of East Asia Monsoon and Mongolian cyclones (Yu et al., 2016). Liu et al. (2019) The large-scale synoptic systems in China have a relatively high occurrence frequency during the cold season (spring and winter), which result in the higher wind speed than warm season (summer and autumn) (Liu et al., (2019)). In addition, at north China, such as Dongying, Penglai, Qingdao, and Lianyungang, Especially at Penglai, there is an obvious low wind speed belt in July and August. By contrast, wind speed is higher in winter (December to February) at Dayang, Dongtou, and Fuqing. As for the Zhuhai stations, wind speed is relatively small throughout the year. These results indicate that the monthly variations of wind speed are significantly different in different regions. It is because of the differences in monsoon and geographical environment (Durisic et al., 2012). TSimilar to diurnal variation, the monthly variation of WPD in eight stations exhibits the same tend as that of mean wind speed. However, the monthly variation of WPD varies by station. tThe monthly WPD at these stationsof Dongying, Penglai, Qingdao, and Lianyungang is relatively high for the period from March to May, as compared to the much lower values from August to October. This result indicates that the wind source of coastline of Shandong province is more adequate in spring and winter monthsseason. By contrast, at south of China, the WS_{120}wind speed is higher in winter (December to February) at Dayang, Dongtou, and Fuqing. As for the Zhuhai stations, wind speed is relatively small throughout the year.

In additionover Dayang and Dongtou, the monthly WPD at Dayang and Dongtou is maximums in December, while is low in March and April. Moreoverand, most of the monthly WPD at Dayang and Dongtou are larger than 200 W/m^2. This may could be likely owing to the fact thatbe due to these two stations are set up on the island, and the wind energy mainly depends on the sea breeze circulations. As for Fuqing and Zhuhai, the WPD maintain a very low value for every month and remain almost constant.These results indicate that the monthly variations of wind speed are significantly different in different regions. It is because of the differences in monsoon and geographical environment (Durisic et al., 2012).

Figure 101 shows the spatial distribution of seasonal WS_{120}mean wind speed and WPD in the coastal regions of China. The shading colors in the background show the corresponding results calculated from the ERA5 data, which is used as reference. Overalln the whole, the spatio-temporal variations of wind speed and ~~wind resource~~WPD calculated from the RWP observations have good consistency with that of ERA5 data. The maximum mean WS_{120}~~wind speed~~ of 6.79 m/s occurs at Dayang in summer and the minimum mean WS_{120}~~wind speed~~ of 4.52 m/s occurs at Zhuhai in autumn. Moreover, the WS_{120}~~mean wind speed~~ at Dongying, Penglai, Qingdao, Lianyungang, Dayang, and Dongtou is relatively higher than that at Fuqing and Zhuhai for all seasons. It indicates that the wind resources may be richer in the coastal region of northern China. ~~As for the seasonal variation, the mean wind speed at Dongying, Penglai, Qingdao, Lianyungang, and Dayang is the larger in spring and summer than other seasons. For other stations, the largest mean wind speed occurs in winter or autumn.~~ According to National Renewable Energy Laboratory standard (Jamil et al., 1995), the WPD of Qingdao, Dayang, and Dongtou are higher than 200 W/m^2 in most seasons, and these three stations could be classified as wind power class II stations. Except for island stations at Dayang and Dongtou, the WPD at Dongying, Penglai, Qingdao, and Lianyungang are much greater than those at Fuqing and Zhuhai, irrespective of seasons. Those results indicated that the wind resources in the Bohai Sea and the Yellow Sea coast are more abundant than those in the South China Sea coast. Furthermore, for the coastal region of Bohai Sea and the Yellow Sea, the wind energy resources are the most abundant in spring while for the East China Sea and the South China Sea coast, the wind energy resources are relatively abundant in summer.

5. Summary and conclusions

This study used the ML algorithms to evaluate the wind energy resource at eight coastal stations based on the wind speed profile and surface meteorological data from May 2018 to August 2020. Moreover, the accuracy of PLM, KNN, SVM and RF ~~models~~ was compared based on the comparison~~correlation and difference~~ between observed WS_{120} and estimated WS_{120}. Finally, the wind energy resource at eight coastal stations was evaluated based on the WS_{120} from RF~~ model~~.

For the four WS_{120} inversion method, the accuracy of three ML models is better than that of~~e~~ PLM. This~~It~~ is probably due to the PLM only dependings on the constant α to establish~~ the mapping~~ relationship between surface wind speed and WS_{120}. In fact, ~~the~~ α is not constant and changes with height, time and meteorological conditions. It~~This~~ results in a relatively low accuracy of the PLM method. In contrast, the ML models consider the influence of environmental parameters to improve accuracy, such as FV and Char etc. Moreover, it can be noted that there are also differences in performance between different ML models. The results indicate that the RF ~~model~~ is the best model to retrieve WS_{120}, followed by SVM~~ model~~; last are KNN~~ model~~. This is caused by different decision strategies of the ML models. The variable importance analysis indicated that the model which can comprehensively consider the influence of most variables has the best performance.

The monthly variation of wind resources varies on the coast of China. The wind resources along the Bohai Sea coast have a~~two~~ peak~~s~~ approximately in May~~ and October~~. By contrast, the wind resources along the Yellow Sea coast keeps relatively stable without pronounced peak. As for the coastal regions of East China Sea and the South China Sea, the wind resources increase from January, reach the maximum in June or July, and then decrease until December. In terms of the diurnal variation of wind resources, the WPD over land station has a peak at daytime from 0900 to 1600 LST, while the WPD over island station exhibits peak value at nighttime from 1800 to 2300 LST. This means that the wind energy at the land and island stations can support the power demand during the noon and midnight, respectively. When the demand and the supply achieve a balance, wind energy will be used more effectively. As for the spatial distribution of wind resource, the Bohai Sea and Yellow Sea coast have more abundant wind resources than the East China Sea and the South China Sea. The seasonal variations of wind resources vary on the coast of China. The coast of the Bohai Sea and Yellow Sea has the richest wind resources in spring or autumn, while the coast of the East China Sea and the South China Sea has the richest wind resources in summer.

Our work comprehensively assesses the wind energy resources on the coast of China using the state-of-the-art ~~machine learning~~ML algorithm, which provides invaluable information for the development of wind energy industry in the coastal regions of China in the future. However, wind energy assessment is only one part of the efficient utilization of wind energy resources. The cost of wind turbines, topography conditions, environment harm, and other factors also need more attention, which deserves further investigation in the future.

Data Availability

The ~~radar wind profiler~~RWP data used in this paper can be provided for non-commercial research purposes upon motivated request (Jianping Guo, Email: jpguocams@gmail.com). The anemometer~~10 m wind speed~~ data can be downloaded in http://www.nmic.cn/data/cdcdetail/dataCode/A.0012.0001.html, last access: 15 November 2022. The RS data can be downloaded in http://www.nmic.cn/data/cdcdetail/dataCode/B.0011.0001C.html, last access: 15 November 2022. The ERA5~~his~~ data can be downloaded in https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

Acknowledgments

This work was ~~jointly ~~supported by the National Natural Science Foundation of China (under grants 42001291~~ and U2142209~~) and the Fundamental Research Funds for the Central Universities (under grants 2042022kf1003)~~ and by the Project through the China Postdoctoral Science Foundation (under grant 2020M682485)~~.

