Estimating hub height wind speed~~Assessment of the wind energy resource on the coast of China~~ based on machine learning algorithms: Implications for ~~the~~ wind energy assessment ~~s~~

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 energy is one of the most essential clean and renewable energy sources in today's world.~~ Accurate estimation of~~The~~ wind speed at wind turbine hub height is of significance~~important~~ for wind energy assessment and exploitation~~application.~~ ~~To achieve the goal of reducingpeaking 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~~generally ~~assumes a~~the ~~constant coefficient to~~ estimates the hub -height~~level~~ wind speed by assuming a constant exponent between surface and hub height wind speeds~~at wind turbine hub height by assuming log profile~~. Th~~is~~e ~~constant assumption~~inevitably ~~may~~ leads to significant uncertainties in estimating wind ~~energy assessment~~speed profile especially under unstable conditions~~in the surface layer, given the large dependence on a variety of factors, such as terrain, time and height~~. To minimize the uncertainties, we here use~~three~~ a~~the~~ machine learning ~~(ML)~~ algorithm~~s~~ known as R~~r~~andom ~~f~~Forest (RF) to estimate the ~~hub -height~~ wind speed at hub heights~~,~~ such as at 120 m (WS_{120}), 160 m (WS_{160}) and 200 m (WS_{200}).~~,~~ These heights go~~es~~ beyond the traditional wind mast limit of 100-120 m. The radar wind profiler and surface synoptic observations at the Qingdao~~eight coastal~~ station~~s~~ from May 2018 to August 2020 are used as key inputs to develop the RF model~~retrieve wind speed at wind turbine hub height and investigate the wind energy resource~~. A deep analysis of the RF model construction has been performed to ensure its applicability. Afterwards, th~~re~~e RF~~ML~~ model~~s~~ and the PLM are used to retrieve the WS_{120}, WS_{160} and WS_{200}~~wind speed at 120 m (WS_{120}), 160 m (WS_{160}) and 200 m (WS_{200}) above ground level (WS_{120})~~.

~~The comparison of results with the radiosonde observations shows the wind speed from RF model is closer to the observed value than that from PLM.~~

~~At three heights,~~ T~~t~~~~T~~he comparison analyse~~i~~s from both RF and PLM models are performed against ~~of results with the observations~~radiosonde wind measurements. At 120 m, ~~At 120 mOverall, tT~~the RF model ~~at 120 m~~ shows a relative higher correlation coefficient R of 0.93 and~~but~~ a smaller RMSE of 1.09 m/s, compared with the R of 0.89 and RMSE of 1.50 m/s for the PLM.

Notably, the metrics used to determine the performance of model declines sharply with height for the PLM model, as opposed to the stably variation for the RF model. This ~~shows~~ suggests ~~the wind speed from~~ the RF model ~~is closer toagrees better with the observed value than that from~~exhibits advantages over the traditional PLM model. This is ~~due to the factlikely~~ because ~~that~~ the RF model ~~that~~ well considers the factors such as surface friction and~~,~~ heat transfer~~ and high-height wind speed constraints exhibits superiority over the traditional PLM~~. The diurnal and seasonal variations of WS_{120}~~,~~, WS_{160} and WS_{200} from RF ~~wind power density (WPD)~~ are then ~~investigated~~analyzed~~estimated~~. T~~For land stations, t~~he hourly ~~mean~~ WS_{120}~~wind speedWSPD~~ is large~~r~~ at daytime from 0900 to 1600 local solar time (LST) and reach a peak at 1400 LST. The seasonal ~~variation of the~~ WS_{120} is large in~~at~~ spring and ~~winter, and~~winter and is low in summer and autumn~~, which. It is due to the joint influence of the East Asia Monsoon and Mongolian cyclones~~. The diurnal and seasonal variations of WS_{160} and WS_{200} are similar to those~~at~~ of WS_{120}. Finally, we investigated the absolute percentage error ~~difference~~ (APE~~D~~) of wind power density between RF and ~~WPD from~~ PLM at different heights. In the vertical direction, the APE is gradually increased as the height increases. Overall, the PLM algorithm has some limitations in estimating wind speed at hub height~~wind energy assessment at high heightsaltitudes. It is suggested toThe use t~~The~~of~~ RF model that combines more observations or auxiliary data is the more suitable for the hub height wind speed estimation~~is suggested~~. These findings obtained here have great implications~~provide insights~~ for 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 emissions of carbon dioxide, sulfur dioxide and other pollutants (Yuan, 2016; Magazzino et al., 2021). ~~Large amounts of anthropogenic emissions of carbon dioxide and other greenhouse gases can affect the earth's radiation budget are a major driver for the global warming, leading to ever-rising air temperature (Shakun et al., 2012; Shi et al., 2021).~~ To tackle this problem, 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; Leung et al., 2012). As one of the largest energy consuming countries in the world, China is currently facing an increasingly serious energy and climate situation (Khatib et al., 2012). The Chinese government proposes to peak its~~the~~ ~~peak~~ carbon dioxide emissions before 2030 and achieve carbon neutrality ~~strategy to deal with energy and environmental issues~~before 2060 (Pei et al., 2022). With the stimulus of policies and the favor of investors, wind power industry in China is flourishing. Therefore, the scientific assessment of wind energy resources in China is of great importance for the healthy development of wind energy industry in the decades to come.

Characterizing~~The variation~~ ~~o~~the ~~f~~ wind speed at wind –turbine hub height is key for wind energy assessment (Yu et al., 2022). The wind turbine is usually~~generally~~ installed at the top of wind mast with a height of 100-120 m above ground level (AGL), which roughly corresponds to the surface layer (Veers et al., 2019). The wind speed data that have been widely ~~can be~~ used for wind energy assessment are mainly obtained from wind mast, Doppler lidar ~~siteurface observations~~ or reanalysis data (Debnath et al., 2021). The 10 m wind data measured by the ground meteorological station can be used for wind energy assessment (Oh et al. 2012; Liu et al., 2019). The wind tower or mast can also provide wind speed observation data below 100 m AGL (Durisic et al. 2012; Liu et al., 2018). Moreover, the reanalysis data, such as the fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis system (ERA5), can ~~also~~ provide the hourly wind speed at a height of 10 or 100 m AGL for wind energy assessment (Laurila et al., 2021; Gualtieri, 2021). However, the wind turbines are increasing in height and rotor diameter with the development of technology, which go beyond the surface layer and enter the Ekman layer ~~(Veers et al., 2019)~~. Such as some offshore wind power plants, the blade tips of the largest wind turbine can reach heights of 250 m AGL (Gaertner et al. 2020). In addition, 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).

Therefore, the wind profile is important for the selection of wind turbine hub height and the assessment of wind energy.

~~Currently~~It is widely recognized that~~,~~ the wind profile is mainly obtained by empirical formulae (Li et al., 2018), such as the power law method (PLM) ~~(Li et al., 2018)~~. The PLM method generally assumes that the wind speed below 150 m in the planetary boundary layer (PBL) varies exponentially with height (Hellman et al. 1914). This~~It~~ means that the wind speed at the wind turbine hub height can be calculated from the surface wind speed based on a constant power law exponent (α). ~~But~~However, the surface layer wind profile is mainly controlled by the surface roughness, friction velocity and the atmospheric stability (Gryning et al., 2007). The surface layer 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). Due to the influence of inhomogeneous underlying surface and ubiquitous atmospheric turbulence, wind speed varies constantly and greatly in the vertical (Tieleman 1992). Especially~~Moreover,~~ ~~In addition, above~~above~~t the top of the~~ surface layer, the factors, such as the Coriolis force~~parameter~~, baroclinity and wind shear, increase the complexity of the wind profile ~~also influence the wind profile~~ (Brümmer 1991). As a result, ~~It leads that Thus,~~ the α has spatiotemporal variability and depends on a variety of factors, such as terrain, time and height (Li et al., 2018). ~~For example, Durisic et al. (2012) investigateds the wind profile at surface layer in the South Banat region based on meteorological mast data, and pointsed out that the α has obvious diurnal and seasonal changes.~~ ~~In addition, above the surface layer, the Coriolis parameter, baroclinity and wind shear also influence the wind profile (Brümmer 1991).~~ Therefore, the assumption of a constant α poses great challenges and uncertainties to wind energy assessment. Some studies use more complex models to improve the PLM, such as the perturbation theory (Sen et al., 2012) and the bivariate wind speed-wind shear model (Jung et al., 2017). These studies confirm that there is a complex nonlinear relationship between surface observations and wind speed at the wind turbine hub height. Therefore, one of the greatest challenges is to develop an accurate method to describe the nonlinear transfer from surface observations to wind speed at wind turbine hub height.

