Extending wind profile beyond the surface layer by combining physical and machine learning approaches

Boming Liu[1], Xin Ma[1*], Jianping Guo[2*], Renqiang Wen[3], Hui Li[1], Shikuan Jin[1], Yingying Ma[1], Xiaoran Guo[2], and Wei Gong[1,4]

[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

[3]CTG Science and Technology Research Institute, China Three Gorges Corporation, 101100, Beijing

[4]Wuhan Institute of Quantum Technology, Wuhan 430206, China

Correspondence to: Dr./Prof. Jianping Guo (Email: jpguocams@gmail.com) and Xin Ma (maxinwhu@whu.edu.cn)

Abstract: Accurate estimation of the wind profile, especially in the lowest few hundred meters of the atmosphere, is of great significance for weather, climate and renewable energy sector. Nevertheless, the Monin–Obukhov similarity theory fails above the surface layer over the heterogeneous underlying surface, resulting in an unreliable wind profile obtained from the conventional extrapolation methods. To solve this problem, we propose a novel method that combines the power law method (PLM) with the random forest (RF) algorithm to extend wind profiles beyond the surface layer, called the ~~Phy-RF~~PLM-RF method. The underlying principle is to treat the wind profile as a power law distribution in the vertical direction, in which the power law exponent (α) is determined by the ~~Phy-RF~~PLM-RF model. First, the ~~Phy-RF~~PLM-RF model is constructed based on the atmosphere sounding data at 119 radiosonde (RS) stations across China and in conjunction with other data such as surface wind speed, land cover type, surface roughness, friction velocity, geographical location, and meteorological parameters from June 2020 to May 2021. Afterwards, the performance of the ~~Phy-RF~~PLM-RF, PLM and RF methods over China are evaluated by comparing them with RS observations. Overall, the wind speed at 100 m of the ~~Phy-RF~~PLM-RF model exhibits high consistency with RS measurements, with a determination coefficient~~correlation coefficient~~ (R^2) of 0.87~~0.93~~ and a root mean squared error (RMSE) of 0.92 m s^{-1}. By contrast, the R^2 and RMSE of wind speed results from PLM (RF) method are 0.75~~0.87~~ (0.83~~0.91~~) and 1.37 (1.04) m s^{-1}, respectively. This indicates that the estimates from the ~~Phy-RF~~PLM-RF method are much closer to observations than those from the PLM and RF methods. Moreover, the RMSE of the wind profiles estimated by the ~~Phy-RF~~PLM-RF model is relatively larger at highlands, while is small in plains. The result indicates that the performance of the ~~Phy-RF~~PLM-RF model is affected by the terrain factor. Finally, the ~~Phy-RF~~PLM-RF model is applied to three atmospheric radiation measurement sites for independent validation, and the wind profiles estimated by the ~~Phy-RF~~PLM-RF model are found consistent with the Doppler wind ~~l~~Lidar observations. This confirms that the ~~Phy-RF~~PLM-RF model has good applicability. These findings have great implications for the weather, climate and renewable energy sector.

1. Introduction

The atmospheric wind field is a critical factor in the transportation of water vapor and matter, influencing weather forecasting and climate change (Stoffelen et al., 2005; 2006). The wind profile~~,~~ is a ~~key~~crucial parameter for measuring the atmospheric wind field, which is related to~~including~~ turbulent mixing, convective transport, and material diffusion in the atmosphere (Solanki et al., 2022; Stoffelen et al., 2020). Particularly in the lowest few hundred meters of the atmosphere, the wind profile plays a significant role in evaluating wind energy resources ~~assessment~~and in understanding the interactions between the atmosphere and the land (Gryning et al., 2007; Veers et al., 2019). Therefore, it is crucial to accurately comprehend the spatial distribution and dynamic variation of wind profiles.

Currently, there are multiple methods for observing wind profiles. Atmospheric reanalysis data, such as the fifth generation ECMWF reanalysis (ERA5) which is based on known physical mechanisms combined with the assimilation of vast amount of observational data, has been widely used to derive the spatiotemporal distribution of wind profiles (Laurila et al., 2021; Gualtieri, 2021). ~~Satellite observations, like Aeolus, can provide the horizontal line of sight wind profile observations, in which can be assimilate into atmospheric models to produce global wind profile products Satellite~~The space-borne Atmospheric Laser Doppler Instrument ~~ALADIN~~ on-board the ~~observations, such as those from~~ Aeolus mission~~,~~ can only provide ~~horizontal~~ line-of-sight wind profile data, which ~~that~~ can be further assimilated into atmospheric models to generate global wind profile products (Stoffelen et al., 2006; Guo et al., 2021). Nevertheless, the accuracy of wind profile products from Aeolus and ERA5 within the planetary boundary layer (PBL) requires improvement due to factors such as atmospheric attenuation and turbulence (Straume et al., 2020; Deng et al., 2022). On the other hand, ground-based wind measurements from towers, radar wind or lidar -based wind profilers~~ground based observations like wind tower, wind profile radar, and wind profile lidar,~~ can obtain highly precise wind profile in the PBL at the observation station (Durisic et al., 2012; Wu et al., 2022). However, single site observations cannot acquire wind profile data on a regional or national scale. Therefore, researchers are endeavoring to develop a theoretical model for wind profile to acquire the large-scale PBL wind profiles.

The wind profile model was initially developed based on~~using~~ the famous Monin–Obukhov similarity theory, which describes the wind profile as functions relying on a stability parameter h/L (Obukhov, 1946; Monin and Obukhov, 1954). The h stands for height while L stands for Obukhov length on the surface. The wind profile model based on similarity theory can be expressed in different forms depending on the varying atmospheric conditions. For the neutral conditions, the wind speed profile model can be simplified as a logarithmic law (Powell et al., 2003; Marusic et al., 2013). For the unstable conditions, the exponential wind speed profile can better describe the wind speed profile in surface layer over homogeneous terrain (Barthelmie et al., 2020). In engineering applications, most studies utilize a power law model to consider the wind profile in the surface layer (Sen et al., 2012; Jung et al., 2021). This can achieve the conversion of surface wind speed to the wind speed at wind turbine hub height. These wind profile models based on the Monin–Obukhov similarity theory have demonstrated effectiveness within the Prandtl layer and the surface layer ~~(approximately 100 m above the ground)~~. The Prandtl layer encompasses the initial tens of meters within the atmospheric boundary layer (Anderson, 2005). The top of surface layer is approximately the height of 100 m above the ground (Veers et al., 2019). Nevertheless, due to factors such as Coriolis parameter, baroclinity, and wind shear, the applicability of the Monin–Obukhov similarity theory breaks down above the surface layer (Optis et al., 2016; Tong et al., 2020). Therefore, extending wind profiles above the surface layer is of significance in applying wind profiles to wind energy assessment and PBL dynamics.

Above the surface layer, the wind profiles are influenced not only by the surface roughness, friction velocity and the atmospheric stability, but also by factors including low-level jets, entrainment processes, and the Coriolis parameter (Gryning et al. 2007; Coleman et al., 2021). To obtain accurate wind profiles above the surface layer, some studies seek to introduce auxiliary variables to account for the influence of these factors. Gryning et al (2007) established a straightforward model that regulates the combined length scale of wind profiles along with their stability correlations. This model is used to calculate wind profiles above the surface. On the other hand, Liu et al. (2022) present an analytical approach based on the Ekman equations and the foundation of the universal potential temperature flux profile. This approach enables one to describe the profiles of wind and turbulent shear stress, which in turn can capture aspects such as the wind veer profile. In addition, some studies use the machine learning (ML) technology to transform surface wind speed and meteorological parameters to wind speeds at different heights. Yu et al. (2022) has devised a transfer method that leverages three ML methods, including the least absolute shrinkage selector operator, random forest (RF) and extreme gradient boost for calculating wind speed at 100 m. Liu et al. (2023) employed the RF model to estimate the wind speed at 120 m, 160 m and 200 m. Nevertheless, the calculation procedure of ML algorithms remains an unexplained process that does not clarify the input parameter's physical significance.

Therefore, it is worth trying to combine ML algorithms with physical models to achieve the inversion of wind profiles above the surface layer.

The present study aims to extend wind profiles beyond the surface layer by combining the physics and machine learning approaches. For this purpose, we attempt to combine the power law method (PLM) with the RF, named ~~Phy-RF~~PLM-RF model, to extend wind profiles beyond the surface layer. The ~~Phy-RF~~PLM-RF model is trained and tested using radiosonde (RS) data and reanalysis gridded meteorological data over China. The performance of the PLM, RF and ~~Phy-RF~~PLM-RF models is also compared. Then, the wind profile generated by the ~~Phy-RF~~PLM-RF model is evaluated against RS observations, followed by independent validation of the model at atmospheric radiation measurement (ARM) sites. The results of our study have great implications for the weather, climate and renewable energy sector.

2. Materials and Data

2.1 Land cover type data

The land cover type data is derived from~~using~~ the Moderate Resolution Imaging Spectroradiometer (MODIS), a satellite-borne instrument that captures images and measures diverse a wide range of surface properties such as land surface temperature, vegetation cover, and atmospheric aerosols (Friedl et al., 2002). The high spatial resolution of the instrument enables the identification of diverse land features, including forests, urban areas, and agricultural fields, thereby making it an important instrument for the purpose of environmental monitoring and land management (Sulla-Menashe et al., 2018). The MODIS provides two land cover type products: MCD12Q1 and MCD12C1. The~~While~~ MCD12Q1 comprises observation data from different regions, which requires self-splicing. The~~,~~ MCD12C1 is an annual concatenated data containing one image per year. Following the previous study (Liu et al., 2020), the land cover type data here is obtained from the MCD12C1, named "MCD12C1.A2021001.061.2022217040006". Fig~~ure .~~1 displays the ~~dominant land cover types'~~ geographic distribution of dominant land cover types in China. ~~The colorbars from bottom to top represent Water, Evergreen Needleleaf Forest, Evergreen Broadleaf Forest, Deciduous Needleleaf forest, Deciduous Broadleaf Forest, Mixed forest, Closed shrublands, Open shrublands, Woody savannas, Savannas, Grasslands, Permanent, Croplands, Urban and built up, Cropland/Natural vegetation mosaic, Snow and ice, and Barren or sparsely vegetated, respectively.~~ The~~se MODIS products~~ land cover type data help us to determine the power law exponent (α).

2.2 Radiosonde measurements

The L-band RS can measure the profiles of atmospheric temperature, pressure, humidity, wind direction and wind speed in-situ. It is taken at 1-min intervals starting from the ground surface up to approximately 30 km above ground level (Guo et al., 2016). The RS observations are conducted at 119 observation stations in China, which is shown in Fig.1. The RS are launched twice per day at around 08:00 and 20:00 local time (LT). Here, the wind speed profiles from RS measurement at 119 stations are obtained as reference value (National Meteorological Science Data Center, 2023). The RS collection time is from 1 June 2020 to 30 May 2021. In addition, the drift of the RS during its ascent has been investigated, as illustrated in Fig. S1. The RS and observation stations' coordinates drifting distance within a height of 0.5 km is less than 0.5 km. This indicates that the drift of RS will not impact obtaining wind profiles in surface layer.

2.3 ERA5 data

The ERA5 is a fifth-generation reanalysis dataset that offers a range of atmospheric parameters, such as temperature, humidity, ~~pressure~~pressure, and radiation (Hersbach et al., 2020). Following previous study (Liu et al., 2023), nine surface parameters have been obtained in this study, 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). These parameters are processed into grid data with a 0.25 × 0.25 size and an hourly time resolution. Based on the longitude and latitude information of the RS and ARM stations, those parameters in the corresponding grid are obtained accordingly. These data are also collected for the period ~~spanning~~ from 1 June 2020 to 30 May 2021.