Author Contributions

The study was completed with cooperation between all authors. J~~ianping ~~G~~uo~~ and ~~Liu ~~B~~oming~~L designed the research framework; ~~Liu Boming~~BL and J~~ianping ~~G~~uo~~ conducted the experiment and wrote the paper; X~~in ~~M~~a~~, H~~ui ~~L~~i~~, S~~hikuan ~~J~~in~~, Y~~ingying ~~M~~a~~, and W~~ei ~~G~~ong~~ analyzed the experimental results and helped touch on the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Ali, S., Lee, S. M., Jang, C. M.: Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island – ~~–~~Incheon, South Korea. Renew Energ., 123:652–~~–~~663, https://doi.org/10.1016/j.renene.2018.02.087, 2018.

Abbes, M., and Belhadj, J.: Wind resource estimation and wind park design in El-Kef region, Tunisia. Energy, 40(1), 348–~~–~~357, https://doi.org/10.1016/j.energy.2012.01.061, 2012.

Akpinar E. K., Akpinar S.: An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. Energy Convers Manage, 46(11):1848–~~–~~67, https://doi.org/10.1016/j.enconman.2004.08.012, 2005.

Altman, N. S.: An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat., 46:~~,~~ 175–185, https://doi.org/10.2307/2685209, 1992.

~~Aukitino, T., Khan, M. G. M., Ahmed, M. R.: Wind energy resource assessment for Kiribati with a comparison of different methods of determining Weibull parameters. Energ Convers Manage, 151:641-60, https://doi.org/10.1016/j.enconman.2017.09.027, 2017.~~

Allabakash, S., Lim, S., Yasodha, P., Kim, H., and Lee, G.: Intermittent clutter suppression method based on adaptive harmonic wavelet transform for L-band radar wind profiler. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 8546–8556, 2019.

Band, S. S., Bateni, S. M., Almazroui, M., Sajjadi, S., Chau, K. W., and Mosavi, A.: Evaluating the potential of offshore wind energy in the Gulf of Oman using the MENA-CORDEX wind speed data simulations. Engineering Applications of Computational Fluid Mechanics, 15(1):, 613–626, https://doi.org/10.1080/19942060.2021.1893225, 2021.

Breiman, L.: Random forests, in: Machine Learning, 45:, 5–32, 2001.

Banuelos-Ruedas, F., Angeles-Camacho, C., Rios-Marcuello, S.: Analysis and validation of the methodology used in the extrapolation of wind speed data at different heights. Renew Sustain Energy Rev., 14(8):2383-91, https://doi.org/10.1016/j.rser.2010.05.001, 2010.

Council GWE. Global wind statistics 2020. Global Wind Report. 2021.

Chang, T. P.: Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. Appl. Energ., 88(1): 272–82, https://doi.org/10.1016/j.apenergy.2010.06.018, 2011.

Coleman, T. A., Knupp K. R., and Pangle P. T.: The effects of heterogeneous surface roughness on boundary-layer kinematics and wind shear. Electronic J. Severe Storms Meteor., 16 (3), 1–29, 2021.

Coomans, D. and Massart, D. L.: Alternative k-nearest neighbor rules in supervised pattern recognition: part 1. k-Nearest neighbour classifcation by using alternative voting rules, Anal. Chim. Acta, 136, 15–27, https://doi.org/10.1016/s0003-2670(01)95359-0, 1982.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20, 273–297, 1995.

Costoya, X., DeCastro, M., Carvalho, D., Feng, Z., and& Gómez-Gesteira, M.: Climate change impacts on the future offshore wind energy resource in China. Renewable Energy, 175, 731–747, 2021.

Durisic Z., Mikulovic J.: Assessment of the wind energy resource in the South Banat region, Serbia. Renew Sust Energ Rev., 16(5):3014–30-23, https://doi.org/10.1016/j.rser.2012.02.026, 2012.

~~Demolli, H., Dokuz, A. S., Ecemis, A., and Gokcek, M.: Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Conversion and Management, 198, 111823, 2019.~~

~~EWEA: Wind Energy-the Facts: A Guide to the Technology, Economics and Future of Wind Power; Routledge: London, UK, 2009.~~

~~Eichhorn M., Scheftelowitz M., Reichmuth M., Lorenz C., Louca K., Schiffler A.: Spatial Distribution of Wind Turbines, Photovoltaic Field Systems, Bioenergy, and River Hydro Power Plants in Germany. Data., 4(1), https://doi.org/10.3390/data4010029, 2019.~~

Fagbenle R. O., Katende J, Ajayi O. O., Okeniyi J. O.: Assessment of wind energy potential of two sites in North-East, Nigeria. Renew Energ., 36(4):1277–12-83, https://doi.org/10.1016/j.renene.2010.10.003, 2011.

Guo, J., Chen X., Su T., Liu L., Zheng Y., Chen D., Li J., Xu H., Lv Y., He B., Li Y., Hu X., Ding A., and Zhai P.: The climatology of lower tropospheric temperature inversions in China from radiosonde measurements: roles of black carbon, local meteorology, and large-scale subsidence. Journal of Climate, 33 (21): 9327–9350, doi: 10.1175/JCLI-D-19-0278.1, 2020.

Guo, J., Liu, B., Gong, W., Shi, L., Zhang, Y., Ma, Y., Xu, X.: First comparison of wind observations from ESA's satellite mission Aeolus and ground-based radar wind profiler network of China. Atmos. Chem. Phys. 21 (4), 2945–2958, https://doi.org/10.5194/acp-21-2945-2021, 2021a.

Guo, J., Zhang, J., Yang, K., Liao, H., Zhang, S., Huang, K., Lv, Y., Shao, J., Yu, T., Tong, B., Li, J., Su, T., Yim, S. H. L., Stoffelen, A., Zhai, P., and Xu, X.: Investigation of near-global daytime boundary layer height using high-resolution radiosondes: First results and comparison with ERA-5, MERRA-2, JRA-55, and NCEP-2 reanalyses, Atmos. Chem. Phys., 21, 17079–17097, https://doi.org/10.5194/acp-21-17079-2021, 2021b.

Gualtieri, G.: Reliability of era5 reanalysis data for wind resource assessment: a comparison against tall towers. Energies, 14(14), 4169, 2021.

Hoffmann L., Gunther G., Li D., Stein O., Wu X., Griessbach S.: From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. Atmos. Chem. Phys., 19(5):3097–3-124, https://doi.org/10.5194/acp-19-3097-2019, 2019.

Hersbach H., Bell B., Berrisford P., Hirahara S., Horanyi A., Munoz-Sabater J.: The ERA5 global reanalysis. Q. J. Roy. Meteor. Soc., 146(730):1999–-2049, 2020.

Hellmann G. Über die Bewegung der Luft in den untersten Schichten der Atmosphare: Kgl. Akademie der Wissenschaften. Reimer 1914.

Hong, L. X., Moller, B.: Feasibility study of China's offshore wind target by 2020. Energy., 48(1):268–-77, https://doi.org/10.1016/j.energy.2012.03.016, 2012.

Jamil, M., Parsa, S., Majidi, M.: Wind power statistics and an evaluation of wind energy density. Renewable Energy, 6(5):623–628, https://doi.org/10.1016/0960-1481(95)00041-h, 1995.