With the development of machine learning (ML) technology, the ML algorithms have been widely used in the field of wind speed and wind power prediction~~forecasting~~ (Magazzino et al., 2021). Chi et al. (2015) compared two wind ~~speed forecasting~~speed-forecasting~~prediction~~ mechanisms in China based on linear regression and support vector machine ~~(SVM)~~ algorithms. They find~~ound~~ that the ML algorithms have better accuracy in solving the nonlinear problem. Lahouar and Slama et al. (2017) use~~d~~ several ~~weather~~meteorological factors to forecast wind power based on a random forest (RF) model. The ~~R~~results indicate~~d~~ that compared with physical and statistical approaches, the ML model~~s~~ can achieve~~obtain~~ better accuracy when coping with problems that cannot be analytically defined. Therefore, it is worth trying to use ML algorithms to retrieve the wind speed at wind turbine hub height from available observations.

Given the abovementioned problems, we attempt to use a ~~machine learning (~~ML~~)~~ algorithm known as RF~~s~~ to retrieve wind speed at wind turbine hub height~~120 m AGL (WS~~$\sim\!\!\sim_{120}$~~)~~ from the radar wind profiler (RWP) and surface synoptic observations~~multi-source RWP measurements~~. A RF ~~algorithm~~model has been trained based on ~~T~~the surface in situ wind speed, high-height~~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 the classical PLM method and th~~re~~e RF~~ML~~ model~~s~~ are~~were~~ then compared. Next, the wind speeds from the~~ most effective~~ RF model ~~were~~are~~as~~ used to ~~assess~~ evaluate the wind power~~ on coast of China~~. The results of our study can provide useful information for the development of wind energy industry in ~~on the~~ coastal~~ of~~ China. The observational data are~~is~~ ~~briefly~~ introduced in section 2. The RF~~ML~~ model~~s~~ construction and wind energy evaluation method are displayed in section 3. Section 4 discusses the accuracy of the RF~~ML~~ model~~s~~ and the variation of wind energy resources. A summary of results is presented in section 5.

2. Materials and Data

2.1 RWP data~~network of China~~

The RWP is a ground-based remote sensing device that is used to~~can observe~~ measure the atmospheric wind profiles from surface to 5-8 km AGL (Liu et al., 2019). ~~The RWP network of China began to develop as of 2008, and the number of RWP stations increased to 134 by the end of 2020 (Liu et al., 2020). The time resolution of RWP data can reach minute level.~~ It~~The RWP~~ has high and low detection modes in the vertical direction, and their corresponding vertical resolutions are 120 and 60 m, respectively (Liu et al., 2020). ~~However~~Nevertheless, the wind profile~~ observations~~ near the ground surface, especially those ~~(~~below ~~24~~300 m AGL~~)~~ are usually highly uncertain~~ removed~~, due to the influence of ground and intermittent clutter (May and Strauch 1998; Allabakash et al., 2019). Therefore, there exists large data gap between ground surface and the lowest measurement height provided by the RWP.

Here, the RWP data are~~is~~ obtained~~provided by a coastal observational station in~~at Qingdao ~~(~~(120.23 °E, 36.33 °N), which is a typical coastal synoptic weather station. The spatial distribution and surface type of this~~ese~~ station~~s~~ are~~is~~ shown in Fig. 1~~, marked by red points~~. Geographically, Qingdao station is located on the south of Shandong ~~Peninsula, and~~Peninsula and ~~isare near the boundary of~~ lies to the west of –the Yellow Sea. To be more specific, ~~T~~this station is set up in the suburb, surrounded by cropland. The altitude of thise station is 12 m above mean sea level. The hourly wind speed (WS_{300}) and direction (WD_{300}) data at 300 m AGL~~s~~ are~~are~~ obtained from 1 May 2018 to 31 August 2020. The original RWP data at 6-min intervals have~~s~~ not been released temporarily, but ~~it~~ can be reasonably requested upon demand by contacting ~~to~~ Dr. Jianping Guo ~~by~~ (E~~e~~mail: ~~(~~jpguocams@gmail.com).

2.2 Anemometer

The wind cup anemometer can measure the instantaneous wind ~~speed, and~~speed and is installed at 10 m AGL (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 at ~~eight~~ Qingdao ~~RWP~~ station~~s~~. The 10 m wind speed ~~data~~ (WS_{10}) and direction (WD_{10}) data can be downloaded in http://www.nmic.cn/data/cdcdetail/dataCode/A.0012.0001.html (last access: 15 November 2022). Here, the WS_{10}~~10 m wind speed data~~ data ~~at the eight RWP stations~~ are~~were~~ also obtained from 1 May 2018 to 31 August 2020. The WS_{10}~~10 m wind speed data~~ data are~~iswas~~ processed into hourly average value to match the RWP data.

2.3 Radiosonde data

The radiosonde (RS) provides the vertical profiles of wind speed and wind direction at 5-8 meter intervals ~~twice a day at 0800 and 2000 local solar time (LST)~~ (Guo et al., 2020). 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 provide observations of wind profiles only twice per day: 0800 and 2000 local solar time (LST). T~~Note that only t~~he ~~station of~~ Qingdao station is equipped with RS and RWP at the same time. The RS data also collect~~ed~~ from 1 May 2018 to 31 August 2020, which can be downloaded from http://www.nmic.cn/data/cdcdetail/dataCode/B.0011.0001C.html (last access: 15 November 2022).

2.4 ERA5 data

The 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 degree, and there are 137 vertical levels in vertical direction. “ERA5 hourly data on single levels from 1959 to present” is a dataset of ERA5, which can provide a series of surface parameters such as temperature, humidity, pressure and radiation etc. (Hersbach et al., 2020). It can be downloaded from the website of https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview (last accessed: ~~on~~ 15 November 2022). Here~~Therefore~~, 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 ~~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~~ Qingdao station, the grid where the RWP station is located is selected and those parameters in the corresponding grid are obtained accordingly. These data are~~were~~ obtained from 1 May 2018 to 31 August 2020 ~~at eight stations~~.

3. Methods

The schematic diagram of surface layer wind profile observations is shown in Fig. 2. The wind mast or tower can provide wind speed data below 100 m AGL (Durisic et al. 2012; Liu et al., 2018). The RWP ~~network of China~~ can measure the wind profiles from the 300 m to a height of 5-8 km AGL (Liu et al., 2019). It leads to a gap (100 to 300 m) in the observation of wind profile. At present, the PLM method is most often applied to extrapolate the surface wind speed to the wind turbine hub height, such as wind speed at 120 m (WS_{120}), 160 m (WS_{160}) and 200 m (WS_{200}) AGL ~~(WS_{120})~~.

3.1 Power law method

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

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

where v_1 and v_2 are the wind speed at height h_1 ~~(10 m)~~ and h_2~~, (120 m)~~, respectively. The α is the power law exponent ~~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 generally is estimated to range from 0.1 to 0.4 (Li et al., 2018). Here, the general value of α for coastal topography is~~was~~ set to 0.15 based on former studies (Patel et al., 2005; Banuelos et al., 2010). However, Jung et al. (2021) pointed out that the error in the wind power density estimation over China can reach to 30 % based on a constant α value. Therefore, we attempt to use ML algorithm~~s~~ to obtain the WS_{120}, WS_{160} and WS_{200}. ~~–~~

3.2 ~~Ma~~RF~~chine learning~~ algorithm~~s~~

RF is an ensemble ML method~~ (Breiman, 2001)~~, which has been widely used in regressive calculation (Breiman, 2001). It is a method to integrate many decision trees into forests and predict the result. Schematic diagram of RF is shown in Fig. S1~~e~~. The RF is composed of many decision trees, and each decision tree is irrelevant. The performance of RF is determined by the aggregation of the results of all the trees (Ma et al., 2021). For RF model, the number of trees (N) is an important parameter to achieve the optimal performance of the model. The further detailed information can be referred to Breiman (2001).

