https://doi.org/10.5194/essd-2026-73
Preprint. Discussion started: 23 March 2026

Global-ABLWind: a global atmospheric boundary layer wind speed profile dataset derived from Aeolus and surface ancillary information

Zhe Tong[1], Boming Liu[1], Xin Ma[1], Jianping Guo[2], Haowei Zhang[3], Haoyu Dong[1], Ge Han[4], Yingying Ma[1], Wei Gong[1,5]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[2]The State Key Laboratory of Severe Weather Meteorological Science and Technology, Chinese Academy of Meteorological Sciences, Beijing 100081, China
[3]Aerospace Information Research Institute, Henan Academy of Sciences, Zhengzhou 450046, China
[4]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[5]Wuhan Institute of Quantum Technology, Wuhan, China

Correspondence to: Boming Liu (liuboming@whu.edu.cn) and Xin Ma (maxinwhu@whu.edu.cn)

Abstract. Accurate wind speed profiles within the atmospheric boundary layer (ABL) are essential for understanding atmospheric processes, climate change, and wind energy assessment. However, existing global ABL wind products lack either sufficient vertical resolution or accuracy, limiting their ability to resolve wind structures throughout the ABL. Here, we propose a physics-constrained machine learning framework designed to reconstruct continuous ABL wind speed profiles by integrating physically interpretable bias-correction mechanisms with dynamical constraints from Aeolus L2C observations. The proposed method enables the reconstruction of high-accuracy wind speed profiles at a vertical resolution of 100 m across the full ABL (0–2 km), overcoming the trade-off between accuracy and vertical resolution that characterizes existing products. Independent validation against RS observations demonstrates that the proposed method achieves high accuracy across all ABL heights. It has an overall correlation coefficient (R) of 0.92 and a root mean square error (RMSE) of 1.94 m s^{-1}, outperforming the original Aeolus L2C product (R = 0.90, RMSE = 2.23 m s^{-1}). Further comparisons at 100 m vertical resolution with the fifth generation ECMWF reanalysis (ERA5) and the power law method confirm the superior accuracy of XGB-Wind, especially in the near-surface layer (0-500 m). Applying the proposed framework to the full Aeolus mission period (from July 2020 to April 2023), we generate a global high-resolution ABL wind speed profile dataset, termed Global-ABLWind. This dataset provides 100 m vertical resolution wind profiles with enhanced accuracy, continuous ABL coverage, and reduced data gaps on a global scale. The dataset is freely available at https://doi.org/10.5281/zenodo.18286457 (Tong et al., 2026) and represents a valuable remote sensing resource for boundary-layer wind studies and wind-related environmental applications.

1 Introduction

The atmospheric boundary layer (ABL) is a critical area for energy, momentum, and material exchange between the Earth's atmosphere and surface (Kalmus et al., 2022; Tong et al., 2024). Its internal wind speed is the core parameter for determining the atmospheric turbulence structure (Liu et al., 2020; Tan et al., 2025) and describing the surface-atmosphere exchange process (Stoffelen et al., 2005, 2020). The wind speed within the ABL directly affects the material and energy exchange between the surface and the free atmosphere, which plays an important role in understanding the variations of atmospheric dynamic field (Li et al., 2022; Peng et al., 2022) and the assessment of wind energy resources (Ramon et al., 2019; Veers et al., 2019). Consequently, accurate ABL wind speed profiles are essential for understanding atmospheric turbulence, pollutant dispersion, and near-surface wind conditions relevant to wind energy applications

Currently, ABL wind speed profiles are obtained primarily from ground-based observations, satellite remote sensing, and reanalysis products (Watson, 2023). Ground-based observations, such as radiosonde (RS) and wind profiler radar, can provide high-accuracy wind profiles but they are limited to the observation site and cannot provide regional large-scale wind profiles (Iwai et al., 2021). Reanalysis products, such as the fifth-generation ECMWF reanalysis (ERA5), combine diverse observations to generate global wind fields at hourly temporal resolution and ~30 km spatial resolution (Hersbach et al., 2020). Nevertheless, near-surface wind estimates from ERA5 are still subject to uncertainties over complex terrain and under strong turbulence conditions (Deng et al., 2022; Pelser et al., 2024). As the only satellite mission capable of directly measuring global wind profiles, the Atmospheric Dynamics Mission Satellite (Aeolus) provides horizontal line-of-sight wind observations that are widely assimilated into numerical models (Stoffelen et al., 2006; Horányi et al., 2015). However, due to atmospheric signal attenuation and surface backscatter contamination, the retrieval performance of Aeolus within the ABL—particularly in the near-surface layer—remains limited (Guo et al., 2021; Abril-Gago et al., 2023). These limitations indicate that current observation systems alone are insufficient to provide accurate and continuous global wind speed profiles throughout the entire ABL, motivating the development of inversion and reconstruction methods to compensate for observational gaps.