Jiang, D., Zhuang, D. F., Huang, Y. H., Wang, J. H., Fu, J. Y.: Evaluating the spatio-temporal variation of China's offshore wind resources based on remotely sensed wind field data. Renew Sust Energ Rev., 24:142–14-8, https://doi.org/10.1016/j.rser.2013.03.058, 2013.

Khatib, H.: IEA World Energy Outlook 2011-A comment. Energ Policy., 48:737–7-43, 2012.

Khosravi, A., Machado, L., and Nunes, R. O.: Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. Applied Energy, 224, 550–-566, 2018.

Leung, D. Y. C., and Yang, Y.: Wind energy development and its environmental impact: A review. Renew. Sust. Energ. Rev., 16(1):1031–103-9, https://doi.org/10.1016/j.rser.2011.09.024, 2012.

Li, J. L., Yu, X.: Onshore and offshore wind energy potential assessment near Lake Erie shoreline: A spatial and temporal analysis. Energy., 147: 1092–-1107, https://doi.org/10.1016/j.energy.2018.01.118, 2018.

Li, Y., Huang, X., Tee, K. F., Li, Q., and Wu, X. P.: Comparative study of onshore and offshore wind characteristics and wind energy potentials: A case study for southeast coastal region of China. Sustainable Energy Technologies and Assessments, 39:, 100711, 2020.

Li, H., Liu, B., Ma, X., Jin, S., Ma, Y., Zhao, Y., Gong, W.: Evaluation of retrieval methods for planetary boundary layer height based on radiosonde data. Atmos. Meas. Tech. 14:; 5977–5986. https://doi.org/10.5194/amt-14-5977-2021, 2021.

Liu, J., Gao, C. Y., Ren, J., Gao, Z., Liang, H., and Wang, L.: Wind resource potential assessment using a long term tower measurement approach: A case study of Beijing in China. Journal of cleaner production, 174:; 917–926, 2018.

Liu, B., Ma, Y., Guo, J., Gong, W., Zhang, Y., Mao, F., Li, J., Guo, X., and Shi, Y.: Boundary layer heights as derived from ground-based Radar wind profiler in Beijing. IEEE Trans. Geosci. Remote Sens., 57 (10):; 8095–8104. doi: 10.1109/TGRS.2019.2918301, 2019.

Liu, B., Guo, J., Gong, W., Shi, L., Zhang, Y., and Ma, Y.: Characteristics and performance of wind profiles as observed by the radar wind profiler network of China. Atmos. Meas. Tech., 13:; 4589–4600, https://doi.org/10.5194/amt-13-4589-2020, 2020.

Liu, B., Ma, X., Ma, Y., Li, H., Jin, S., Fan, R., and Gong, W.: The relationship between atmospheric boundary layer and temperature inversion layer and their aerosol capture capabilities. Atmos. Res., 271:; 106121, https://doi.org/10.1016/j.atmosres.2022.106121, 2022.

Liu, R., Liu, S., Yang, X., Lu, H., Pan, X., Xu, Z.: Wind dynamics over a highly heterogeneous oasis area: An experimental and numerical study. Journal of Geophysical Research: Atmospheres, 123:; 8418– 8440. https://doi.org/10.1029/2018JD028397, 2018.

Liu, Y., Xiao, L. Y., Wang, H. F., Dai, S. T., Qi, Z. P.: Analysis on the hourly spatiotemporal complementarities between China's solar and wind energy resources spreading in a wide area. Sci. China. Technol. Sc. 56:; 683–692, https://doi.org/10.1007/s11431-012-5105-1, 2013.

Liu J. K., Gao C. Y., Ren J. Z., Gao Z. Q., Liang H. W., Wang L. L.: Wind resource potential assessment using a long term tower measurement approach: A case study of Beijing in China. J. Clean Prod., 174:917-26, 2018.

Liu, F., Sun, F., Liu, W., Wang, T., Wang, H., Wang, X., and& Lim, W. H.: On wind speed pattern and energy potential in China. Applied Energy, 236:, 867—876, 2019.

Laurila, T. K., Sinclair, V. A., and& Gregow, H.: Climatology, variability, and trends in near-surface wind speeds over the North Atlantic and Europe during 1979–2018 based on ERA5. International Journal of Climatology, 41(4), 2253—2278, 2021.

May, P. T., and Strauch, R. G.: Reducing the effect of ground clutter on wind profiler velocity measurements. Journal of Atmospheric and Oceanic Technology, 15(2):, 579—586, 1998.

Maronga, B., and Reuder, J.: On the formulation and universality of Monin–Obukhov similarity functions for mean gradients and standard deviations in the unstable surface layer: Results from surface-layer-resolving large-eddy simulations. Journal of the Atmospheric Sciences, 74(4):, 989—1010, 2017.

Mo, H. M., Hong, H. P., and Fan, F.: Estimating the extreme wind speed for regions in China using surface wind observations and reanalysis data. Journal of Wind Engineering and Industrial Aerodynamics, 143:, 19—33, 2015.

Ming Z, Kun Z, and Jun D.: Overall review of China's wind power industry: Status quo, existing problems and perspective for future development. Renew. Sust. Energ. Rev., 24:379–3-86, 2013.

Mohandes, M. A., and Rehman, S.: Wind speed extrapolation using machine learning methods and LiDAR measurements. IEEE Access, 6, 77634-77642, 2018.

Magazzino, C., Mele, M., & Schneider, N.: A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO2 emissions. Renewable Energy, 167:, 99—115, 2021.

Ma, Y., Zhu, Y., Liu, B., Li, H., Jin, S., Zhang, Y., Fan, R., and Gong, W.: Estimation of the vertical distribution of particle matter (PM2.5) concentration and its transport flux from lidar measurements based on machine learning algorithms, Atmos. Chem. Phys., 21:, 17003–17016, https://doi.org/10.5194/acp-21-17003-2021, 2021.

Oh, K. Y., Kim, J. Y., Lee, J. K., Ryu, M. S., and Lee, J. S.: An assessment of wind energy potential at the demonstration offshore wind farm in Korea. Energy., 46(1):555–5-63, 2012.

Pei, Z. P.,; Han, G.,; Ma, X.,; Shi, T. Q.,; Gong, W.: A Method for Estimating the Background Column Concentration of CO2 Using the Lagrangian Approach. IEEE Transactions on Geoscience and Remote Sensing, 60, doi:10.1109/TGRS.2022.3176134, 2022.

Patel, M. R.: Wind and solar power systems: design, analysis, and operation. CRC press; 2005.

Pishgar-Komleh S. H., Keyhani A., Sefeedpari P.: Wind speed and power density analysis based on Weibull and Rayleigh distributions a case study: Firouzkooh county of Iran. Renew Sust Energ Rev., 42: 313-22, https://doi.org/10.1016/j.rser.2014.10.028, 2015.

Rocha P. A. C., de Sousa R. C., de Andrade C. F., da Silva M. E. V.: Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. Appl. Energ., 89(1):395–-400, 2012.

Shu Z. R., Li Q. S., He Y. C., Chan P. W.: Observations of offshore wind characteristics by Doppler-LiDAR for wind energy applications. Appl. Energ., 169:150–-63, 2016.