3.2.1 Inputs for RF

In the construction of the RF model, it is necessary to obtain the relevant ~~a series of~~ variables that may affect the surface wind profile according to the physical mechanism and previous research. At present, the PLM is often used to calculate the wind speed at hub height. It confirms that the wind speed at hub height is related to the wind speed at other heights (Durisic et al., 2012; Li et al., 2018). Therefore, the WS_{10}, WD_{10}, WS_{300} and WD_{300} are selected as inputs. The surface wind profile also depends on the surface roughness, friction velocity and the atmospheric stability~~ and surface stress~~ ~~and on the atmospheric stratification~~ (Smith, 1988; Gryning et al., 2007), ~~,~~so that ~~Char,~~ FSR,~~;~~ FV and Char ~~and Temp~~ are also regarded as inputs~~ variable~~. The higher FSR causes a slower wind speed in the surface layer. ~~At a land station, surface roughness is derived from the vegetation type (Li et al., 2021).~~ The FV is a theoretical wind speed at the Earth's surface, which is used to calculate the way wind changes with height in near surface~~the surface layer~~ (Stull, 1988).~~The FV is used to calculate the way wind changes with height in the surface layer (Stull, 1988).~~ Moreover, considering that the generation of wind is closely associated with uneven heating of the Earth's surface by solar radiation (Solanki et al., 2022), the Rn, LHF and SHF are also selected as input variables. Additionally, some studies use atmospheric temperature and ~~atmospheric~~ pressure as input to improve the accuracy of wind speed prediction (Chi et al., 2015). Here, we also regard DP, Temp~~, Temp~~ and Press as the input variables. The reference value, also included as input in the RF~~ML~~ model~~s~~, is the ~~WS_{120}~~ WS_{120}, WS_{160} and WS_{200} measured from RS.

3.2.2 Feature selection

To estimate the WS_{120}, WS_{160} and WS_{200}, we need to build RF model on 120 m (RF_{120}), 160 m (RF_{160}) and 200 m (RF_{200}), respectively. For each~~RFeach ML~~ model, it is necessary to select the main features from the input~~ value~~s to avoid data redundancy and reduce the complexity of the model (Ma et al., 2021). Following the~~ previous~~ research of Gregori et al. (2022), the ~~variable~~inputs,~~s~~ which cannot ~~that~~

~~after being discarded do not~~ cause a 2% reduction in correlation coefficient, are regarded as irrelevant feature and removed ~~(Gregori et al., 2022).~~ Figure 3 shows the importance analysis of input~~ variable~~s for three RF~~ML~~ models. ~~The relative importance of the variable indicates the dependence of the model on this parameter.~~ The relevant~~main~~ features~~ of each model~~~~input variables with importance lager than 0.1~~ are~~were~~ marked by red bars. The irrelevant features are marked by blue bars, which are not regarded as final inputs in three RF~~final ML~~ models~~s~~. For three RF models,– the ~~main~~relevant features are both WS_{10}, FV, Char, SHF and WS_{300}~~although the importance values of WS_{10} and FV are large, the importance values of some inputs are also relatively large with varies from 0.1-0.15~~. It indicates~~d~~ that ~~except for surface stress,~~ the factors such as surface friction, heat transfer and high-height~~altitude~~ wind speed constraints are~~will also be~~ considered in the construction ~~inversion process~~ of RF models. In addition, it is surprising that FSR has such low importance in three RF models construction. FSR is a measure of surface resistance, which directly affects the near-surface wind speed (Smith, 1988). At a land station, the FSR~~surface roughness~~ is derived from the vegetation type (Li et al., 2021)~~(Li et al., 2021)~~. The surface type of Qingdao station is cropland. Li et al. (2021) confirms that the FSR at cropland is most likely to 0.3 m. In training data, the FSR from ERA5 also approximates a constant value (0.3 m). Since the constant variable has no meaning for RF model construction, the RF model divides FSR into irrelevant variable. Therefore, the final inputs for three RF models are WS_{10}, FV, Char, SHF and WS_{300}.

3.2.3~~4~~~~3~~ ~~Model Ttraining~~tuning parameters ~~RF~~

RF algorithm requires to setup the ~~hyperparameters~~N in order to avoid overfitting in the training dataset (Ma et al., 2021). ~~The most important hyperparameter of RF is the number of tree (N).~~ Here, we use~~d~~ the RF algorithm for regression in MATLAB R2020b. The code and usage of RF are referred to the MATLAB help cente~~re~~ (https://ww2.mathworks.cn/help/stats/treebagger.html, last access: 15 November 2022). The specific tuning parameter~~training~~ process of ~~each~~RF model is presented as follows:~~.~~ The N value varies from 1-500 with an interval of ~~2~~10. Correlation coefficient (R) and root mean square error (RMSE) are~~were~~ used to evaluate the accuracy of the model. We need to set an appropriate N value to maximize R and minimize RMSE. ~~–~~Fig.~~ures~~ S2~~c and 2f~~ shows the tuning parameter~~s~~ process for the ~~number of tree (~~N of three RF models~~)~~. For RF_{120}, i~~I~~t can find that the R increased with N value increased, while the R is~~was~~ almost unchanged when N value is greater than 100. When N equals ~~300~~200, R reaches the maximum value (0.~~81~~82) and RMSE reaches the minimum value (1.~~64~~ 68 m/s). Therefore, the N value is set to ~~300~~ 200 for RF_{120}. Moreover, according to the same tuning parameter~~s~~ process~~principle~~, the N values are set to 300 and 150 for RF_{160} and RF_{200}, respectively. After determining the final inputs and N values~~ hyperparameter~~, the three RF models have been trained and tested. At Qingdao Station, a total of 746 sample data are obtained after data matching. We use the 5-fold crossover to train RF models. The test results are discussed in section 4.1.

3.2.4~~3~~ Sensitivity analysis

The accuracy and generalization of the RF model depend on training and testing samples (Ma et al., 2021). However, the training and testing samples are~~is~~ obtained at 0800 and 2000 LST. It needs to discuss whether the RF model also applies to other times. This depends on whether the RF model has enough generalization for the training samples, and whether the inputs at other times have appeared in the training samples. Fig. S3-S5~~4~~ shows the difference between estimated wind speed ~~WS_{120}~~ and observed wind speed~~WS_{120}~~ of three RF models, which as a function of the inputs. For three RF models, the deviations are~~was~~ relatively stable and~~ did~~ not change with the increase of inputs. It indicates that three RF models have good generalization for the training and testing samples. This is~~could be likely due to the fact that~~because the RF tends to increase random disturbance in the sample space, parameter space and model space, thereby reducing the impact of "cases" and improving the generalization ability (Breiman, 2001). Moreover, Fig.~~ure~~ S6 shows the distribution of inputs at different time. The red dashed lines represent the maximum and minimum values of each variable at training samples. In the range of the red line, three RF models can provide stable output due to its good generalization ability. It can be found that almost all the inputs have appeared in training samples. Therefore, three RF models have sufficient generalization and can be used at~~in~~ other times.

3.3~~4~~ Assessment methods of wind energy

For the wind speed at hub height~~obtained WS_{120}~~, a series of indicators have~~need to~~ been 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).

3.3~~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):

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

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

where v is the WS_{120}; k and c are the shape parameter of the Weibull distribution. Higher c indicates larger wind speed, while the k indicates wind stability. 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):

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

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

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:

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

3.3~~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):

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

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). In addition, the absolute percentage error (APE) is used to quantify the differences in wind energy assessment based on different methods. The APE is~~was~~ calculated by:

$$APE = \frac{|WPD_{RF} - WPD_{PLM}|}{WPD_{RF}} * 100\ \% \tag{8}$$

where WPD_{RF} and WPD_{PLM} are calculated by the wind speed from RF and PLM, respectively.