2.4 ARM data

The ARM user facility is established by the U.S. Department of Energy (Lubin et al., 2020; Zhang et al., 2022). It sets up observation stations and instruments globally for atmospheric observation experiments, making publicly assessable the atmospheric observations, including temperature, wind, radiation, and cloud properties (Liu et al., 2022). The wind profile data from the Doppler wind ~~l~~Lidars deployed at the Eastern North Atlantic (ENA), North Slope of Alaska (NSA), and Southern Great Plains (SGP) stations are collected to independently compare with the proposed method. Statistical parameters such as coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE) are used to quantify the comparison results. Fig.~~urure~~ure ~~e~~S2 presents the geographic locations and land cover types of the three lidar stations. ~~It can be seen that~~ ENA is situated on an Atlantic ~~Ocean Island ocean~~Ocean Island Ocean as its primary land cover, NSA is situated on Alaska's north coast with grassland as its ground cover, and SGP is located in the Great Plains in the central United States where grassland is also the dominant land cover. The Doppler wind ~~l~~Lidar observations covers the period from 1 June 2020 to 30 May 2021. Moreover, these wind profiling measurements are processed as hourly averages correspond with other data.

3. Methods

3.1 Power law method

The PLM assumes that wind speed increases exponentially with height (Hellman et al. 1914). The wind profile can be calculated based on the surface wind speed (v_0) using the following formula:

$$v_i = v_0 \times \left(\frac{h_i}{h_0}\right)^{\alpha} \tag{1}$$

where v_i represents wind speed at height h_i. The h_0 is the measurement height of v_0. Here, the v_0 is observed by an anemometer at a height of 10 m above the ground. The α is the power law exponent, which varies with land cover type, ~~height~~height, and time (Li et al., 2018).

The α is usually set as a constant (0.14) for the purpose of approximating the wind profile at station where there are no observations or empirical formulas available. Fig~~ures.~~ures 2a and 2c show the ~~root mean squared error (~~RMSE~~)~~ and difference between PLM (α=0.14) results and RS measurements for wind speed at 100 m (WS$_{100}$). The average RMSE and difference over China are 1.49±0.39 and -0.23±0.68 m s^{-1}, respectively. The results indicate that PLM results (α=0.14) underestimate wind profiles at almost a quarter of the sites (Fig. 2c). These results suggest that the estimation of wind profiles based on a constant α value is subject to large errors. Some studies also confirm it (Jung et al., 2021; Liu et al., 2023). Furthermore, other studies demonstrate that the α values differ based on the land cover type due to varying surface roughness (Durisic et al., 2012). An empirical lookup table is summarized with respect to the setting of α, as shown in Table S1. The value of α ranges from 0.1 to 0.4 with increasing surface roughness. Based on the MODIS land cover type dataset, the corresponding value of α can be obtained for all RS sites. Fig~~ures-~~ures 2b and 2d show the RMSE and the difference between the PLM (dynamic α) results and RS measurements. Compared with the PLM (α=0.14), the results of PLM (dynamic α) have improved. However, the results of PLM (dynamic α) are still underestimated at most stations in the Northeast and Inner Mongolia regions.

3.2 Random **F**~~f~~orest model

RF model is a nonlinear fitting algorithm and has been used to calculate wind profiles (Yu et al., 2022; Liu et al., 2023). Here, the RF model is also used to fit the surface parameters to obtain the wind profile. The input variables include surface wind

speed (WS), surface wind direction (WD), land cover type (Type), altitude (Alt), longitude (Lon), latitude (Lat), month (M), hour (H), Char, FSR, FV, DP, Temp, Pres, Rn, LHF, and SHF. The reference value is the wind speed provided by RS. In addition, the parameter tuning of the RF model directly affects the performance and generalization ability of the model (Zhu et al., 2021). The tuning parameter process for the Estimator number and Min Leafsize is shown in Fig. S3. The RMSE is minimum (1.02 m s^{-1}) and R is maximum (0.91) when the Estimator number is 300 and the Min Leafsize is 5. Therefore, the Estimator number and the Min Leafsize is set to 300 and 5 for the RF model, respectively.

3.3 Combining physical and RF model

In this study, we propose a novel method, termed ~~Phy-RF~~PLM-RF that combines PLM~~physical~~ and RF model, to estimate wind profiles. Its principle is to treat the wind profile as a power law distribution in the vertical direction, where the α is fitted by using the RF model. The details for this method go as follows:

3.3.1 Physical constraint

Previous studies have confirmed that the wind profile in the surface layer adheres to a power law distribution (Liu et al., 2023). The primary reason for the error is the uncertainty in the value of α. Therefore, to achieve more accurate results, it is necessary to first analyze the reasons of the error. Fig~~ures~~ures .3a~~3 and 3~~b shows~~display~~ the differences in WS$_{100}$ estimated by PLM (α=0.14) and PLM (dynamic α) relative to RS observations. ~~The color bar indicates the various land cover types.~~ Based on MODIS land cover type data, the 119 RS sites are classified as urban area, woodland, shrubs, grassland, and smooth surface. It is found that regardless of land cover type, the difference in wind speed decreases as the surface wind speed increases. Similarly, the differences between assumed α and observed α at 100 m decrease with increasing surface wind speed (Fig~~s~~. 3c~~. and 3~~d). These results indicate that there is a relationship between the error of the PLM results and the surface wind speed. This may be due to the limited influence of surface friction on the wind profile. When the wind speed within the PBL is low, factors such as surface friction and Coriolis force complicate the vertical distribution of the wind profiles, leading to the low surface wind speed and large errors in the PLM (Wang et al., 2023). On the contrary ~~the other hand~~, when the wind speed within the PBL is high, the effect of surface friction can be neglected to some extent. This results in the real wind profile being closer to the power law distribution, thereby reducing the error of the PLM results.

To quantify the effect of surface wind speed on α, the α bias (assumed value minus observed value) at 100 m is examined as a function of surface wind speed under different land cover types is investigated, as shown in Fig.4. The gray dots and black lines indicate the sample points and the logarithmic curve, respectively. The coefficients of determination between the surface wind speed and the difference of α on all types, urban area, woodland, shrubs, grassland, and smooth surface are 0.92, 0.97, 0.94, 0.97, 0.93 and 0.84, respectively. It indicates that there is a good correlation between the surface wind speed and difference of α. Therefore, the correction factor of α ($\Delta\alpha$) can be defined statistically based on the land cover type and surface wind speed. The correction functions of α for different land cover types are also plotted in Fig. 4. For each sample, the $\Delta\alpha$ can be calculated by the correction functions, and then entered the fitting of the RF model as a physical constraint to improve the accuracy. In addition, the α bias as a function of surface wind speed at different heights is also investigated, as shown in Fig. S4. At 50, 100, 150, 200, 250 and 300 m, the coefficients of determination between the surface wind speed and the difference of α are larger than 0.9. This indicates that the $\Delta\alpha$ can be constructed by the surface wind speed to improve the inversion accuracy of the wind speed at high-altitude.

3.3.2 Model construction

For the ~~Phy-RF~~PLM-RF model, the wind profile is considered as a power law distribution, and the α is fitted by the RF model. The inputs include $\Delta\alpha$, WS, WD, Type, Alt, Lon, Lat, M, H, Char, FSR, FV, DP, Temp, Pres, Rn, LHF, and SHF. The reference value is the α calculated from RS observations. The tuning parameter evolution for the ~~Phy-RF~~PLM-RF model is shown in

Fig. S5. The RMSE reaches a minimum (0.91 m s^{-1}) and R reaches a maximum (0.93) when the Estimator number is 500 and Min Leafsize is 5. Therefore, the Estimator number and Min Leafsize are set to 500 and 5, respectively.

To comprehend the model's physical meaning, the importance analysis of the inputs is performed for the RF and ~~Phy-RF~~PLM-RF models, as shown in Fig. 5. The relevant features that can affect the accuracy of the model accuracy are marked with red bars. For the RF model, the relevant features are WS, Type, SHF, FV, WD and FSR. The importance of WS, Type and SHF is greater than other features. WS is the surface wind speed. Type is the value of α based on the land cover type. From the perspective of physical meaning, the RF model calculates wind profiles through complex fitting methods based on surface wind speed and meteorological conditions. In contrast, for the ~~Phy-RF~~PLM-RF model, Δα, FV, SHF, Type, WS, FSR and Temp are both the relevant features. The importance of Δα is the largest, but the importance of Type and WS are ranked fourth and fifth. In addition, the importance of FV ranks second. FV is used to calculate the way the wind changes with height in the lowest levels of the atmosphere (Liu et al., 2023). These results indicate that the ~~Phy-RF~~PLM-RF model calculates the way wind speed changes in the vertical direction. In addition, the SHF and FSR are both relevant features in the construction of the RF and ~~Phy-RF~~PLM-RF models. This indicates that surface roughness and solar radiation are factors that need to consider in the calculation of wind profiles.

3.4 Sensitivity analysis

The average values and standard deviations of the difference~~variance~~ between the assumed α and the observed α are illustrated for the primary input features (Fig. 6). Green, blue and red represent the PLM, RF, and ~~Phy-RF~~PLM-RF models, respectively. The differences in deviations for ~~Phy-RF~~PLM-RF models decrease slightly with increasing surface wind speed. Moreover, the means and deviations of differences for the ~~Phy-RF~~PLM-RF model are relatively stable and do not vary with the land cover types. These results indicate that both the RF and ~~Phy-RF~~PLM-RF models exhibit good generalization across different land cover types and surface wind speeds. This is due to that fact that the RF model considers random perturbations in the sample space to improve generalization ability (Breiman, 2001). In addition, due to the samples are only obtained at 08:00 and 20:00 LT, it notes that whether the performance of the ~~Phy-RF~~PLM-RF model is affected by time. The RS observation stations are geographically distributed in several time zones, but they are all observed at the same time. This means that although the recording time of the RS measurements is 08:00 and 20:00 LT, the training and test samples contain observations data from multiple time periods. Therefore, Rn is used as a measure of time to investigate the applicability of the methods (Fig. 6c). For three methods, the means of the difference are relatively ~~stable~~stable, and the standard deviations of the difference decrease slightly as Rn increases. This indicates that the generalization of the ~~Phy-RF~~PLM-RF model within the sample is reliable. However, the Rn in China at noon can reach 1.5 ~ 2*10^6 J/m^2, which exceeds the upper limit of the input values in the current sample. It indicates that the generalization of the ~~Phy-RF~~PLM-RF model at noon time cannot be proven based on the existing training and test samples. Therefore, ~~it~~we ~~have to~~must ~~can only~~ rely on the Doppler wind lidar~~LiDAR~~ ~~data~~observations from the ARM sites for comparison ~~–~~to evaluate the performance of the ~~Phy-RF~~PLM-RF model at noon. Specific comparisons will be discussed in Sect. 4.4.

4. Results and discussion

In this section, the performances of PLM, RF and ~~Phy-RF~~PLM-RF models are firstly compared by conducting intercomparison analyses. The wind profiles calculated by the ~~Phy-RF~~PLM-RF model are then evaluated by comparing with the RS observations. Finally, the ~~Phy-RF~~PLM-RF model is applied to three ARM sites for independent validation.