Saleh H, Aly A. A., Abdel-Hady S.: Assessment of different methods used to estimate Weibull distribution parameters for wind speed in Zafarana wind farm, Suez Gulf, Egypt. Energy. 44(1):710-–719, https://doi.org/10.1016/j.energy.2012.05.021, 2012.

Shakun, J.D., Clark, P.U., He, F., Marcott, S.A., Mix, A.C., Liu, Z., Otto-Bliesner, B., Schmittner, A. and Bard, E.: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. Nature, 484(7392): 49–54, 2012.

~~Shoaib, M., Siddiqui, I., Rehman, S., Khan, S., and& Alhems, L. M.: Assessment of wind energy potential using wind energy conversion system. Journal of cleaner production, 216, 346-360, 2019.~~

Shi, T., Han, G., Ma, X., Gong, W., Chen, W., Liu, J., Bu, L.: Quantifying CO2 uptakes over oceans using LIDAR: a tentative experiment in Bohai bay. Geophys. Res. Lett. 48 (9), 2020GL091160, 2021.

Shoaib, M., Siddiqui, I., Rehman, S., Khan, S., and Alhems, L. M.: Assessment of wind energy potential using wind energy conversion system. Journal of cleaner production, 216:, 346–360, 2019.

Solanki, R., Guo J., Lv Y., Zhang J., Wu J., Tong B., and Li J.: Elucidating the atmospheric boundary layer turbulence by combining UHF Radar wind profiler and radiosonde measurements over urban area of Beijing. Urban Climate, 43:, 101151, doi: 10.1016/j.uclim.2022.101151, 2022.

Stull, R. B.: An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers, Dordrecht, 1988.

Su, X., Wang, L., Gui, X., Yang, L., Li, L., Zhang, M., and Wang, L.: Retrieval of total and fine mode aerosol optical depth by an improved MODIS Dark Target algorithm. Environment International, 166:, 107343, 2022a.

Su, X., Wei, Y., Wang, L., Zhang, M., Jiang, D., and Feng, L.: Accuracy, stability, and continuity of AVHRR, SeaWiFS, MODIS, and VIIRS deep blue long-term land aerosol retrieval in Asia. Science of The Total Environment, 832:, 155048, 2022b.

Tieleman, H. W.: Wind characteristics in the surface layer over heterogeneous terrain. Journal of Wind Engineering and Industrial Aerodynamics 41(1): 329-340, 1992.Wen Y, Kamranzad B, Lin PZ. Assessment of long-term offshore wind energy potential in the south and southeast coasts of China based on a 55-year dataset. Energy. 224, https://doi.org/10.1016/j.energy.2021.120225, 2021.

Veers, P., Dykes, K., Lantz, E., Barth, S., Bottasso, C. L., Carlson, O., and... & Wiser, R.: Grand challenges in the science of wind energy. Science, 366(6464):, eaau2027, 2019.

Yuan, J.: Wind energy in China: Estimating the potential. Nature Energy, 1(7):, 1-2, 2016.

~~Yu J. X., Fu Y. Q., Yu Y., Wu S. B., Wu Y. D., You M. J.: Assessment of Offshore Wind Characteristics and Wind Energy Potential in Bohai Bay, China. Energies, 12(15), 2019.~~

Yu, L., Zhong, S., Bian, X., and~~&~~ Heilman, W. E.: Climatology and trend of wind power resources in China and its surrounding regions: A revisit using Climate Forecast System Reanalysis data. International Journal of Climatology, 36(5):~~,~~ 2173-2188, 2016.

Zheng C. W., Zhuang H., Li X., Li X. Q.: Wind energy and wave energy resources assessment in the East China Sea and South China Sea. Sci. China Technol. Sc., 55(1):163–1~~-~~73, 2012.

Zhang, J., Zhang, M., Li, Y., Qin, J., Wei, K., and Song, L.: Analysis of wind characteristics and wind energy potential in complex mountainous region in southwest China. Journal of Cleaner Production, 274:~~,~~ 123036, 2020.

Tables:

Table 1 Detailed information of the radar wind profiler observational stations.

Station Name	Station ID	Longitude (°E)	Latitude (°N)	Altitude (~~k~~m)	Surface types
Dongying	54736	118.67	37.44	11.1	Land
Penglai	54752	120.76	37.79	60.7	Land
Qingdao	54857	120.23	36.33	12	Land
Lianyungang	58044	119.24	34.54	4	Land
Dayang	58474	122.04	30.64	49	Island
Dongtou	5876~~S~~0	121.15	27.83	71	Island
Fuqing	58942	119.39	25.72	51.7	Land
Zhuhai	59488	113.2	22.07	30	Land

Table 2~~3~~ Statistics for the Weibull distribution of WS_{120} at~~on~~ the eight stations from 1 May 2018 to 31 August 2020.

Station	WS_{120} (m/s)	Standard deviation (m/s)	Weibull Shape factor k	Weibull Scale factor c (m/s)
Dongying	5.54	1.77	3.46	6.16
Penglai	5.27	2.39	2.35	5.95
Qingdao	5.86	2.45	2.58	6.59
Lianyungang	5.81	1.75	3.68	6.43
Dayang	6.64	2.99	2.38	7.49
Dongtou	5.89	2.66	2.37	6.65
Fuqing	5.39	2.44	2.37	6.08
Zhuhai	4.68	1.78	2.87	5.25

Figures:

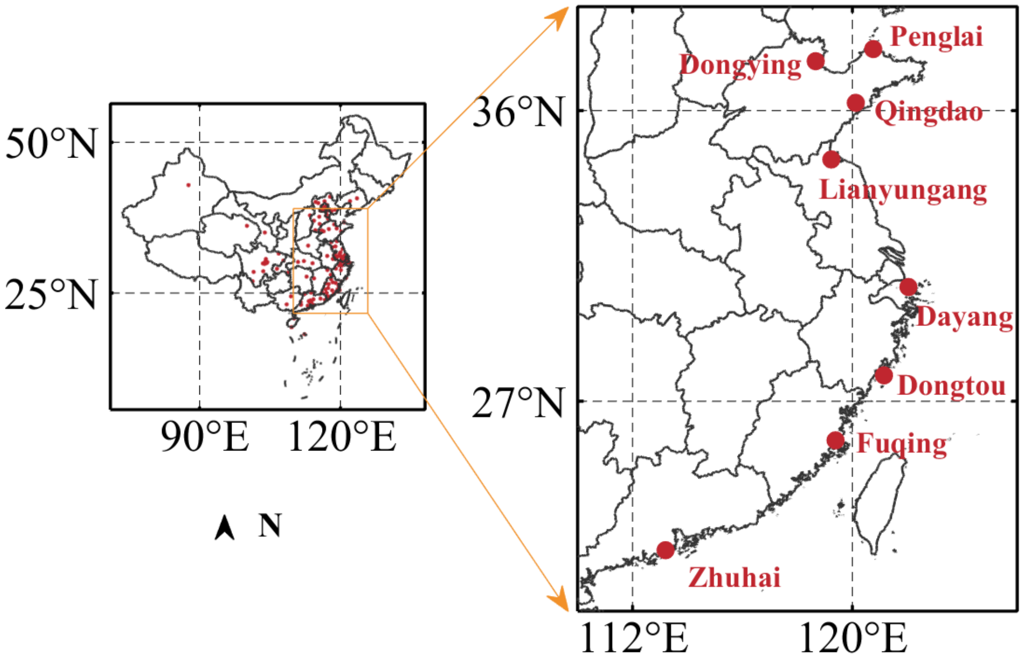

Figure 1. Geographical ~~location~~ distribution of the eight radar wind profiler observational stations (red dots) in the coast of East China.