4. Results and discussion

4.1 Intercomparison of ~~WS_{120}~~ wind speed using different ~~methods~~methods.

Figure 4 shows the wind profile from different methods under different time. The red, black and blue lines represent the mean wind speed from RS, PLM and RF, respectively. For the PLM, the retrieved results below 80 m AGL are consistent with the RS observations. Gryning et al. (2007) also pointed out that the wind profile based on surface-layer theory is valid up to a height of 50–80 m. Above 80 m AGL, the wind speeds retrieved by PLM deviate from the RS observations. This deviation is increasing with the height. The comparison results between PLM and RS at 120 m, 160 m and 200 m AGL (Fig. 5) are also confirmed it. This is due to the fact that above surface layer, the Coriolis force, baroclinity and wind shear increase the complexity of the wind profile (Brümmer 1991). Moreover, most of estimated results from PLM are underestimated when the observed ~~WS_{120}~~wind speed is high, especially at 200 m AGL. T~~t~~he reason is that~~is is due to~~ the~~e~~ surface wind profile~~WS_{120}~~ 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. The s~~S~~imple exponential relationship is unable to obtain the surface wind profile~~WS_{120}~~ with high accuracy, especially at high wind speed condition. By comparison, the WS_{120}, WS_{160} and WS_{200} retrieved from RF are closer to RS observations. Compared with PLM, the R and RMSE between the observed ~~WS_{120}~~ wind speed and the estimated ~~WS_{120}~~ wind speed from RF at three heights are significantly improved (Fig. 5). ~~TEspecially for the RF, the highest R (0.934) and the smallest RMSE (1.090 m/s) show that the RF is the best model to retrieve WS_{120}.~~ This is due to the fact that ~~t~~the surface friction (FV), ~~These results indicate that considering~~ heat transfer (SHF) and high-height~~altitude~~ wind speed constraints (WS_{300}) are considered in construction ~~inversion process~~of RF, which can improve the accuracy of the model. Moreover, it notes that three RF models tend to slightly overestimate small values and underestimate high values. ~~This may be due to~~The reason is the small number of training samples at high and low values, resulting in the reduction of RF model generalization. Overall, it can be seen from the metrics of R and RMSE that the wind speed~~accuracy~~ from~~of~~ RF model~~s~~ is better than that from~~of~~ PLM.

In addition, for both PLM and RF, the retrieved wind profile at 2000 LST~~nighttime~~ is closer to the RS ~~observed~~ observations. The comparisons between the observed wind speed and the estimated wind speed for PLM and RS under different time is shown in Fig. S7. The fitting results of PLM and RF at 2000 LST are slightly higher than that at 0800 LST. It indicates that the performance of PLM and RF vary with hour of the day. This~~e~~ is because~~due to~~ the wind profile depends not only on the surface friction but also on the atmospheric stratification (Gryning et al., 2007). ~~SinceThis is due to the~~ The surface layer is in an unstable stratification due to heat transfer caused by solar radiation during daytime, while the surface layer tends to stable stratification due to surface radiation cools during nighttime (Yu et al., 2022; Solanki et al., 2022). The WS_{120}, WS_{160} and WS_{200}~~wind speed WS_{120} is~~are more vulnerable to the surface turbulence due to the unstable stratification during daytime. Therefore, the performance of PLM and RF at nighttime is better than that at daytime.

Figure 6 shows the comparisons between the observed results ~~WS_{120}~~ and the estimated results ~~WS_{120}~~ for ~~four methods~~PLM and RF under different season. The red, green, blue and black represent the spring, summer, autumn and winter, respectively.~~–~~ At three heights, the performance of PLM is the best in winter and the worst in summer. It shows that the performance of PLM is affected by seasonal factors, which is~~ likely~~ due to the wind shear varying dramatically with season (Banuelos-Ruedas et al., 2010). Pérez et al. (2005)~~Previous study~~ indicates~~d~~ that the surface layer wind speed profile is mainly affected by the convection produced by surface heating in summer~~ (Pérez et al., 2005)~~. The ~~WS$_{120}$ is~~WS$_{120}$, WS$_{160}$ and WS$_{200}$ ~~affect~~affect by the surface due to the unstable stratification, which leads that the PLM performs worst in summer. In contrast, during winter, the surface temperature is generally lower than the air temperature aloft creating a stable inversion (Yu et al., 2022; Liu et al., 2022). The WS$_{120}$, WS$_{160}$ and WS$_{200}$ ~~WS$_{120}$ is~~are disconnected from the surface due to stable stratification. It leads that the PLM performs best in winter. As for RF, although the performance~~fitting result~~ in spring is slightly lower than that in other seasons,~~.~~ the fitting results at four seasons are significantly improved compared with the ~~other three method~~PLMs. This indicates that RF is least affected by seasons. The reason is that the RF model is less subjective than PLM because they are data driven. Overall, in terms of stability and accuracy, the RF is more suitable for~~the best model to~~ estimating wind speed at hub height~~retrieve WS$_{120}$~~.

4.2 ~~*Characteristics*~~ *Vertical profiles of wind speed at surface layer*

Figure 7~~8~~ shows the diurnal and seasonal variations of WS$_{120}$~~-~~, WS$_{160}$ and WS$_{200}$~~WPD at Qingdao~~~~eight stations~~. The diurnal and seasonal variations of wind speed at three heights are on average similar to each other. From the perspective of daily variation, the ~~WS$_{120}$~~ wind speed is larger at daytime from 0900 to 1600 LST, while is lower at nighttime from 0000 to 0400 LST. This daily cycle ~~of WS$_{120}$~~ is mainly affected by the solar radiation and the sea-land breeze. On the one hand, the surface is heated by solar radiation at daytime, warming the low-level air. The convection formed by rising warm air mass results 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, the difference of specific heat capacity between sea and land can form the difference of thermal properties between sea and land. 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 are~~were~~ also observed at three other stations in China (Liu et al., 2013). From the perspective of seasonal variation, the wind speed ~~seasonal distribution of WS$_{120}$~~ is large in~~at~~ spring and ~~winter, and~~winter and is low in summer and autumn. This is ~~due to~~because the influence of East Asia Monsoon and Mongolian cyclones (Yu et al., 2016). 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).

The histograms of WS$_{120}$~~-~~, WS$_{160}$ and WS$_{200}$ with corresponding Weibull distributions ~~at eight coastal stations~~ are plotted in Fig. ~~9~~8~~7~~. The blue bar and pink lines represent occurrence probability and Weibull distributions, respectively. Moreover, the mean ~~WS$_{120}$~~ wind speed and Weibull distribution parameters for three heights ~~all eight stations~~ are listed in Table 1~~2~~. The occurrence ~~probability~~probabilities of WS$_{120}$~~-~~, WS$_{160}$ and WS$_{200}$ are~~is~~ both the unimodal distribution, with a peak probability in medium wind speed (about 5 m/s) and a low probability in high and low wind speed. The mean WS$_{120}$, WS$_{160}$ and WS$_{200}$ are 5.84, 6.26 and 6.57 m/s, which~~wind speed~~ gradually increases with height~~at island stations is slightly higher than that at coastal land stations~~. ~~This is due to t~~The lower wind speed near the ground is caused by the influence of underlying surface roughness and surface friction~~atmospheric stability, resulting in the lower wind speed near the grounddifference between sea and land breeze~~ (Li et al., 2018; Li et al., 2020). In addition, there is a deviation between the probability density function and the frequency of occurrence at some stations, which is ~~due to the fact that~~because Weibull distribution generally has a long tail effect or a right skewed distribution (Pishgar-Komleh et al., 2015; Ali et al., 2018). Overall, the Weibull distribution matches with the frequency of wind speed at all stations. Therefore, the Weibull distribution parameters can be applied for the wind energy assessment.