4.1 Intercomparison of different methods~~.~~

Fig~~ure.~~ 7 displays the scatter plot between the estimated WS$_{100}$ and the observed WS$_{100}$ for three methods at different times. Overall, the R^2 (RMSE) of WS$_{100}$ from PLM, RF and ~~Phy-RF~~PLM-RF under all times ~~are~~is 0.75~~87~~ (1.37 m s^{-1}), 0.83~~91~~ (1.04

m s^{-1}) and 0.~~87~~93 (0.92 m s^{-1}), respectively. The accuracy of the RF and ~~Phy-RF~~PLM-RF models is better than that of the PLM. For the PLM, most of the estimated WS$_{100}$ are underestimated when the observed wind speed~~value~~ is high. This is because the PLM relies on the exponential relationship to calculate the WS$_{100}$. However, the wind profile is affected by turbulence, surface friction and other factors (Tieleman 1992; Solanki et al., 2022). The exponential law based on constants is unable to obtain the WS$_{100}$ with high accuracy. In contrast, the performance of the RF and ~~Phy-RF~~PLM-RF models improved significantly. The RF and ~~Phy-RF~~PLM-RF models consider more environmental factors, such as SHF and FV, in the inversion process. They improve the accuracy of the model by considering the effects of surface friction and surface radiation flux on the wind profiles. Briefly, these two methods rely on a dynamic α to invert the wind profiles. Each site uses the α that varies with environmental factors, resulting in improved accuracy of inversion ~~accuracy~~. In particular, for the ~~Phy-RF~~PLM-RF model, the ~~α-~~correction function of α can be used to obtain a value of α closer to the observation, resulting in the highest R^2 (0.~~87~~94) and the lowest MAE~~RMSE~~ (~~1.00~~0.60 m s^{-1}). In addition, the MAE~~RMSE~~ of WS$_{100}$ from PLM, RF and ~~Phy-RF~~PLM-RF at 08:00 (20:00) LT are 1.03~~1.39~~ (1.01~~1.35~~), 0.79~~1.05~~ (0.77~~1.03~~) and 0.60~~0.94~~ (0.60~~0.91~~) m s^{-1}, respectively. The comparison results at 08:00 and 20:00 LT are also show that the performance of ~~Phy-RF~~PLM-RF is the best, followed by RF; last are PLM.

Figure 8 shows the R^2, MAE~~difference~~ and RMSE between the estimated WS$_{100}$ and the observed WS$_{100}$ for three methods under different months. ~~For each method,~~ ~~T~~the R^2 is relatively consistent between months, irrespective of methods (Fig. 8a). ~~Overall, the mean difference of the WS$_{100}$ from three methods is close to 0, but the quantile range of the difference is not consistent.~~ For PLM, the monthly mean MAE ~~quantile ranges of the difference~~ are higher~~wider~~ during the cold season (October–April) than during the warm season (June–September). This is because the wind speed variations are more complex during the cold season. The large-scale synoptic systems have a relatively high frequency of occurrence during the cold season (Liu et al., 2019). Compared with PLM, the RF and ~~Phy-RF~~PLM-RF models have the stable accuracy over the 12 months, i.e., the difference between the months is relatively small. ~~In addition, the quantile range of the Phy-RF~~PLM-RF ~~model is narrower than that of the other models. The average difference for the WS$_{100}$~~The monthly mean MAE of ~~Phy-RF~~PLM-RF model does not show significant seasonal differences. It indicates that the ~~Phy-RF~~PLM-RF model is not affected by seasonal variation. This is because that the RF models are data-driven (Zhu et al., 2021; Ma et al., 2021). After correcting the α based on the RF model, the ~~Phy-RF~~PLM-RF model can effectively overcome the influence of seasonal factors. Figure ~~.~~8c~~b~~ shows that the WS$_{100}$ from ~~Phy-RF~~PLM-RF model has a smaller RMSE among the months, and the RMSE is relatively stable over the 12 months. The results indicated that the ~~Phy-RF~~PLM-RF model outperforms both PLM and RF in terms of accuracy and stability. Therefore, the ~~Phy-RF~~PLM-RF model may be a more suitable choice for estimating wind profiles in China than either RF or PLM.

4.2 ~~Evaluation the w~~Wind speed evaluation of the ~~Phy-RF~~PLM-RF model

Figure~~.~~ 9 shows the spatial distributions of the mean wind speed from ERA5 (color shaded) and ~~Phy-RF~~PLM-RF model (colored dots) at 100 m for different periods. In general, the mean ~~wind speeds~~WS$_{100}$ of ERA5 and ~~Phy-RF~~PLM-RF model are generally similar. About the seasonal variation, the WS$_{100}$ is low in summer and fall and high in spring and winter. This is due to the frequently large-scale synoptic systems in the cold season (Liu et al., 2019). From the perspective of spatial distribution, the WS$_{100}$ is the highest in Inner Mongolia and Northeast China, followed by coastal areas, and the lowest in inland areas. There are two reasons for the high wind speed in Inner Mongolia and Northeast China. One is that the climate in these areas is dry and cold, especially in winter. The low temperature and high air density lead to the formation of a strong pressure gradient (Liu et al., 2019). When the pressure gradient is large, ~~cyclonic~~cyclonic, and anticyclonic weather will occur, resulting in higher wind speed. Another reason is that these areas are susceptible to the influence of the Siberian monsoon and warm currents from the Pacific (Yu et al., 2016). This monsoon will cause an increase in wind speed as it passes through Inner Mongolia and Northeast China. In addition, ~~the spatial distributions of wind speed estimated by the Phy-RF~~PLM-RF ~~model at different heights~~the comparisons between the WS$_{100}$ from ERA5 and from PLM-RF model for different periods are shown in

7

Fig. S6. Although the output of the PLM-RF model has a good correlation with the WS_{100} from ERA5, there exist~~are~~ still some differences. Most of the WS_{100} from the PLM-RF model are greater than that of ERA5 when the wind speed is high. This is because the $\Delta\alpha$ is introduced in the PLM-RF model, which makes the model tend to produce ~~output~~ large~~higher~~ output values~~results~~.~~In the vertical direction, the wind speed gradually increases with height. Especially at 200 to 300 m, the wind speed is significantly higher than the WS_{100}. This is because the surface wind profile is affected by the friction velocity and surface roughness (Gryning et al., 2007). The obstructions such as trees, buildings, and hills cause turbulence and reduce the wind speed (Solanki et al., 2022).~~

Figure 10 shows the spatial distributions of the difference between the estimated wind speed and the RS observation for the ~~Phy-RF~~PLM-RF model at different heights. At 50 and 100 m, most sites (more than 90%) show a mean difference of less than 0.2 m s^{-1}, with an overall mean difference of -0.02±0.02 m s^{-1} and -0.15±0.05 m s^{-1}, respectively. In contrast, above 100m, the average differences are negative at almost all sites. The mean differences for all sites at 150, 200, 250 and 300 m are -0.19±0.08, -0.24±0.10, -0.24±0.10, and -0.25±0.11 m s^{-1}, respectively. Compared to the results of PLM (Fig. 2c and 2d), the accuracy of the wind speed in the ~~Phy-RF~~PLM-RF model has improved. Overall, the wind speed estimated by the ~~Phy-RF~~PLM-RF model is slightly underestimated compared to the observed values. Moreover, the average difference gradually increases with the increasing height. This is because the wind profile above the surface layer is not logarithmic but increases faster in response to the reduction in surface friction force (Gryning et al., 2007; Liu et al., 2023). The ~~mean~~ MAE and~~ RMSE and MAE between the estimated and observed wind speeds at different heights can be seen in Fig. S7 and S8. These results also confirm~~s~~ that the performance of the ~~Phy-RF~~PLM-RF model decreases with increasing height ~~(Fig. S7)~~. This is because the wind profile above the surface layer is affected by the influence of the low-level jets, entrainment processes, and the Coriolis parameter (Coleman et al., 2021). In addition, the spatial distributions of RMSE and MAE indicate ~~is similar, indicating~~ that the performance of the PLM-RF model may be influenced by the terrain ~~factor~~.~~The introduction of $\Delta\alpha$ aimed to enhance the performance of the Phy-RFPLM-RF model, but the results indicate that the Phy-RFPLM-RF model slightly underestimates at high altitudes.~~

4.3 Effect of terrain

To evaluate the effect of terrain factor on the performance of ~~Phy-RF~~PLM-RF model, the plain terrain is defined, which the topographic relief is less than 50 m within a radius of 5 km around the observation station. The RS sites are divided into two categories: plains (marked by red dots) and highlands (marked by black dots), as show in Fig. S9~~8~~. Fig~~ure.~~ 11a shows the mean α observed from RS at different height. Blue and red boxes represent the results under plains and highland areas, respectively. The mean α in highlands is greater than that in plains. It indicates that the variation of wind profiles in highlands is more complex than that in plains. Previous studies have also shown that the valley winds and low-level jets can complicate the wind profiles in the PBL (Solanki et al., 2021; Wang et al., 2023). Fig~~ure~~ 11b shows the difference between the estimated and observed wind speed for ~~Phy-RF~~PLM-RF model at different height. The difference in highlands is obviously larger than that in plains. Moreover, the similar phenomena were also found in the results of RMSE (Fig. 11c). The RMSE in highlands is relatively large, while is relatively small in plains. This may be due to differences in terrain. The terrain in plains is mainly flat, while the terrain in highlands is mainly mountainous (Chen et al., 2016). The wind profile is not only affected by factors such as surface friction and solar radiation, but also constrained by the terrain (Panofsky et al., 1964; Jung et al., 2021). In the construction of the ~~Phy-RF~~PLM-RF model, the influence of the terrain factor was not considered, resulting in a higher RMSE of the ~~Phy-RF~~PLM-RF model on the highlands.

4.4 Independent validation

Figure 12 displays the vertical wind speed distribution using different methods at three ARM sites. At the NSA site, the wind profiles calculated by the ~~Phy-RF~~PLM-RF model are similar to the observed values at 08:00 and 20:00 LT, but are slightly overestimated at 14:00 LT. The performance of the ~~Phy-RF~~PLM-RF model at 14:00 LT is inferior to that of PLM (Fig. 12c). Similarly, this phenomenon occurs at sites of SGP as well. The results of the ~~Phy-RF~~PLM-RF model are significantly overestimated at 14:00 LT. These results indicate that the performance of the ~~Phy-RF~~PLM-RF model is influenced by hourly variations. However, to our surprise, the result of the PLM-RF model is very consistent with the observations at the ENA site, even at 14:00 LT (Fig. 12g). This may be due to the differences in land cover types between the sites. Although the PLM-RF model has some overestimation at 14:00 LT, the comparisons at other times indicate that the wind profiles of the PLM-RF model are still ~~similar to~~like the observed results (Fig. 12a, 12e and 12i). ~~T~~Overall, ~~t~~he PLM-RF model's wind profiles exhibit greater proximity to the observed values when compared to the results generated by PLM at three ARM sites.