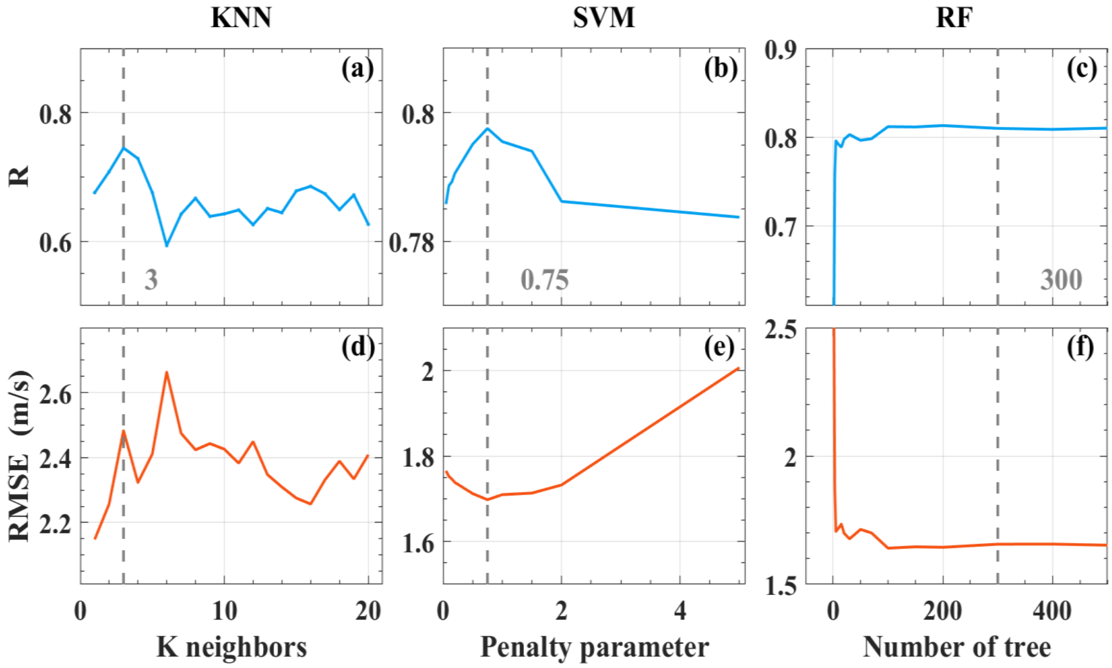

Figure 23. The parameter tuning process for (a, d) KNN, (b, e) SVM and (c, f) RF models. The blue and red lines represent the variation of R and RMSE, respectively. The gray dotted lines and texts indicate the optimal parameters for of their corresponding models.

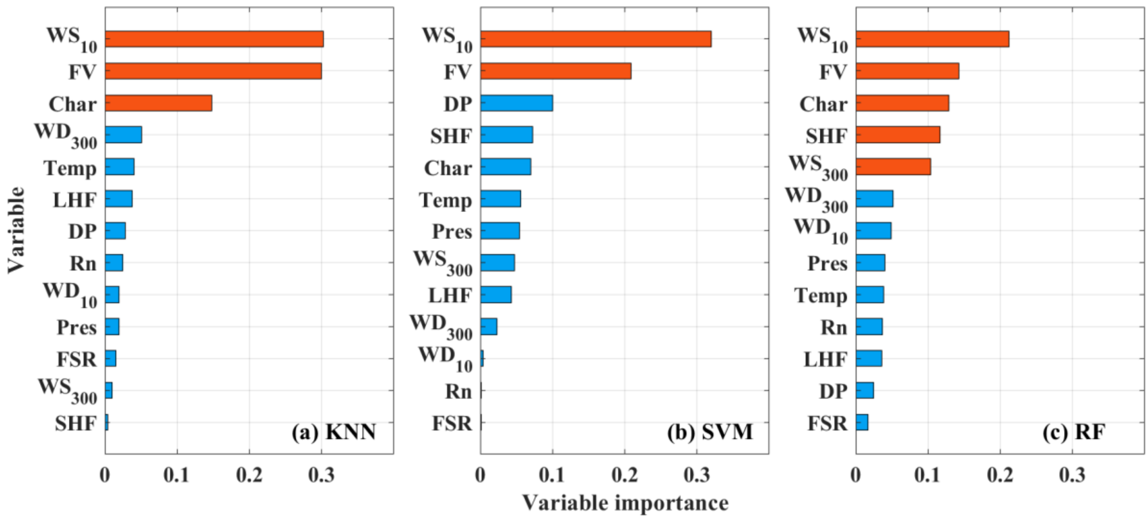

Figure 34. Importance analysis of input variables for three models: (a) KNN, (b) SVM, and (c) RF– ~~models~~.

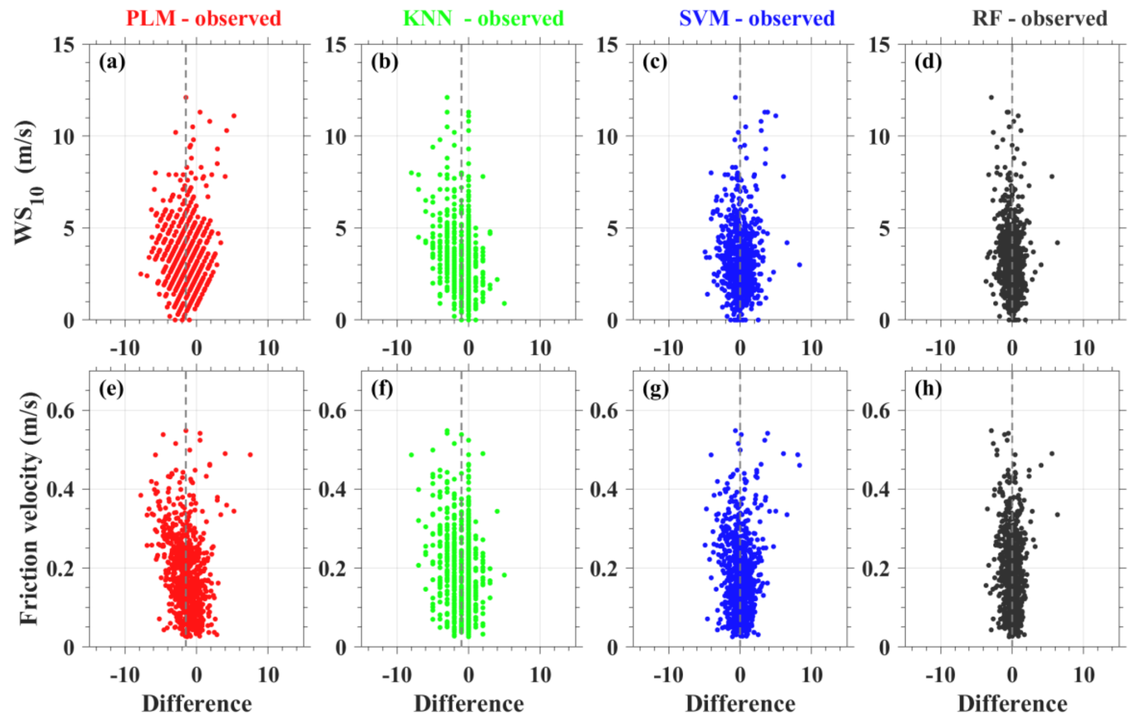

Figure 4. Scatter plots showing the difference of observed WS_{120} and estimated WS_{120} as a function of WS_{10} (a-d) and friction velocity (FV, e-h). The red, green, blue and black points represent the difference for PLM-observed, KNN-observed, SVM-observed and RF-observed, respectively. The gray line represents the mean difference.