4.3 ~~Vertical profiles of wind speed at surface layer~~ Influence of wind speed from different methods on WPD

~~4.3 VariInfluence on wind energy assessmentation of wind resource~~

Figure 9 shows the diurnal variations of WPD from PLM and RF at 120 m, 160 m and 200 m AGL. The red solid and dotted lines represent the variation of WPD from RF and PLM, respectively. The gray bar represents the absolute percentage error ~~difference~~ (APE~~D~~) of WPD between RF and PLM. The diurnal pattern of WPD from RF is ~~similar to~~like that from PLM. At three heights, t~~t~~he hourly mean WPD is larger at daytime from 0900 to 1600 LST with a peak at 1400 ~~LST, and~~LST and is lower at nighttime from 0000 to 0400 LST. On the contrary, the– APE~~D~~ is lower at daytime (0800 to 1800 LST) and larger at nighttime (2000 to 0600 LST). At 120 m, the mean APE~~D~~ at daytime and nighttime are~~is~~ 14.09 ~~40~~% and 35.80 ~~20~~%, respectively. Considering that the results from RF are underestimated at high wind speed condition, the APE~~D~~ of WPD between~~ the~~ PLM and~~ the~~ actual observation at~~in~~ daytime should be slightly greater than 14.09 ~~20~~%. Moreover, the diurnal variations of APE~~D~~ at 160 m and 200 m AGL ~~are similar to~~generally resemble ~~that~~ the features obtained at 120 m AGL. But the APE~~D~~ of WPD between RF and PLM increases with the height. These results indicate that the PLM is more suitable for wind energy assessment in the daytime, and the error of wind energy assessment based on PLM is gradually increased as the height increases.

Figure 10 shows the monthly variations of WPD from PLM and RF at 120 m, 160 m and 200 m AGL. The monthly variation~~pattern~~ of WPD from RF is also similar to that from PLM. T~~t~~he monthly WPD is relatively high for the period from March to May, as compared to the lower values from June ~~August~~ to October. At 120 m, the APE~~D~~ is largest in~~at~~ summer and is lowest ~~at~~in winter. The seasonal APE~~D~~ during spring, summer, autumn and winter are 23.65 %, 40.83 %, 19.67 % and 12.62 %, respectively. The monthly variations of APE~~D~~ at 160 m and 200 m are consistent with that at 120 m. It indicates that the PLM is more suitable for wind energy assessment ~~at~~in autumn and winter. In addition, the APE~~D~~ during spring at 120 m, 160 m and 200 m are 23.65 %, 28.12 % and 34.22 %, respectively. Due to the performance of RF model is the worst in spring, the APE~~D~~ of WPD between PLM and real value during spring may increases. Jung et al. (2021) also finds that the global median absolute percentage error in the wind energy estimations is 36.9% assuming the power law exponent being 0.14. Overall, the PLM has some limitations in wind energy assessment above 100 m. When using PLM to evaluate wind energy at high height~~altitude~~, it is necessary to pay attention to its errors. Moreover, the use of RF model that takes ~~into account~~the –factors such as surface friction, heat transfer and high-height~~altitude~~ wind speed constraints into account is suggested to evaluate wind energy~~ at high altitude~~.

5. Summary and conclusions

The traditional methods such as the ~~power law method (~~PLM~~)~~ used to estimate wind speed at hub height generally assume a constant exponent α in establishing the power law relationship between wind speeds at surface and hub -height, which inevitably leads to large uncertainties. To confront this challenge, t~~T~~his study ~~used~~ uses the ~~random forest (RF)ML~~ algorithm~~s~~ to retrieve the~~evaluate~~ ~~$WS_{120}$the wind energy resource at eight coastal stations~~wind profile based on the ~~wind speed profile~~RWP and surface meteorological data from May 2018 to August 2020. ~~Moreover, the accuracy wind estimates fromof PLM, KNN, SVM and RF was areisare compared against based on the comparison between model observed WS_{120} and estimatesed and observationsresults and observed value WS_{120}. Finally, tThe diurnal and seasonal variations of wind speed and WS_{120}, WS_{160} and WS_{200} from RF are then analyzed. Finally, the wind energy resource at 120 m, 160 m and 200 mat three heights eight coastal stations wasare evaluated based on the wind speed from PLM and RF based on the WS_{120} wind speed from RF.~~

The comparison against observations indicates that the WS_{120} estimated~~s~~ from RF are~~is~~ ~~closer to the observed value than that~~better than those from PLM, given~~.~~ the ~~Tthe~~ relative higher R (0.93 versus

0.89) and smaller RMSE (1.09 m/s versus 1.50 m/s)~~ of RF are better than the R (0.89) and RMSE (1.50 m/s) of PLM RF model ~~. ~~Moreover~~Particularly, the performance of PLM declines with height. Especially at 200 m, the R and RMSE ~~between WS_{200} fromof~~from PLM~~ and observed value~~ ~~is~~are change~~d~~ to 0.78 and 2.42 m/s, respectively. In contrast, the RF model maintains good accuracy at different heights. The R (RMSE) for RF model at~~ 120 m,~~ 160 m and 200 m are~~ 0.93 (1.09 m/s),~~ 0.91 (1.29 m/s), and 0.91 (1.48 m/s), respectively. These results show that above the surface layer, the wind speeds from PLM deviate from the observed value. The RF model is more suitable for retrieving the hub height wind speed, w~~W~~hen the hub height is extended above the surface layer. Overall, the ~~accuracy of three ~~RF~~ML~~ model shows advantages over~~s is better than that of~~the traditional PLM. This is ~~probably due to the fact that~~because the RF~~ML~~ model~~s~~ ~~w~~well considers the influence of near-surface environmental parameters, ~~to improve accuracy, ~~such as friction velocity~~ FV~~ and charnock coefficient~~Char etc~~. Moreover, the heat transfer and high-height~~altitude~~ wind speed constraints are also considered in the construction of RF model. Based on the wind speed ~~WS_{120}, WS_{160} and WS_{200}~~ from RF, the diurnal and seasonal variations of wind energy are then analyzed. The hourly mean WPD is larger ~~at daytime ~~from 0900 to 1600 LST with a peak at 1400 LST. The WPD is relatively high ~~at~~in spring and winter, as compared to the lower values ~~at~~in summer and autumn. Finally, the differences of WPD between RF and PLM at different heights ~~time scales ~~are investigated. At 120 m, the mean APE~~D~~ of WPD between RF and PLM at daytime and nighttime are 14.09 % and 35.80 %, respectively. Moreover, the seasonal APE~~D of WPD between RF and PLM~~ at 120 m is largest ~~at~~in summer (40.83 %) and is lowest ~~at~~in winter (12.62 %). In addition, the mean APE at 120 m, 160 m and 200 m are 24.19 %, 27.99 % and 32.57 %, respectively. These results indicate that there are some errors in the wind energy evaluation based on wind speed from PLM. Therefore, when retrieving high height~~altitude~~ wind speed, it is suggested to combine more observation or auxiliary data to build a more accurate model, such as RF model. In the absence of other observation data, it is necessary to pay attention to the~~ its~~ errors when using PLM to evaluate wind energy~~speedenergy~~ at high height~~altitude~~. –

Our work ~~obtains~~provides a new pathway to fill the data gap of wind speed at the hub height for the high capability of ~~–the accurate wind profile~~ ~~comprehensively assesses the wind energy resources on the coast of China using ~~the state-of-the-art ML algorithm, which ~~provides ~~lays a solid foundation ~~new mean~~~~invaluable information ~~for more robust~~the development of ~~wind energy assessment~~industry in the coastal regions of China in the future~~. However, ~~wind energy~~the high-precision wind profile estimate ~~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 ~~RWP~~ output data and codes used in this paper can be provided for non-commercial research purposes upon motivated request (Jianping Guo, Email: jpguocams@gmail.com). The anemometer 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 data can be downloaded in https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (under grants 42001291), the Fundamental Research Funds for the Central Universities (under grants 2042022kf1003), and the Open Grants of the State Key Laboratory of Severe Weather (under grants 2021LASW-B09).

Author Contributions

The study was completed with cooperation between all authors. JG and BL designed the research framework; BL and JG conducted the experiment and wrote the paper; XM, HL, SJ, YM, and WG 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.

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.

Brümmer B.: Wind shear at tilted inversions. Boundary-Layer Meteorol 57:295–308, 1991.

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.

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.

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.

Debnath, M., Doubrawa, P., Optis, M., Hawbecker, P., and Bodini, N.: Extreme wind shear events in US offshore wind energy areas and the role of induced stratification, Wind Energ. Sci., 6, 1043–1059, 2021.