To further evaluate the performance of the PLM-RF model, the diurnal variations of R^2, MAE and RMSE between the WS_{100} calculated by PLM-RF model and the WS_{100} observed by Doppler wind ~~l~~Lidar are ~~investigated~~shown in Figure 13. At SGP site, the R^2 is higher at nighttime and lower at daytime. These results confirm that the performance of the PLM-RF model at SGP site is influenced by diurnal variations. This is because that ~~t~~The generalization of the RF algorithm depends on the training and test samples (Zhu et al., 2021). As mentioned in Sect. 3.4, the training and test samples of the ~~Phy-RF~~PLM-RF model do not actually contain any in situ measurements~~data~~ from the period 11:00 to 15:00 LT~~time periods~~. This means that the ~~Phy-RF~~PLM-RF model has no generalization at noon, resulting in poor accuracy of the ~~Phy-RF~~PLM-RF model during daytime~~at 14:00 LT~~. On the contrary, the performance of the PLM-RF model is stable at NSA and ENA sites. This is due to ~~However, to our surprise, the result of the Phy-RFPLM-RF model is very consistent with the observations at the ENA site, even at 14:00 LT (Fig. 12g). This may be due to the differences in land cover types between the sites.~~ the~~NSA and~~ SGP site~~s~~ are located over~~on~~ land, with significant diurnal variations in wind speed. The wind speed at daytime is relatively low, even lower than the estimated value of PLM. In contrast, the ENA site is located on the island, and the diurnal variation of wind speed is not significant. The wind speed throughout the day is higher than the estimated value of PLM. For the ~~Phy-~~PLM-RF model, since the training data is mainly composed of relatively high wind speed at nighttime, the model exhibits significant overestimation correction. The model can accurately calculate wind speed when the actual value is larger than the estimated value of PLM, while it will significantly overestimate the actual value if it is lower than the estimated PLM. ~~Although the Phy-RFPLM-RF model has some overestimation at 14:00 LT, the comparisons at other times indicate that the wind profiles of the Phy-RFPLM-RF model are still similar to the observed results (Fig. 12a, 12e and 12i). Overall, the Phy-RFPLM-RF model's wind profiles exhibit greater proximity to the observed values when compared to the results generated by PLM at three ARM sites.~~

Overall~~Additionally~~, the wind speed results retrieved by the ~~Phy-RF~~PLM-RF model are consistent with the Doppler wind ~~l~~Lidar's ~~observations~~measurements at different heights. These results indicate that the ~~Phy-RF~~PLM-RF model has good spatial applicability and can be used to obtain the wind profiles on different land cover types.

5. Summary and conclusions

The traditional wind profile model was constructed based on the Monin–Obukhov similarity theory. As a result, the wind profile based on the similarity theory is only effective within the surface layer. To address this challenge, this study proposes a ~~Phy-RF~~PLM-RF method that combines the traditional PLM with the RF algorithm to extend the wind profiles beyond the surface layer.

The reasons for the errors in PLM above the surface layer are first analyzed. The result indicates that the error of PLM is mainly attributable to the α setting. This is because the wind profile above the surface is affected by factors such as surface

roughness, friction velocity, low-level jets, and Coriolis parameter, resulting in the complexity of the α. Moreover, the surface wind speed has a certain impact on the variation of α. At the height of 50, 100, 150, 200, 250 and 300 m, the coefficients of determination between the surface wind speed and the difference of α are greater than 0.9. It may be due to the limited influence of surface friction on the wind profile. When the PBL wind is high, the effect of surface friction can be neglected to some extent, resulting in the real wind profile being closer to the power law distribution. Based on this physical constraint, the ~~Phy-RF~~PLM-RF method considers the wind profile as a power law distribution in the vertical direction, and the α at different heights is fitted by the RF model to calculate the wind profile.

The performance of PLM, RF and ~~Phy-RF~~PLM-RF methods are then compared based on the RS observations over China from 1 June 2020 to 30 May 2021. The R^2 (~~MAE~~RMSE) of the WS_{100} from PLM, RF and ~~Phy-RF~~PLM-RF models were 0.75~~87~~ (1.0~~2~~37 m s−1), 0.83~~91~~ (0.78~~1.04~~ m s−1) and 0.87~~93~~ (0.60~~92~~ m s−1), respectively. It shows that the ~~Phy-RF~~PLM-RF model has better accuracy and stability compared to PLM and RF. Especially under high wind speed events, the output of PLM is significantly low, while the PLM-RF model can effectively correct this underestimation. The ~~Phy-RF~~PLM-RF model can be understood as the PLM based on dynamic α. The RF model is used to adjust α at different heights based on factors such as surface wind speed, land cover type, and meteorological parameters to achieve high-precision wind profile inversion.

~~Afterwards, the wind profiles calculated by the Phy-RFPLM-RF model are then evaluated from a temporal and spatial perspective. From a temporal distribution perspective, the average difference for the Phy-RFPLM-RF model does not exhibit significant seasonal variations. It indicates that~~Overall, the advantages of the PLM-RF model ~~is~~is that it can provide more accurate wind profiles than the PLM, especially when the actual wind speed is high. Moreover, the ~~Phy-RF~~PLM-RF model is not affected by seasonal variation. This is because the RF model is data driven. The training sample of the ~~Phy-RF~~PLM-RF model contains enough samples from four seasons. The PLM-RF model is recommended for areas with high wind speeds, such as coastal areas. ~~In contrast,~~The limitations of the ~~Phy-RF~~PLM-RF model is that the performance of PLM-RF model is affected by diurnal variation and terrain. The generalization of the RF model depends on whether the training samples contain sufficient sample inputs. The training samples of the ~~Phy-RF~~PLM-RF model do not contain in situ measurements~~data~~ from the time period −11:~~:~~00 to 15:~~:~~00 LT ~~time periods~~, resulting in relatively poor accuracy during this period. ~~Therefore, it is not recommended to use the Phy-RFPLM-RF model during 11:00 to 15:00 LT before adding observation data to retrain the model. From a spatial distribution perspective,~~ Similarly, the RMSE of the wind profiles is relatively larger at highland areas~~.~~. ~~It indicates that the Phy-RFPLM-RF model is affected by the terrain factor.~~ ~~which~~This is likely due to the fact that the influence of terrain~~factor~~ was not considered in the construction of the ~~Phy-RF~~PLM-RF model. Therefore, it is not recommended to use the PLM-RF model for ~~at~~the ~~time~~ period from −11:~~:~~00 to 15:~~:~~00 LT ~~and~~ over highland areas before ~~add~~including observation data to cons~~retrain the model.~~~~Overall, the Phy-RFPLM-RF model can provide more accurate wind profiles than the PLM and RF models.~~

Our study extends the wind profile beyond the surface layer by combining physical and ML approaches, which has great implications for the weather, climate and renewable energy sector. However, due to the limitations in data size and terrain factors, the performance of the PLM-RF model above water surfaces is uncertain.~~of data time and terrain factors, it should pay attention to the application scenarios of the Phy-RFPLM-RF model.~~ In the future, the global RS observation data will be used to train and test the PLM-RF model, and evaluate its performance on a global scale~~topographic conditions need more attention and deserve further investigation~~.

Data Availability~~——————~~

The output data and codes used in this paper can be provided for non-commercial research purposes upon reasonable request (Jianping Guo, email: jpguocams@gmail.com). The RS data can be downloaded from

Acknowledgments

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.

Competing interests

The authors declare that they have no conflict of interest.

References

Atmospheric Radiation Measurement (ARM) user facility data: Doppler Lidar Horizontal Wind Profiles, ARM [data set], https://adc.arm.gov/discovery/#/results/instrument_class_code::dlprof-wind (last access: 18 September 2023), 2023.

Anderson, J. D.: Ludwig Prandtl's boundary layer. Physics today, 58(12), 42-48, https://doi.org/10.1063/1.2169443, 2005.

Breiman, L.: Random forests in: Machine Learning, 45: 5–32, https://doi.org/10.1023/a:1010933404324, 2001.

Barthelmie, R. J., Shepherd, T. J., Aird, J. A., and Pryor, S. C.: Power and wind shear implications of large wind turbine scenarios in the US Central Plains. Energies, 13(16), 4269, https://doi.org/10.3390/en13164269, 2020.

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, https://doi.org/10.55599/ejssm.v16i3.80, 2021.

Chen, M., Gong, Y., Li, Y., Lu, D., and Zhang, H.: Population distribution and urbanization on both sides of the Hu Huanyong Line: Answering the Premier's question. Journal of Geographical Sciences, 26, 1593-1610, https://doi.org/10.1007/s11442-016-1346-4, 2016.

Deng, X., He, D., Zhang, G., Zhu, S., Dai, R., Jin, X., and Li, X.: Comparison of horizontal wind observed by wind profiler radars with ERA5 reanalysis data in Anhui, China. Theoretical and Applied Climatology, 150(3), 1745-1760, https://doi.org/10.1007/s00704-022-04247-6, 2022.

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

ECMWF: ERA5 hourly data on single levels from 1959 to present, ECMWF [data set], https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview (last access:7 March 2023), 2023.

435   Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X. Y., Muchoney, D., Strahler, A. H., and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results. Remote sensing of Environment, 83(1-2), 287-302, https://doi.org/10.1016/s0034-4257(02)00078-0, 2002.

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,
440   https://doi.org/10.1007/s10546-007-9166-9, 2007.

Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., and Zhai, P.: The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data. Atmospheric Chemistry and Physics, 16(20), 13309-13319, https://doi.org/10.5194/acp-16-13309-2016, 2016.

Guo, J., Liu, B., Gong, W., Shi, L., Zhang, Y., Ma, Y., Xu, X.: First comparison of wind observations from ESA's satellite
445   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, 2021.

Gualtieri, G.: Reliability of ERA5 reanalysis data for wind resource assessment: a comparison against tall towers. Energies, 14(14), 4169, https://doi.org/10.3390/en14144169, 2021.

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

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, https://doi.org/10.1002/qj.3803, 2020.

Jung, C., and D. Schindler.: The role of the power law exponent in wind energy assessment: A global analysis. International Journal of Energy Research 45.6, 8484-8496, https://doi.org/10.1002/er.6382, 2021.

455   Lubin, D., Zhang, D., Silber, I., Scott, R. C., Kalogeras, P., Battaglia, A., and Vogelmann, A. M.: AWARE: The atmospheric radiation measurement (ARM) west Antarctic radiation experiment. Bulletin of the American Meteorological Society, 101(7), E1069-E1091, https://doi.org/10.1175/bams-d-18-0278.1, 2020.

Li, J. L. and 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.

460   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, Appl. Energ., 236, 867–876, https://doi.org/10.1016/j.apenergy.2018.12.056, 2019.

Liu, L., and Stevens, R. J.: Vertical structure of conventionally neutral atmospheric boundary layers. Proceedings of the National Academy of Sciences, 119(22), e2119369119, https://doi.org/10.1073/pnas.2119369119, 2022.

Liu, B., Guo, J., Gong, W., Shi, L., Zhang, Y., and Ma, Y.: Characteristics and performance of wind profiles as observed by
465   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. Atmospheric Research, 271, 106121, https://doi.org/10.1016/j.atmosres.2022.106121, 2022.

470   Liu, B., Ma, X., Guo, J., Li, H., Jin, S., Ma, Y., and Gong, W.: Estimating hub-height wind speed based on a machine learning algorithm: implications for wind energy assessment, Atmos. Chem. Phys., 23, 3181–3193, https://doi.org/10.5194/acp-23-3181-2023, 2023.

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, https://doi.org/10.1002/joc.6957, 2021.

Luo, B.; Yang, J.; Song, S.; Shi, S.; Gong, W.; Wang, A.; Du, L.: Target Classification of Similar Spatial Characteristics in Complex Urban Areas by Using Multispectral LiDAR. Remote Sens., 14, 238. https://doi.org/10.3390/rs14010238, 2022.

Monin, A. S., and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere. Contrib. Geophys. Inst. Acad. Sci. USSR, 151(163), e187, 1954.

Marusic, I., Monty, J. P., Hultmark, M., and Smits, A. J.: On the logarithmic region in wall turbulence. Journal of Fluid Mechanics, 716, R3, https://doi.org/10.1017/jfm.2012.511, 2013.

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, https://doi.org/10.1175/jas-d-16-0186.1, 2017.

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 ($PM_{2.5}$) 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.