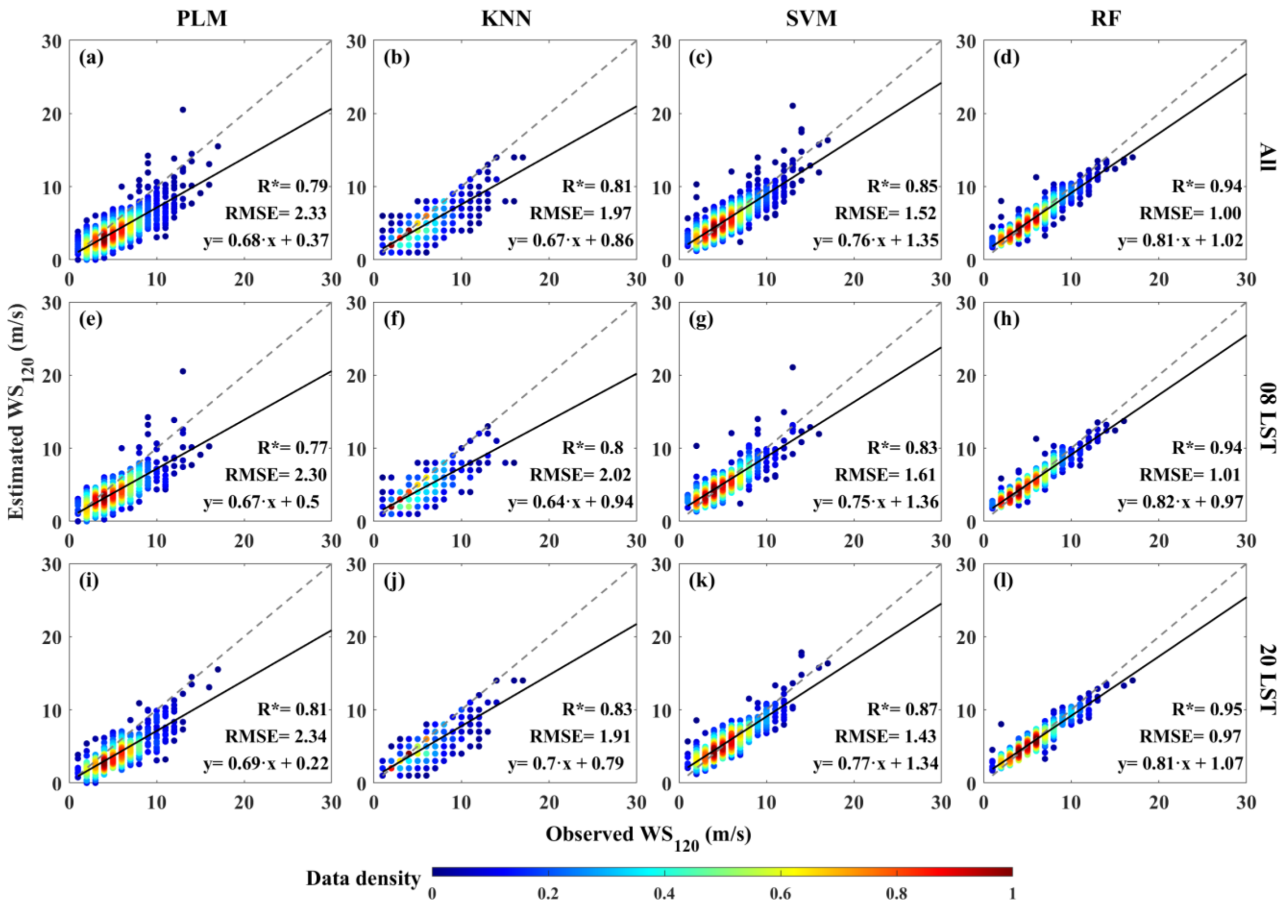

Figure 54~~54~~. Comparison~~Correlation coefficient~~s between observed WS_{120} and estimated WS_{120} based on the (a, e, i) PLM, (b, f, j) KNN, (c, g, k) SVM and (d, h, l) RF models under different time. The gray and black line is the reference and regression line, respectively. The ~~colorbar~~color bar represents the data density. The asterisk indicates that the correlation coefficient (R) has passed the t-test at a confidence level of 95%~~passed the statistical significance difference test (P < 0.05)~~.

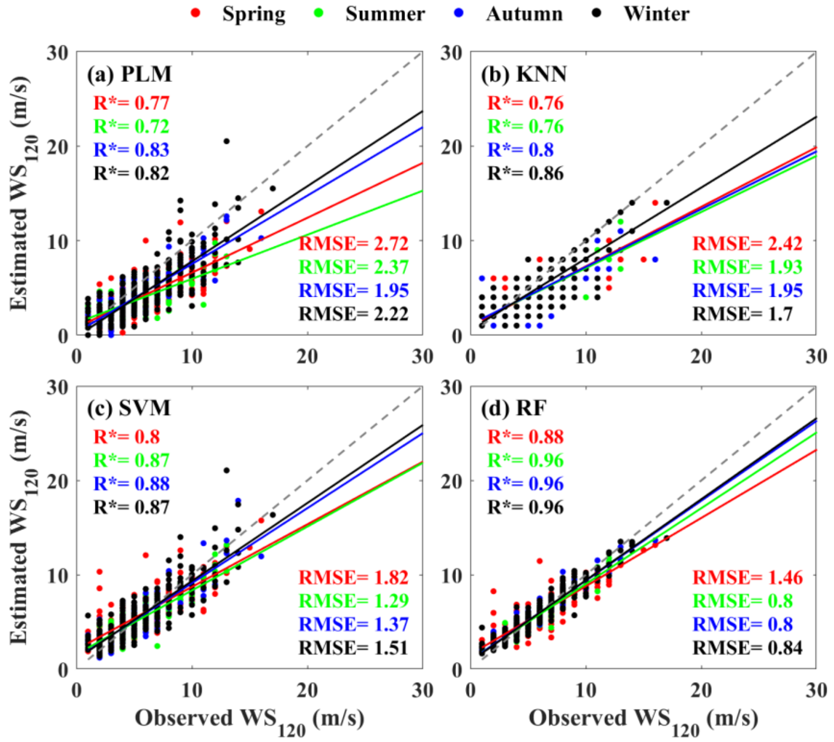

Figure 6. Comparisonorrelation coefficients between observed WS_{120} and estimated WS_{120} based on the (a) PLM, (b) KNN, (c) SVM and (d) RF models under different season. The– red, green, blue and black represent spring, summer, autumn and wintergray and black line is the reference and regression line, respectively. The colorbar represents the data density. The asterisk indicates that the correlation coefficient (R) has passed the statistical significance difference test ($P < 0.05$).t-test at a confidence level of 95%.