Durisic Z., Mikulovic J.: Assessment of the wind energy resource in the South Banat region, Serbia. Renew Sust Energ Rev., 16(5):3014–3023, https://doi.org/10.1016/j.rser.2012.02.026, 2012.

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):1277–1283, https://doi.org/10.1016/j.renene.2010.10.003, 2011.

Gryning, S. E., Batchvarova, E., Brümmer, B., Jrgensen, H., and Larsen, S.: On the extension of the wind profile over homogeneous terrain beyond the surface boundary layer. Boundary-Layer Meteorology, 124(2), 251-268, 2007.

Gaertner, E., Rinker, J., Sethuraman, L., Zahle, F., Anderson, B., Barter, G., Abbas, N., Meng, F., Bortolotti, P., Skrzypiński, W. R., Scott, G., Feil, R., Bredmose, H., Dykes, K., Shields, M., Allen, C., and Viselli, A.: Definition of the IEA 15-Megawatt Offshore Reference Wind Turbine, Golden, CO: National Renewable Energy Laboratory, NREL/TP-5000-75698, https://www.nrel.gov/docs/fy20osti/75698.pdf, 2020.

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):3097–3124, 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., 24:142–148, https://doi.org/10.1016/j.rser.2013.03.058, 2013.

Jung, C.~~hristopher~~, and D~~irk~~. Schindler.: The role of the power law exponent in wind energy assessment: A global analysis. International Journal of Energy Research 45.6, 8484-8496, 2021.

Khatib, H.: IEA World Energy Outlook 2011-A comment. Energ Policy., 48:737–743, 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):1031–1039, 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, J., Guo, J., Xu, H., Li, J., and Lv, Y.: Assessing the surface-layer stability over china using long-term wind-tower network observations. Boundary-Layer Meteorology, 180(1), 155-171, 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, 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., 24:379–386, 2013.

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):555–563, 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.

Pérez, I. A., García, M. A., Sánchez, M. L., & De Torre, B. (2005). Analysis and parameterisation of wind profiles in the low atmosphere. Solar Energy, 78(6), 809-821.

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.

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, 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):163–173, 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 ~~2~~ 1 Statistics for the Weibull distribution of WS_{120}~~-~~, WS_{160} and WS_{200} ~~at the eight stations~~ from 1 May 2018 to 31 August 2020.

Height (m)~~Station~~	Mean wind speed (m/s)~~WS_{120} (m/s)~~	Standard deviation (m/s)	Weibull Shape factor k	Weibull Scale factor c (m/s)
120~~Dongying~~	5.84~~5.54~~	2.54~~1.77~~	2.47~~3.46~~	6.58~~6.16~~
160~~Penglai~~	6.26~~5.27~~	2.59~~2.39~~	2.60~~2.35~~	7.05~~5.95~~
200~~Qingdao~~	6.57~~5.86~~	2.80~~2.45~~	2.52~~2.58~~	7.40~~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:

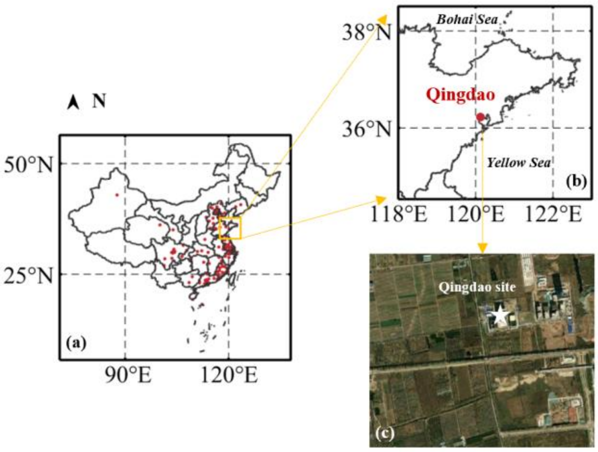

Figure 1. (a, b) Geographical distribution and (c) surface type of the radar ~~eight radar~~ wind profiler ~~observational~~ station~~s~~ ~~(red dots) in the coast of East China~~at Qingdao.

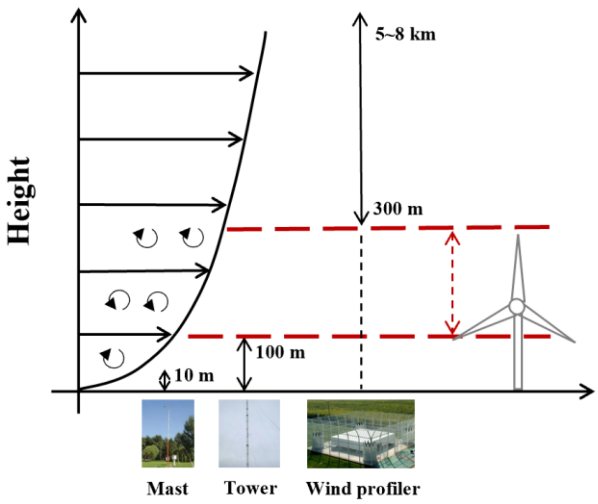

Figure. 2 The schematic diagram of surface layer wind profile observations.

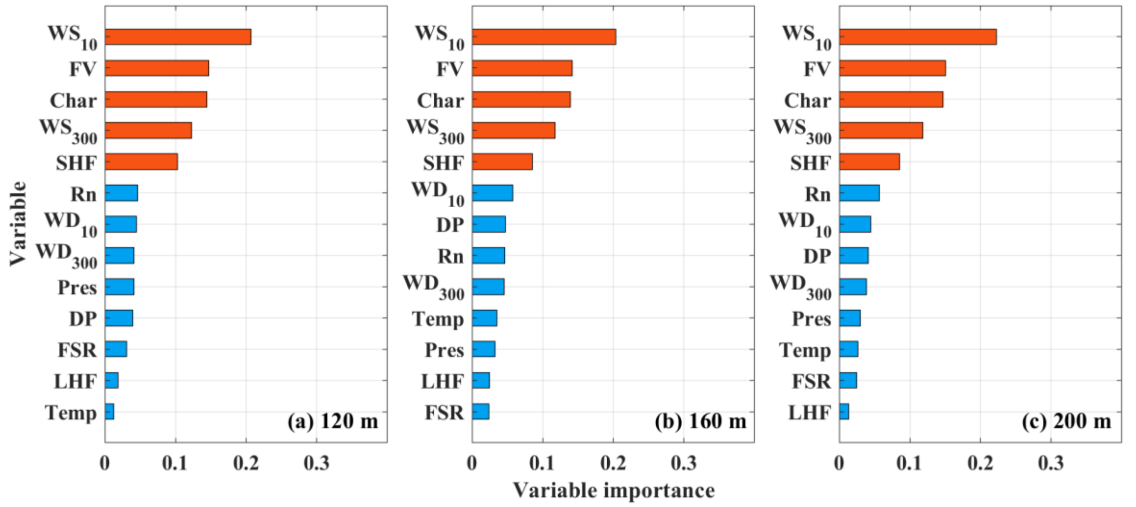

Figure 3. Importance analysis of input~~ variable~~s for RF~~three~~ model at~~s:~~ (a) 120 m~~KNN~~, (b) 160 m~~SVM~~, and (c) 200 m~~RF~~.