National Meteorological Science Data Center: Radiosonde observation data, China Meteorological Administration [data set], http://www.nmic.cn/data/cdcdetail/dataCode/B.0011.0001C.html (last access: 7 March 2023), 2023.

Obukhov, A. M.: Turbulence in an atmosphere with inhomogeneous temperature, Tr. Inst. Teor. Geofis. Akad. Nauk. SSSR, 1, 95-115, 1946.

Optis, M., Monahan, A., and Bosveld, F. C.: Limitations and breakdown of Monin–Obukhov similarity theory for wind profile extrapolation under stable stratification. Wind Energy, 19(6), 1053-1072, https://doi.org/10.1002/we.1883, 2016.

Panofsky, H. A., and Townsend, A. A.: Change of terrain roughness and the wind profile. Quarterly Journal of the Royal Meteorological Society, 90(384), 147-155, https://doi.org/10.1002/qj.49709038404, 1964.

Pei, Z., Han, G., Mao, H., Chen, C., Shi, T., Yang, K., and Gong, W.: Improving quantification of methane point source emissions from imaging spectroscopy. Remote Sensing of Environment, 295, 113652, https://doi.org/10.1016/j.rse.2023.113652, 2023.

Powell, M. D., Vickery, P. J., and Reinhold, T. A.: Reduced drag coefficient for high wind speeds in tropical cyclones. Nature, 422(6929), 279-283, https://doi.org/10.1038/nature01481, 2003.

Pérez, I. A., García, M. A., Sánchez, M. L., and De Torre, B.: Analysis and parameterisation of wind profiles in the low atmosphere. Solar Energy, 78(6), 809-821, https://doi.org/10.1016/j.solener.2004.08.024, 2005.

Sulla-Menashe, D., and Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. Usgs: Reston, Va, Usa, 1, 18, 2018.

Straume, A. G., Rennie, M., Isaksen, L., de Kloe, J., and Parinello, T.: ESA's space-based Doppler wind lidar mission Aeolus– First wind and aerosol product assessment results. In EPJ Web of Conferences (Vol. 237, p. 01007). EDP Sciences, https://doi.org/10.1051/epjconf/202023701007, 2020.

Stoffelen, A., Pailleux, J., Källén, E., Vaughan, J. M., Isaksen, L., Flamant, P., and Ingmann, P.: The atmospheric dynamics mission for global wind field measurement. Bulletin of the American Meteorological Society, 86(1), 73-88, https://doi.org/10.1175/bams-86-1-73, 2005.

Stoffelen, A., Marseille, G. J., Bouttier, F., Vasiljevic, D., De Haan, S., and Cardinali, C.: ADM‑Aeolus Doppler wind lidar observing system simulation experiment. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 132(619), 1927-1947, https://doi.org/10.1256/qj.05.83, 2006.

Stoffelen, A., Benedetti, A., Borde, R., Dabas, A., Flamant, P., Forsythe, M., and Vaughan, M.: Wind profile satellite observation requirements and capabilities. Bulletin of the American Meteorological Society, 101(11), E2005-E2021, https://doi.org/10.1175/bams-d-18-0202.1, 2020.

Sen Z, Altunkaynak A, Erdik T.: Wind velocity vertical extrapolation by extended power law. Adv Meteorol, 1–6. https://doi.org/10.1155/2012/178623, 2012.

Solanki, R., Guo, J., Li, J., Singh, N., Guo, X., Han, Y., and Liu, B.: Atmospheric-boundary-layer-height variation over mountainous and urban sites in Beijing as derived from radar wind-profiler measurements. Boundary-Layer Meteorology, 181, 125-144, https://doi.org/10.1007/s10546-021-00639-9, 2021.

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.

Tieleman, H. W.: Wind characteristics in the surface layer over heterogeneous terrain, J. Wind Eng. Ind. Aerod., 41, 329–340, https://doi.org/10.1016/0167-6105(92)90427-c, 1992.

Tong, C., and Ding, M.: Velocity-defect laws, log law and logarithmic friction law in the convective atmospheric boundary layer. Journal of Fluid Mechanics, 883, A36, https://doi.org/10.1017/jfm.2019.898, 2020.

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, eaau2027, https://doi.org/10.3389/fenrg.2020.624646, 2019.

Wu, S., Sun, K., Dai, G., Wang, X., Liu, X., Liu, B., Song, X., Reitebuch, O., Li, R., Yin, J., and Wang, X.: Inter-comparison of wind measurements in the atmospheric boundary layer and the lower troposphere with Aeolus and a ground-based coherent Doppler lidar network over China, Atmos. Meas. Tech., 15, 131–148, https://doi.org/10.5194/amt-15-131-2022, 2022.

Wang, S., Guo, J., Xian, T., Li, N., Meng, D., Li, H., and Cheng, W.: Investigation of low-level supergeostrophic wind and Ekman spiral as observed by a radar wind profiler in Beijing. Frontiers in Environmental Science, 11, 1195750, https://doi.org/10.3389/fenvs.2023.1195750, 2023.

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, Int. J. Climatol., 36, 2173–2188, https://doi.org/10.1002/joc.4485, 2016.

Yu, S., and Vautard, R.: A transfer method to estimate hub-height wind speed from 10 meters wind speed based on machine learning. Renewable and Sustainable Energy Reviews, 169, 112897, https://doi.org/10.1016/j.rser.2022.112897, 2022.

Yang, S., Yang, J., Shi, S., Song, S., Luo, Y., and Du, L.: The rising impact of urbanization-caused CO_2 emissions on terrestrial vegetation. Ecological Indicators, 148, 110079, https://doi.org/10.1016/j.ecolind.2023.110079, 2023.

Zhu, Y., Ma, Y., Liu, B., Xu, X., Jin, S., and Gong, W.: Retrieving the Vertical Distribution of PM2.5 Mass Concentration from Lidar Via a Random Forest Model. IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2021.3102059, 2021.

550   Zhang, D., Comstock, J., and Morris, V.: Comparison of planetary boundary layer height from ceilometer with ARM radiosonde data. Atmospheric Measurement Techniques, 15(16), 4735-4749, https://doi.org/10.5194/amt-15-4735-2022, 2022.

Zhang, Y., Wang, W., He, J., Jin, Z., and Wang, N.: Spatially continuous mapping of hourly ground ozone levels assisted by Himawari-8 short wave radiation products. GIScience and Remote Sensing, 60(1), 2174280, https://doi.org/10.1080/15481603.2023.2174280, 2023.
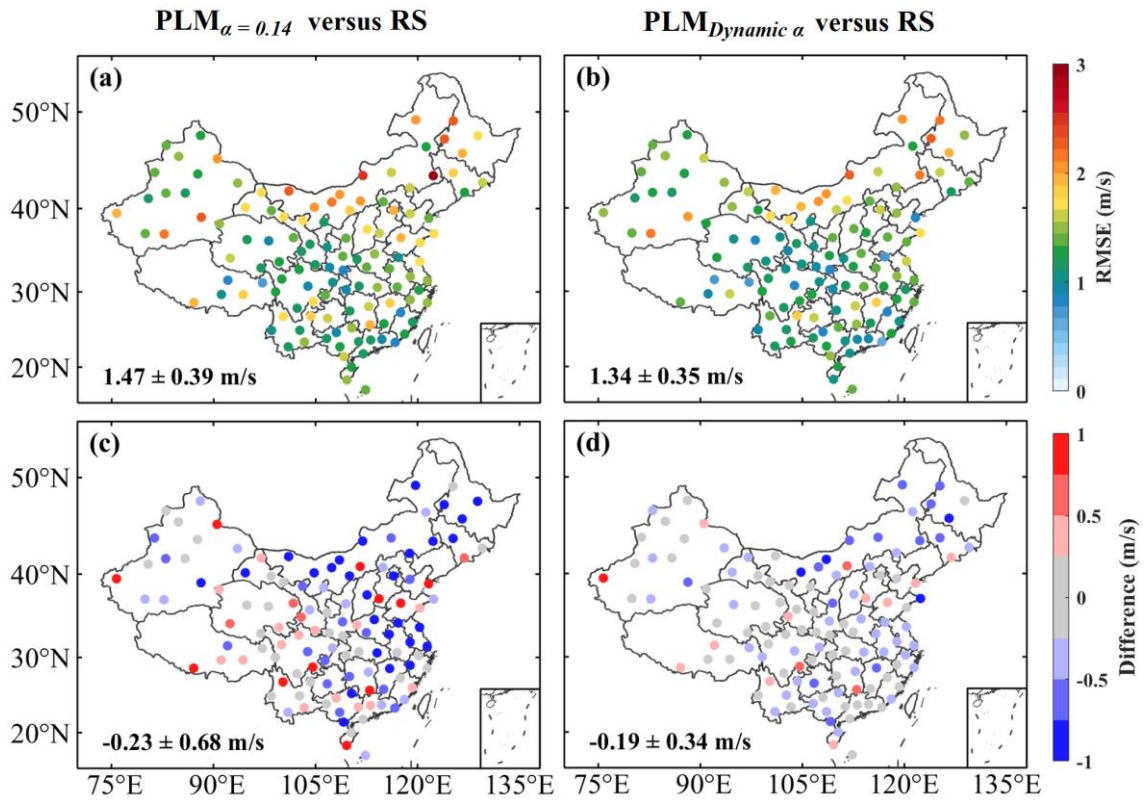
555

Figures:
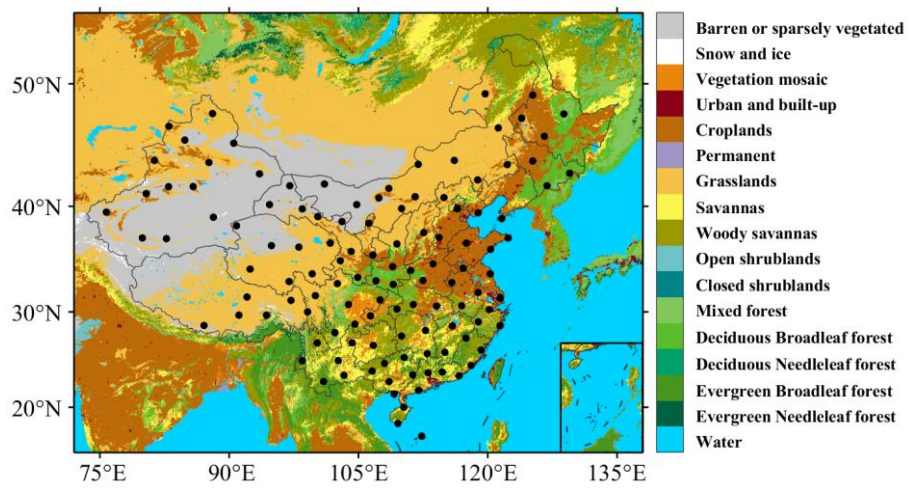
Figure 1. Geographical distribution of the radiosonde stations in China, which is overlaid over surface land cover type (color shading) from the MODIS observations. The color bars represent the different surface land cover types.