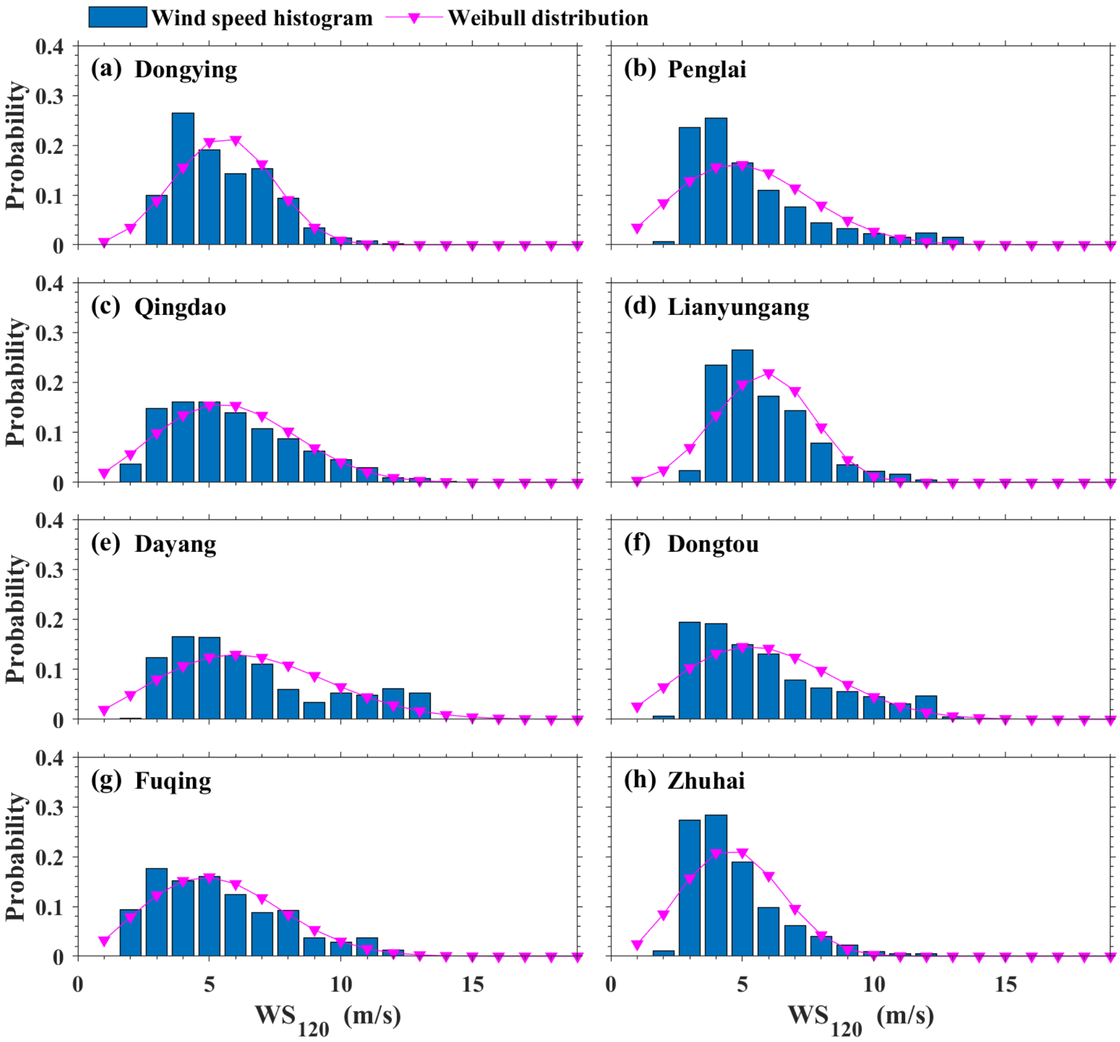

Figure 78. Probability distribution and Weibull distribution of WS_{120} at the eight stations from 1 May 2018 to 31 August 2020. The blue bar and pink lines represent occurrence probability and Weibull distributions, respectively.

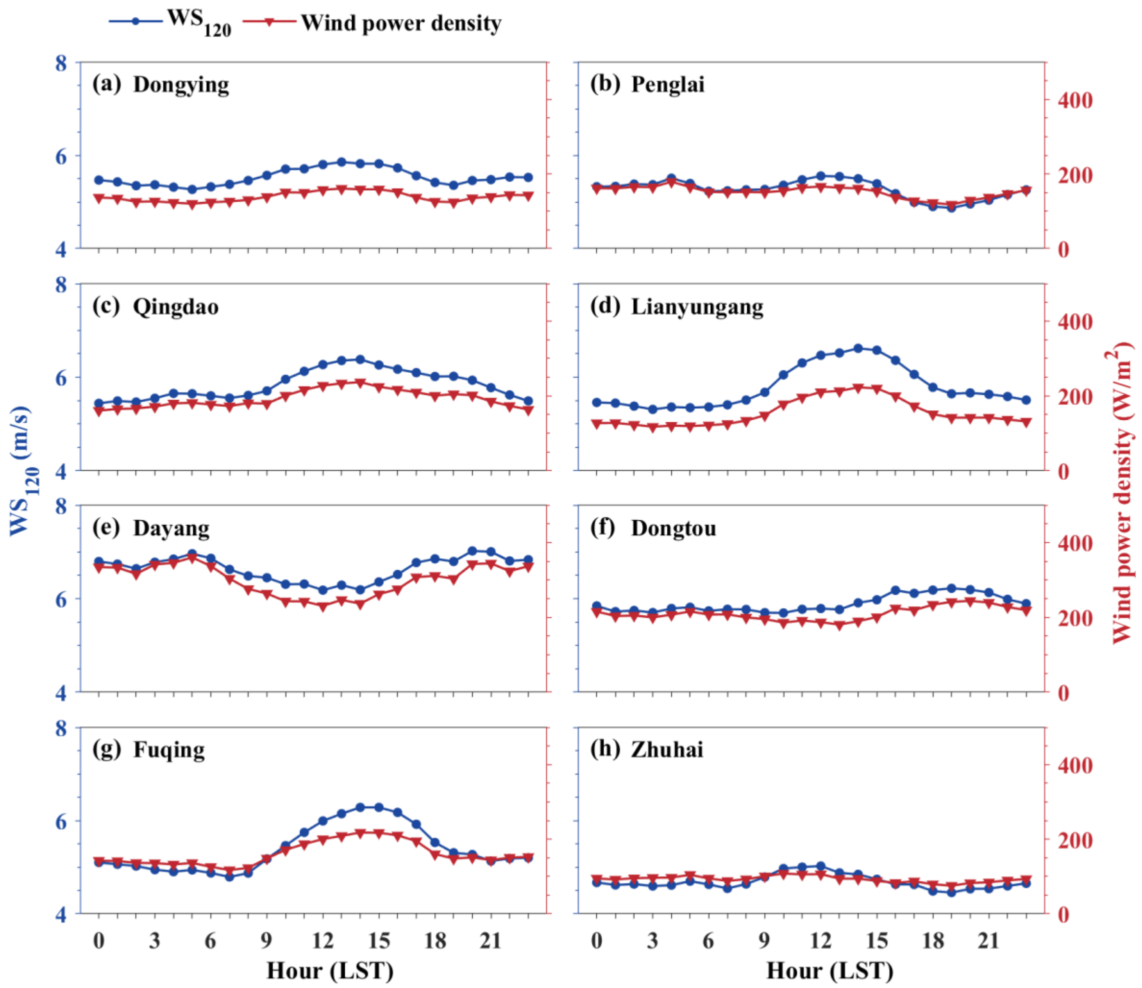

Figure 8~~9~~. Diurnal variation of the WS_{120} and wind power density for the eight RWP stations as shown in Figure 1~~ shown in Fig. 1~~. The blue and red lines denote~~are~~ the mean wind speed and wind power density, respectively.