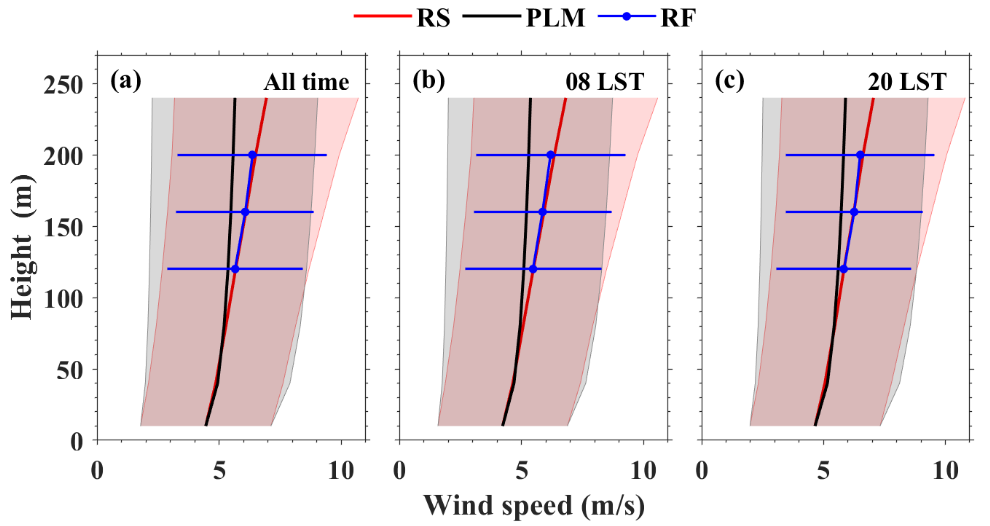

Figure 4. Vertical profiles of the ~~mean~~ wind speed from different methods at (a) all time, (b) 0800 and (c) 2000 LST. Red, black and blue lines represent mean wind profile from RS, PLM and RF, respectively. Corresponding color shading areas represent the standard deviation.

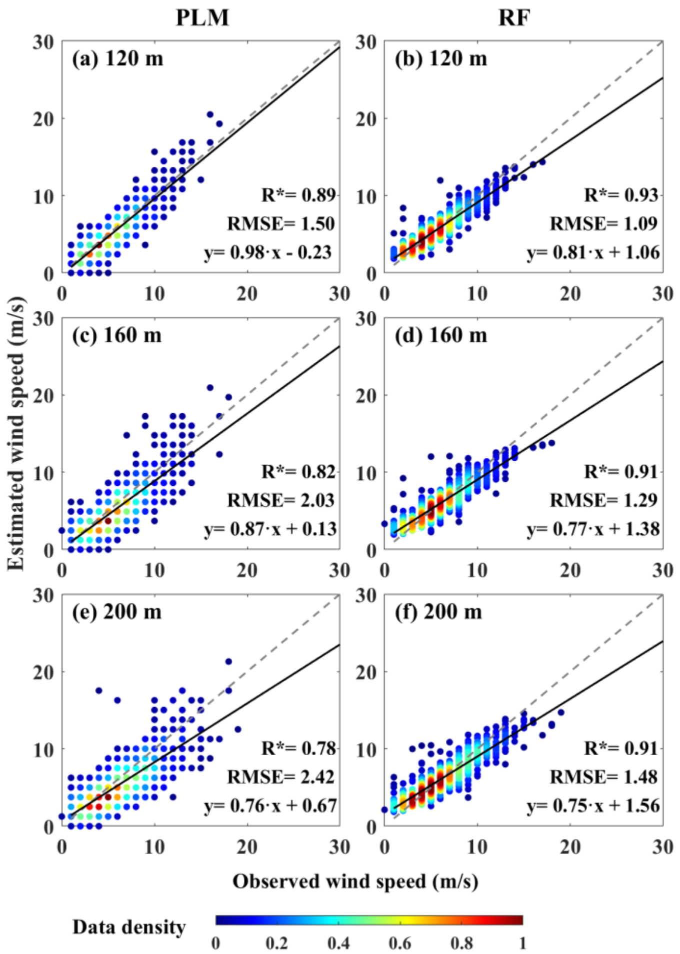

Figure 5. Comparisons between observed ~~WS_{120}~~ wind speed and estimated ~~WS_{120}~~ wind speed for (a, c, e) PLM and (b, d, f) RF at 120 m, 160 m and 200 m. ~~based on the (a, e, i) PLM, (b, f, j) KNN, (c, g, k) SVM and (d, h, l) RF models under different time.~~ The gray and black line is the reference and regression line, respectively. The 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%.

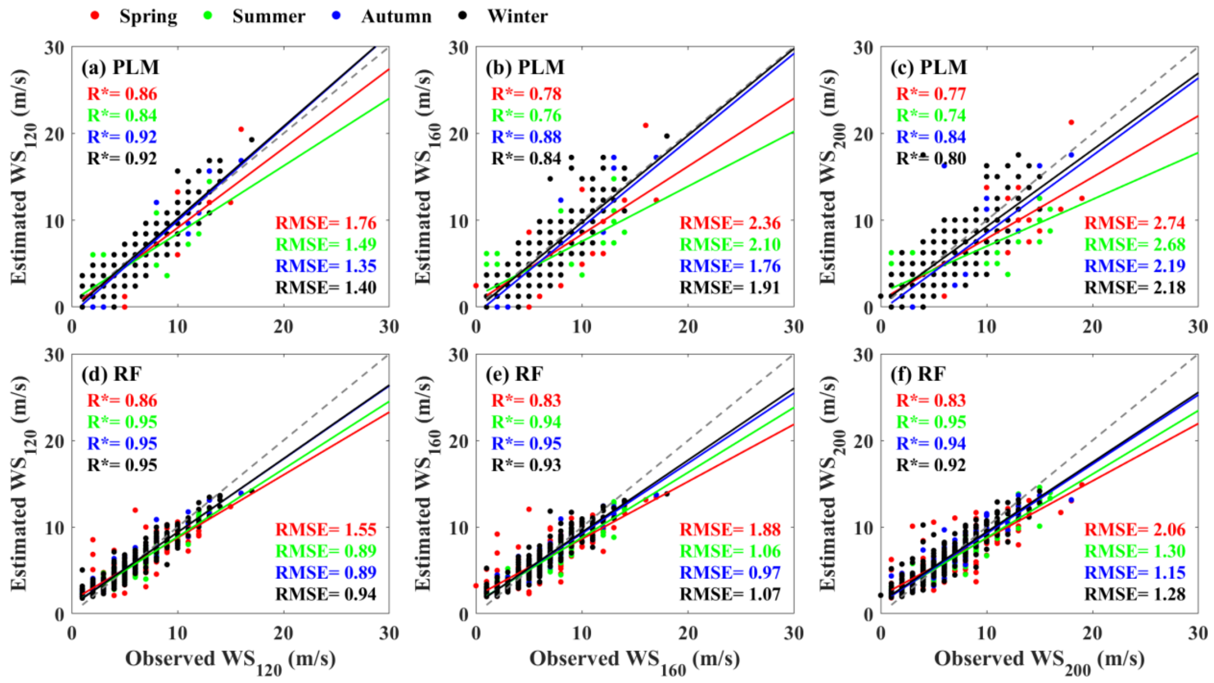

Figure 6. Comparisons between observed wind speed and estimated wind speed for (a, b, c) PLM and (d, e, f) RF at 120 m, 160 m and 200 m~~observed WS_{120} and estimated WS_{120} based on the (a) PLM, (b) KNN, (c) SVM and (d) RF models~~ under different season. The red, green, blue and black represent spring, summer, autumn and winter, respectively. The asterisk indicates that the correlation coefficient (R) has passed the t-test at a confidence level of 95%.