560

Figure. 2 The spatial distributions of (a, b) RMSE and (c, d) difference for the WS_{100} estimated by the traditional PLM (constant α of 0.14) and PLM (dynamic α), respectively. Also shown in the bottom left corner are the mean values of RMSE and difference.
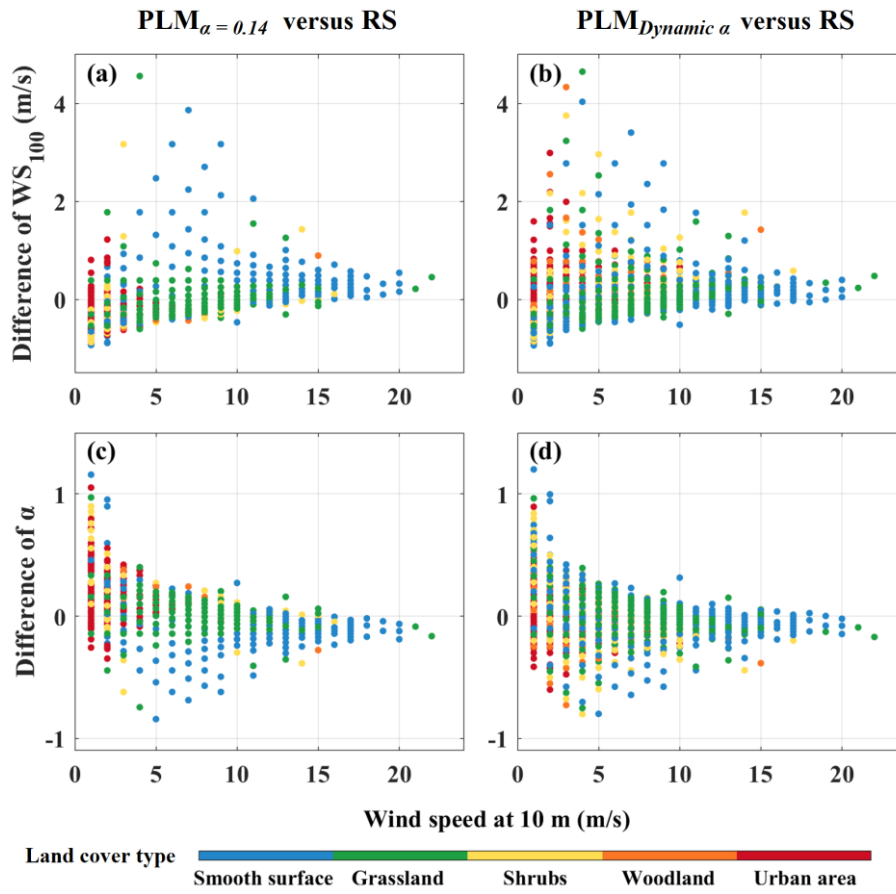
Figure 3. The differences in (a, b) WS_{100} and (c, d) α estimated by PLM (α=0.14) and PLM (dynamic α) relative to RS observations shown as a function of land cover type (color shading). The color bar indicates the various land cover types.
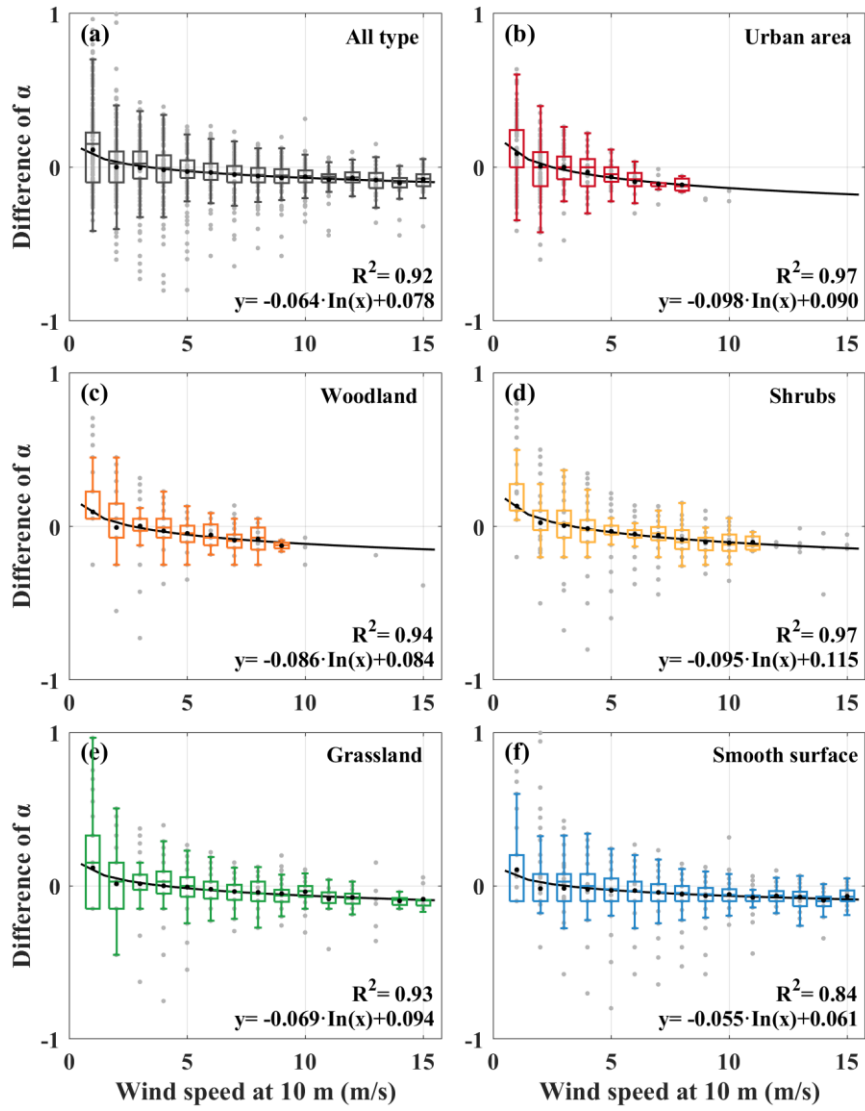
Figure 4. The α bias (assumed value minus observed value) at 100 m as a function of surface wind speed under (a) all types of land cover, (b) urban area, (c) ~~wooldland~~woodland, (d) shrubs, (e) grassland, and (f) smooth surface.
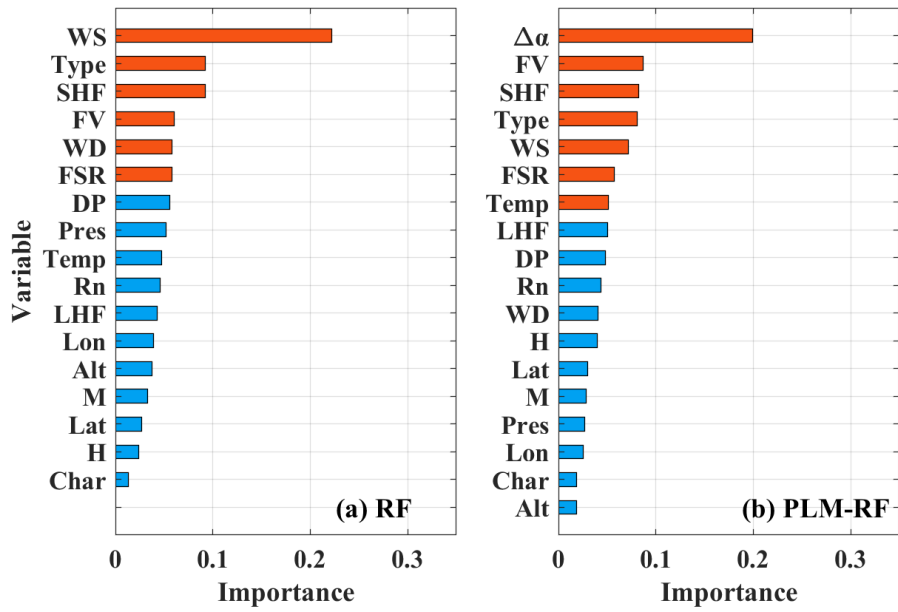
Figure 5. The importance ~~analysis~~ scores of ~~input~~predictor variables for the (a) RF and (b) ~~Phy-RF~~PLM-RF models. WS=surface wind speed; WD=surface wind direction; Type=land cover type; Alt=altitude; Lon=longitude; Lat=latitude; M=month; H=hour; Char=charnock coefficient; FSR=forecast surface roughness; FV=friction velocity; DP=dew point; Temp=temperature; Pres=pressure; Rn=net solar radiation; LHF=latent heat flux; SHF=sensible heat flux and $\Delta\alpha$=correction factor of α.
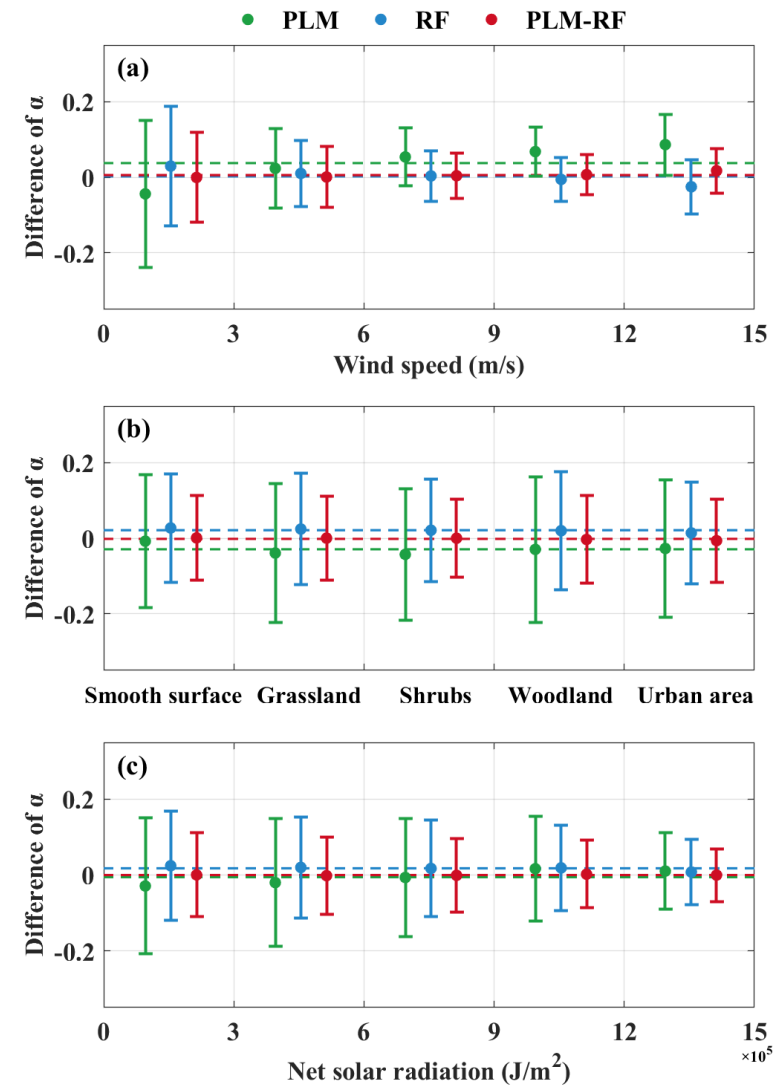
Figure 6. The means and standard deviations of the difference between assumed α and observed α at 100 m as a function of (a) surface wind speeds, (b) land cover type and (c) net solar radiation. Green, blue, and red lines represent the results from the PLM, the RF, and the ~~Phy-RF~~PLM-RF, respectively.
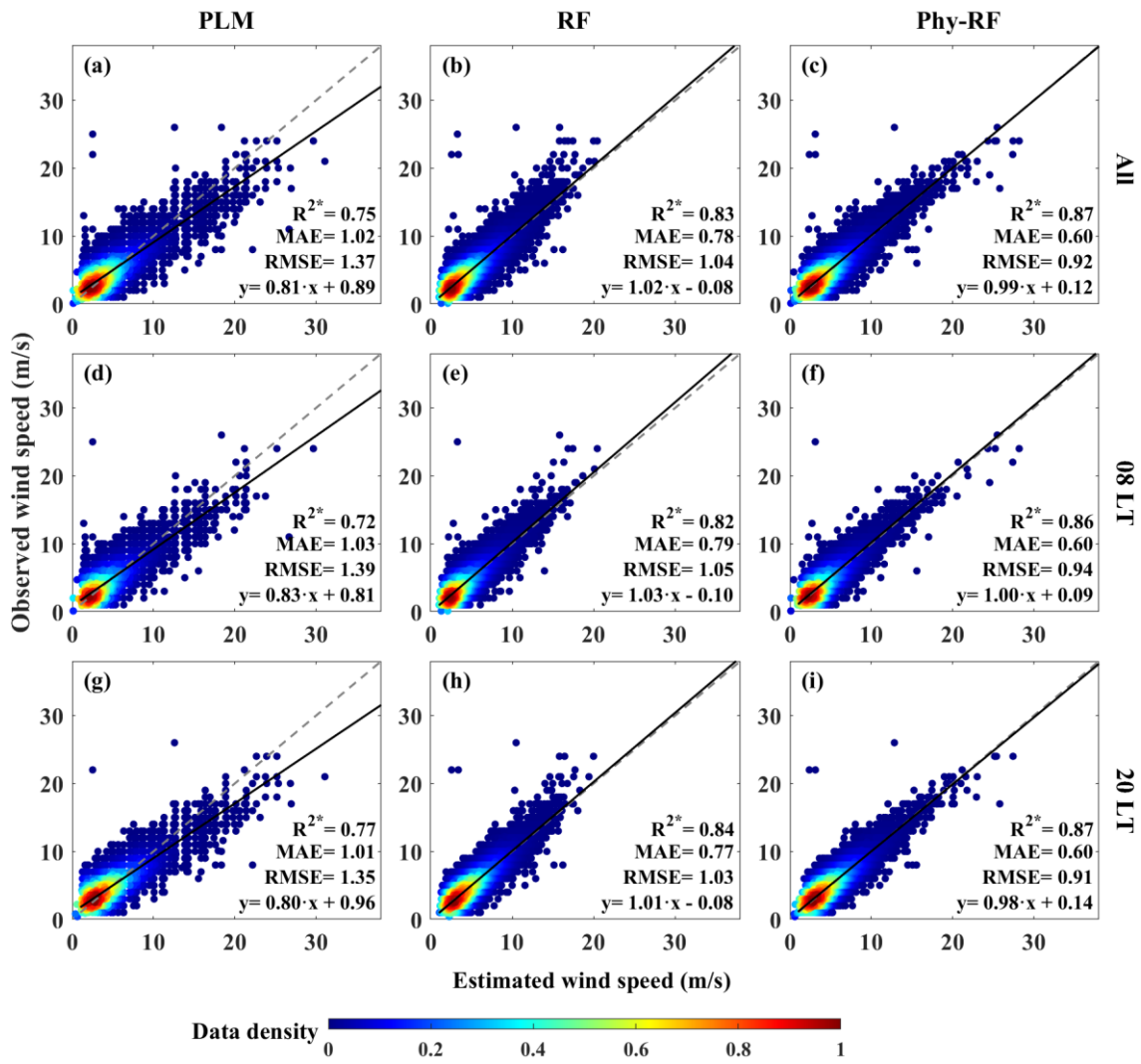