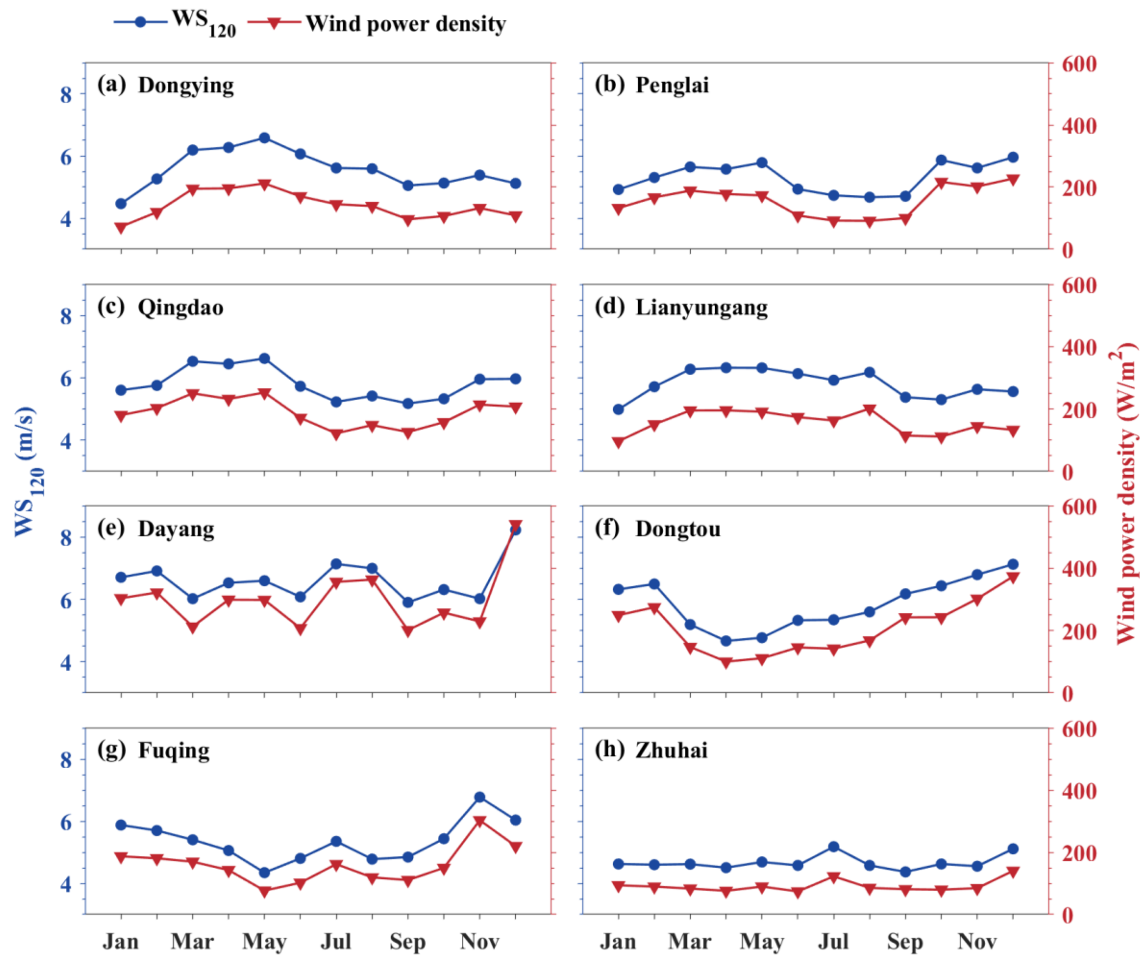

Fig. 910. Similar towith Fig. 89, but for the monthly variation.

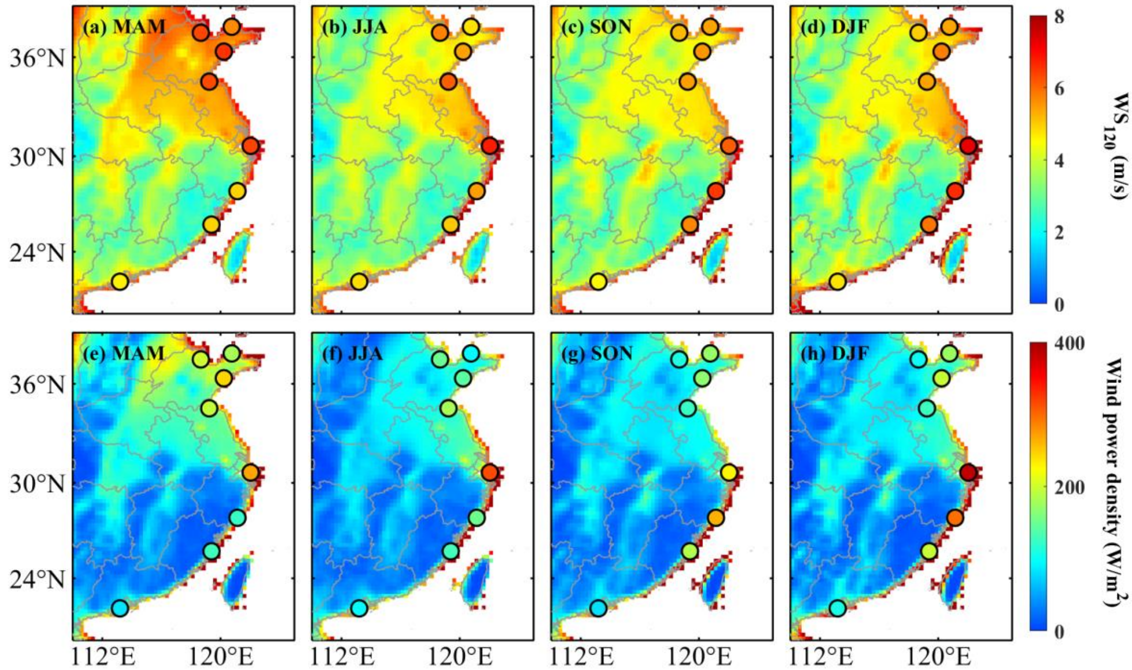

Fig. 101. Spatial distribution of the seasonal mean wind speed and wind power density at 10200 m AGL along the coastline of China. The circles represent the WS_{120} observations directly from the eight RWP stations. The shading colors in the background show the corresponding results calculated from the ERA5 reanalysis reanalysisdata.