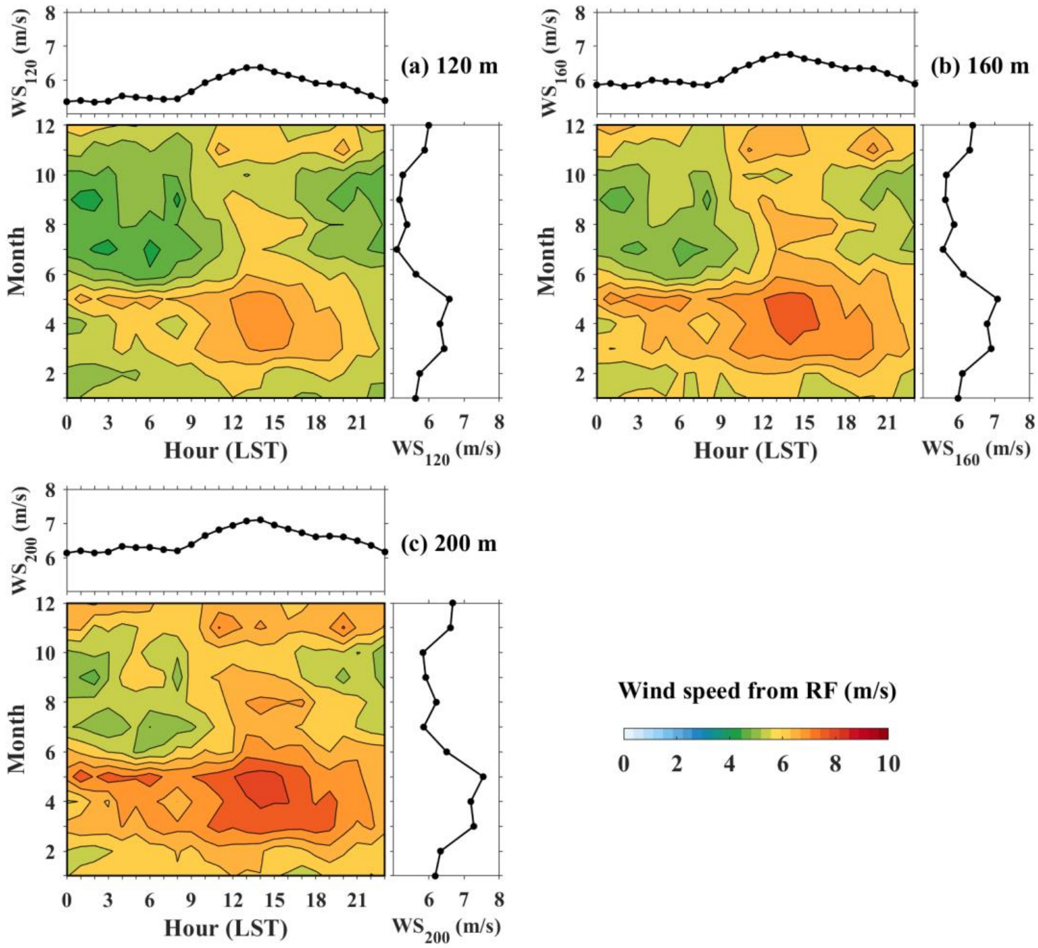

Figure 7. Monthly and diurnal cycles of (a) WS~~ws~~$_{120}$~~-~~, (b) WS_{160} and (c) WS_{200} from 1 May 2018 to 31 August 2020. Color bar represents the wind speed from RF model.

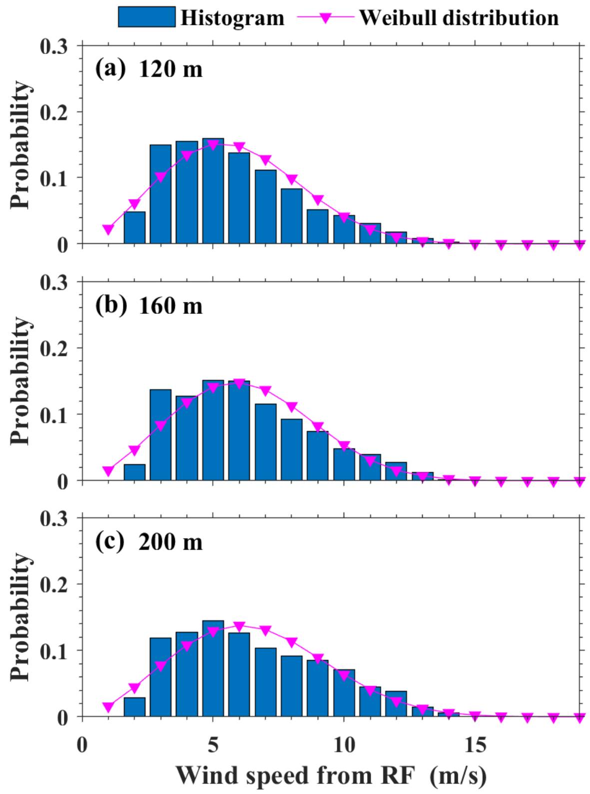

Figure 8. Probability distribution and Weibull distribution of (a) WS_{120}, (b) WS_{160} and (c) WS_{200}~~WS_{120} at the eight stations~~ from 1 May 2018 to 31 August 2020. The blue bar and pink lines represent occurrence probability and Weibull distributions, respectively.

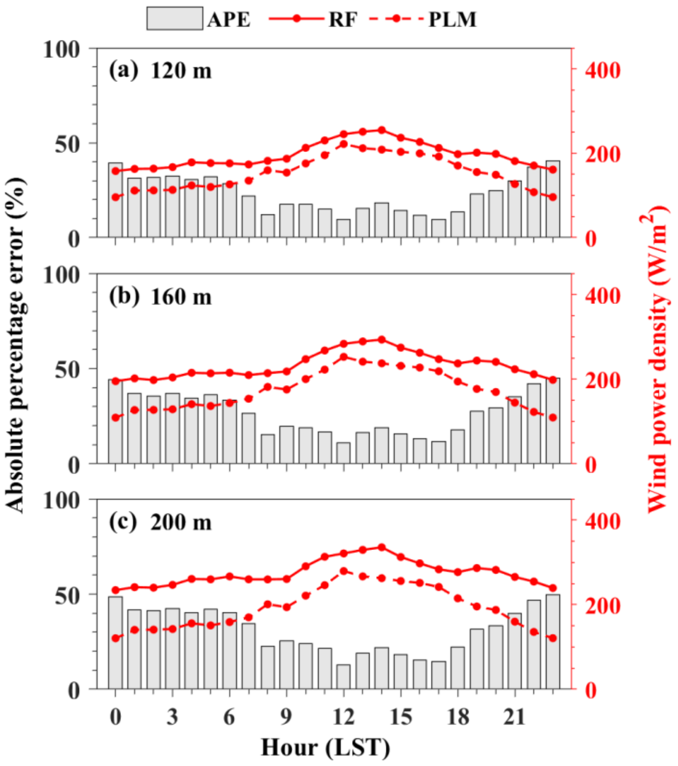

Figure 9. Diurnal variation of the ~~WS120 and~~ wind power density (WPD) ~~for the eight RWP stations~~at (a) 120 m, (b) 160 m and (c) 200 m ~~as shown in Figure 1~~. ~~The blue and red lines denote the mean wind speed and wind power density, respectively.~~The red solid and dotted lines represent the WPD from RF and PLM, respectively. The gray bar represents the absolute percentage error (APE) of WPD between RF and PLM.

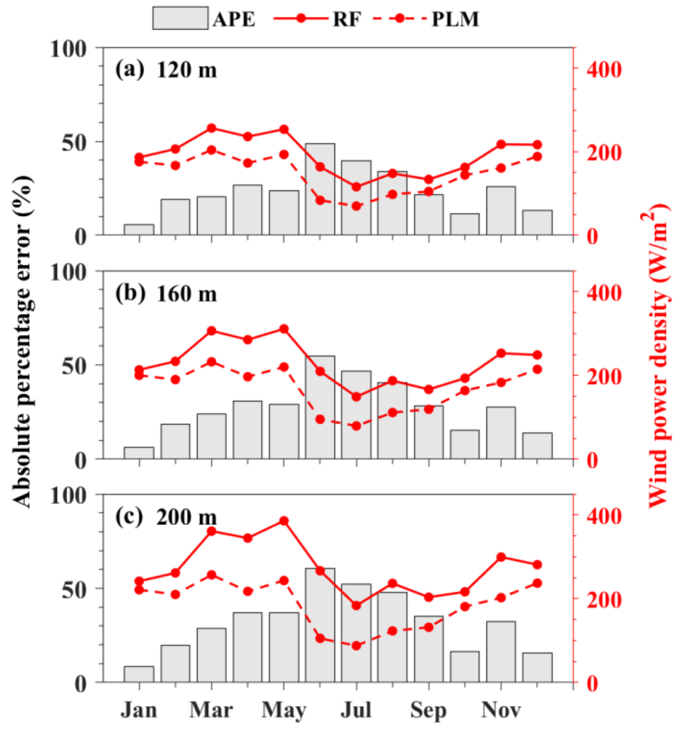

Fig. 10. Similar to Fig. 9, but for the monthly variation.