Figure 7. Comparisons between observed WS_{100} and estimated WS_{100} for (a, d, g) the PLM, (b, e, h) RF and (c, f, i) ~~Phy-RF~~PLM-RF models under all time, 0800 LT and 2000 LT. The gray and black lines are the reference and regression lines, respectively. The color bar represents the data density. The asterisk indicates that the correlation coefficient (R) has passed the t test at a confidence level of 95 %.
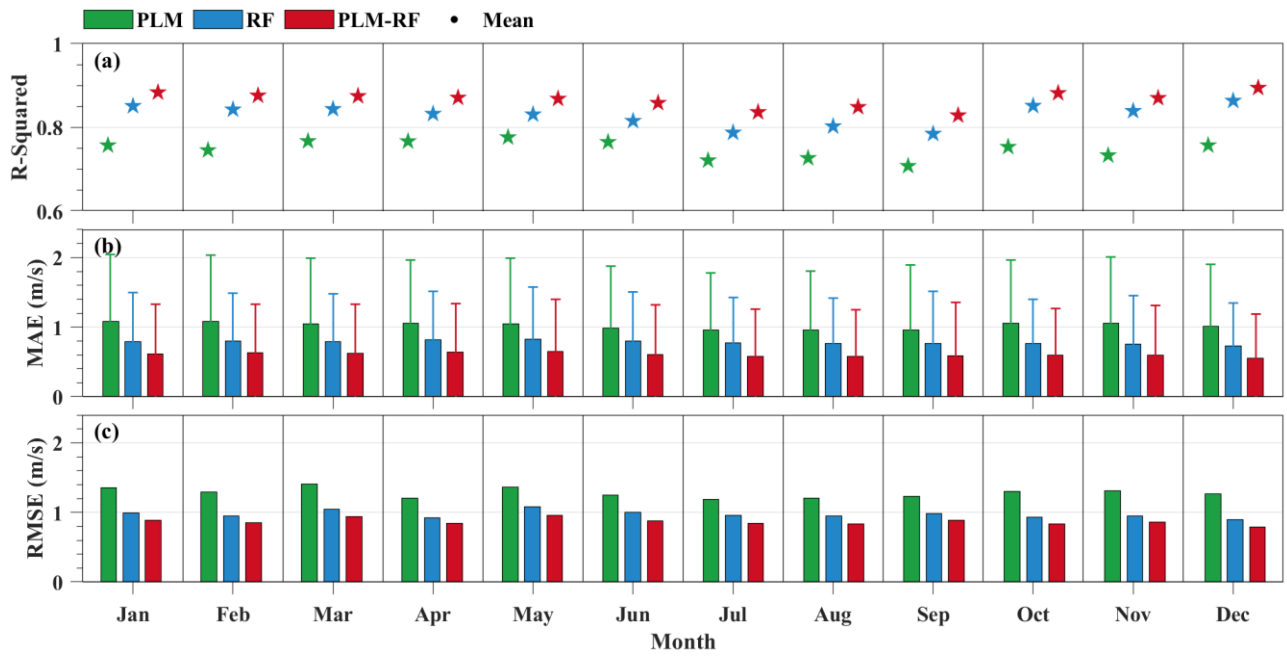
Figure 8. Annual cycle of the ~~difference~~ (a) R^2, (b) MAE and (c) RMSE ~~(b)~~ between the estimated WS_{100} and observed WS_{100} for the PLM, RF and ~~Phy-RF~~PLM-RF models. The green, blue and red colors represent the PLM, RF and ~~Phy-RF~~PLM-RF methods, respectively.
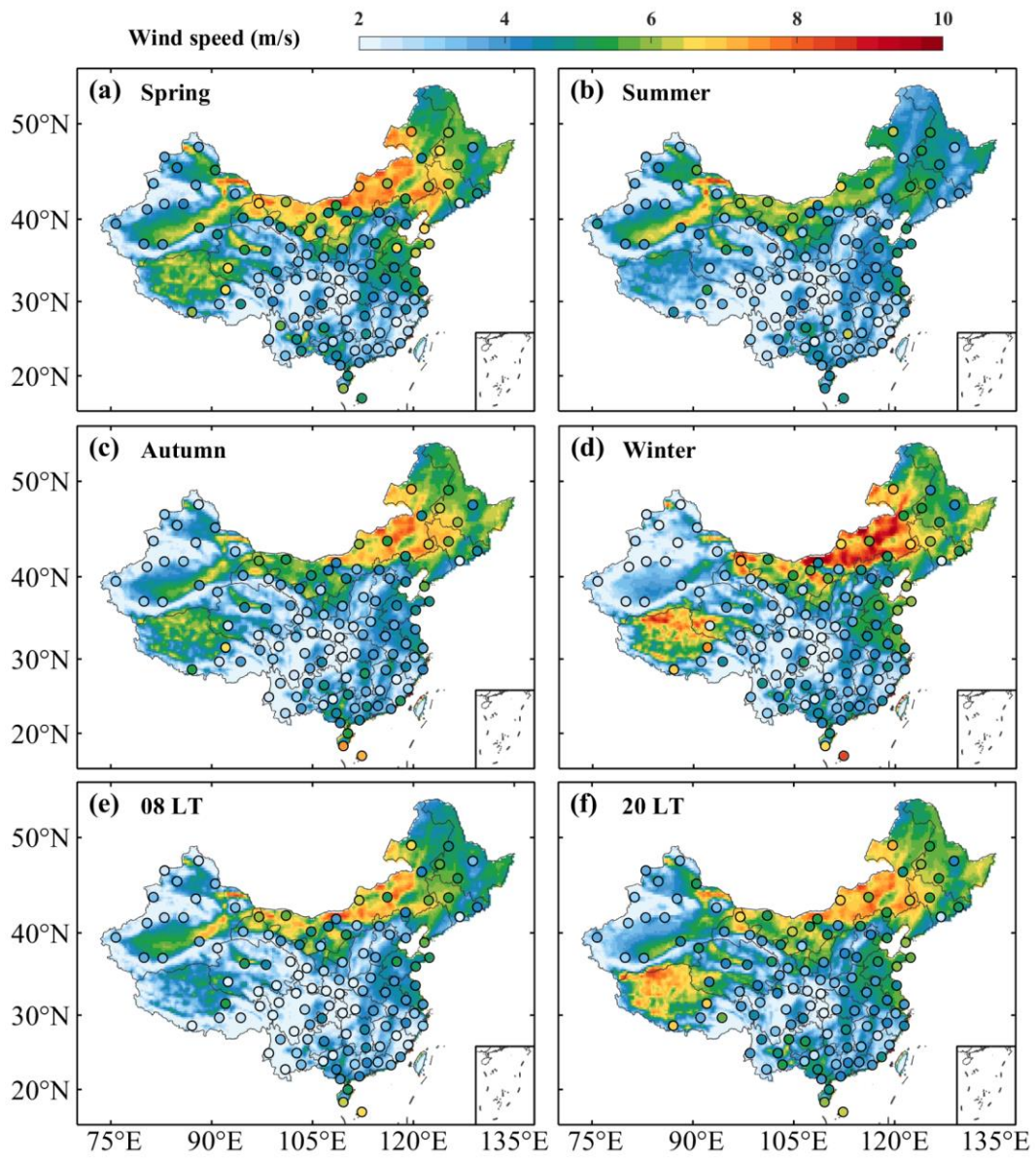
Figure 9. Spatial distributions of the mean WS_{100} from ERA5 (color shaded) and ~~Phy-RF~~PLM-RF model (color dots) in (a) spring, (b) summer, (c) autumn, (d) winter, (e) 08:00 LT, and (f) 20:00 LT.
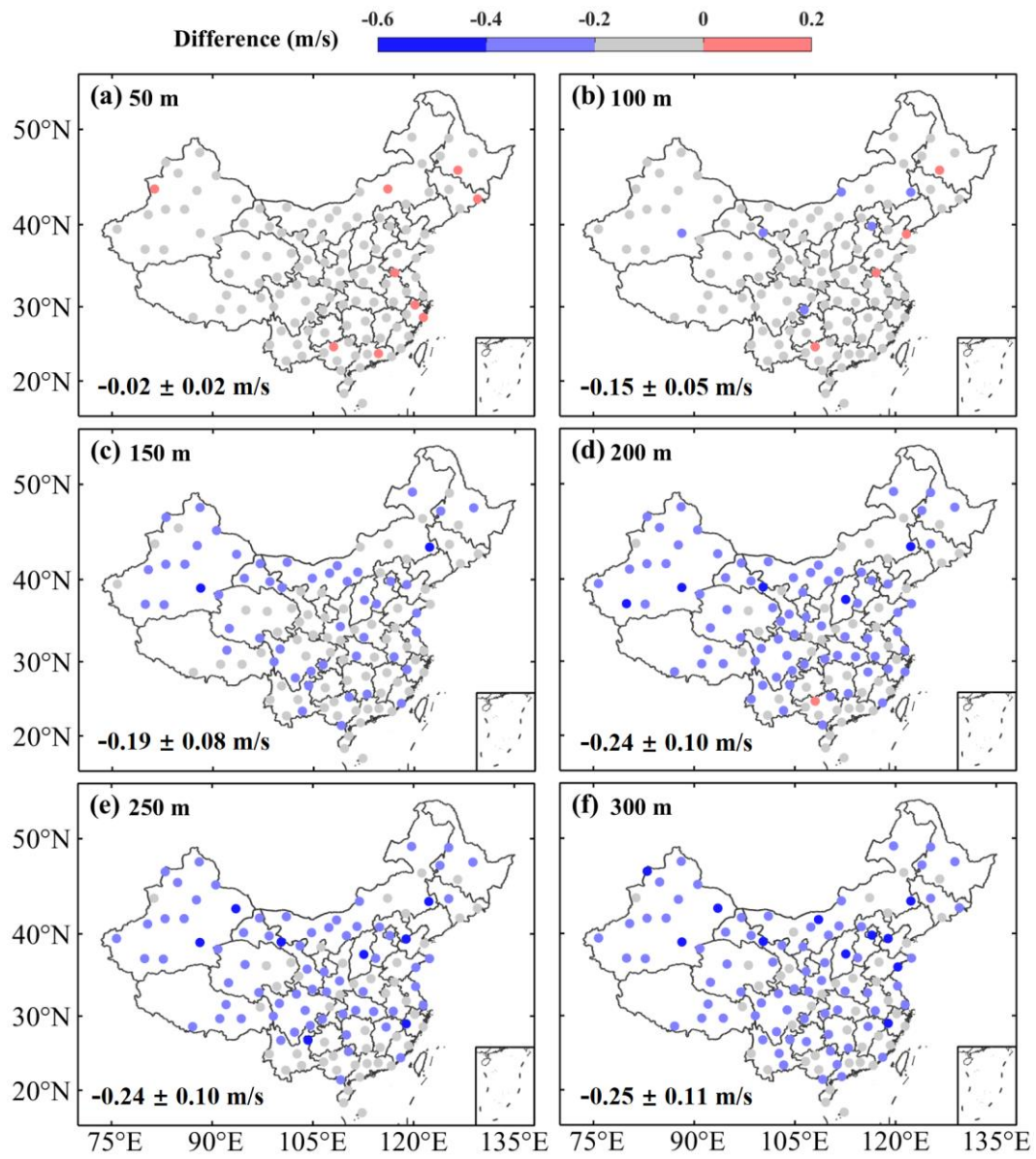
610 **Figure 10**. The spatial distributions of the difference between the estimated wind speed and observed wind speed for the ~~Phy-RF~~PLM-RF model over 120 radiosonde stations in China at different heights: (a) 50 m, (b) 100 m, (c) 150 m, (d) 200 m, (e) 250 m, (f) 300 m.
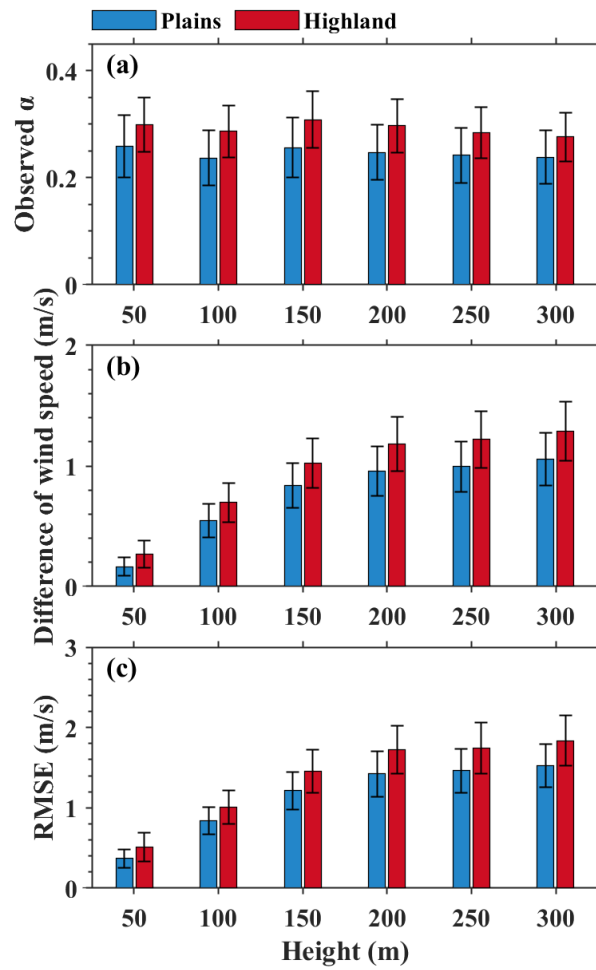
Figure 11. (a) The mean α observed from RS as a function of height. The (b) difference and (c) RMSE between the estimated wind speed and observed wind speed as a function of height. Blue and red boxes represent the results under plains and highland areas, respectively.
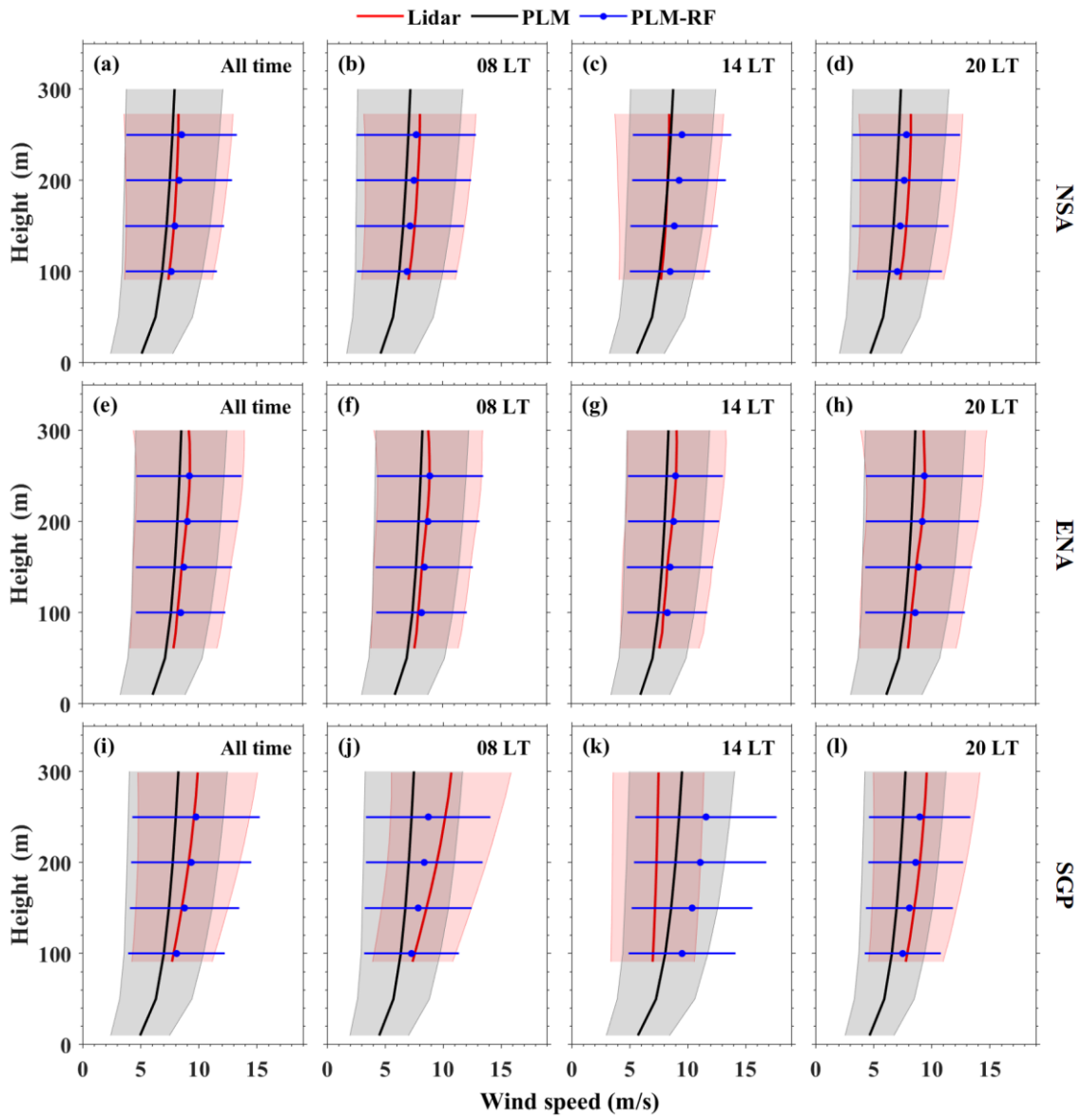
Figure 12. Vertical profiles of the wind speed from different methods at three ARM sites: (a-d) NSA, (e-h) ENA, and (i-l) SGP. Red, black, and blue lines represent mean wind profile from the Doppler wind lLidar, the PLM, and the Phy RFPLM-RF, respectively, and their corresponding color shading areas represent the standard deviation.

625
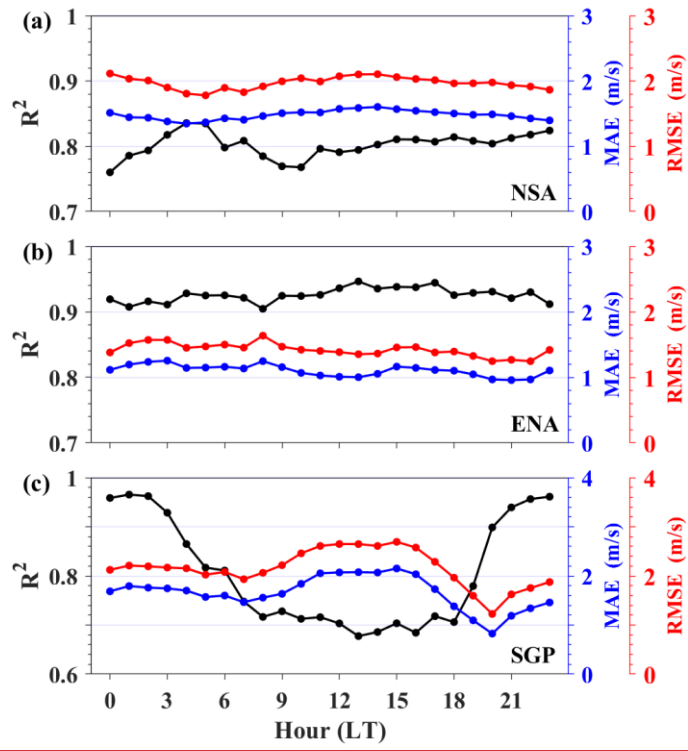
Figure 13. Diurnal variations of R^2 (black curves), MAE (blue lines) and RMSE (red lines) between the WS_{100} calculated by PLM-RF model and the WS_{100} observed by the Doppler wind Lidars at (a) NSA, (b) ENA, and (c) SGP sites.