To overcome the lack of continuous ABL wind observations, numerous wind profile inversion and extrapolation models have been developed. The power law method (PLM) and logarithm law method are widely used in scientific research and engineering applications (Marusic et al., 2013; Quick et al., 2025). These models are constructed based on the Monin-Obukhov similarity theory, assuming that the wind speed profile changes with height as a power or logarithmic function. However, these wind profile models are only applicable to the lowest several dozen meters within the ABL under specific atmospheric conditions (Gualtieri, 2019; Ghannam and Bou-Zeid, 2021). Under complex atmospheric conditions, the effects of low-level jet, entrainment process and boundary layer height will lead to the failure of the two wind profile models above the surface layer (about 100 m) (Baas et al., 2009; Kent et al., 2018). Liu and Stevens (2022) proposed a model by combining the Ekman equations and the basis function of the universal potential temperature flux profile to characterize wind and turbulent shear stress profiles and hence can capture aspects such as wind veer profile. Meanwhile, some researchers introduced meteorological auxiliary variables to retrieve wind speed profiles beyond the surface layer through machine learning. Optis et al. (2021) and Hatfield et al. (2023) input surface wind speed and multiple meteorological variables into a random forest (RF) model to increase the maximum extrapolated height to 200 and 100 m, respectively. An extrapolation method was developed by Yu and Vautard (2022) with three machine learning methods to estimate 100 m wind speed and found that RF and extreme Gradient Boost (XGBoost) performed better than traditional algorithm. Liu et al. (2024) combined the power law with machine learning (RF) to calculate wind speed beyond the surface layer. Frech et al. (2025) used the RF model to generate wind profile products up to 200 m for the US coast using sea surface wind observations from multiple satellites. Nevertheless, these approaches remain fundamentally limited by the lack of independent constraints in the upper ABL, preventing reliable reconstruction of wind profiles throughout the entire ABL. Near-surface meteorological variables alone are insufficient to constrain wind speed in the upper ABL, reflecting the fundamentally ill-posed nature of ABL wind profile retrieval. In contrast, Aeolus observations provide independent wind information in the upper ABL and lower troposphere, offering a unique opportunity to constrain wind profile retrievals beyond the surface layer (Stoffelen et al., 2020; Rennie et al., 2021). Leveraging Aeolus as an upper-level dynamical constraint may therefore compensate for the ill-posed nature of ABL wind profile inversion based solely on near-surface variables, which provides the possibility for reconstructing high-precision global ABL wind speed profile data.

In this study, we present a new global ABL wind speed profile dataset (Global-ABLWind) by an advanced machine learning model named XGB-Wind that integrates data-driven approaches with physically interpretable constraints. This dataset spans from July 2020 to April 2023, covering the entire ABL from the surface to a height of 2 km worldwide. The performance of the proposed methods and Global-ABLWind is evaluated by global RS data, and systematically compared with Aeolus L2C products, ERA5, and PLM. The dataset generated fills the gap in high-precision global ABL wind speed profile and is of great significance to the study of boundary layer climate, weather and wind energy resources.

2 Data and materials

2.1 Data Collections

Global Aeolus L2C, RS, and ERA5 data were used for this study. According to the principle of time matching, the above data were collected from July 2020 to April 2023. Figure 1 shows the geographical distribution of the RS station and Aeolus ground tracks. The black lines represent the Aeolus ground tracks, and the red dots represent RS stations. L2C wind is used to provide upper-level wind constraints. RS data are used as labels to train the model. ERA5 meteorological variables are applied to calculate the bias correction at the bottom layer and as features to input into the model.

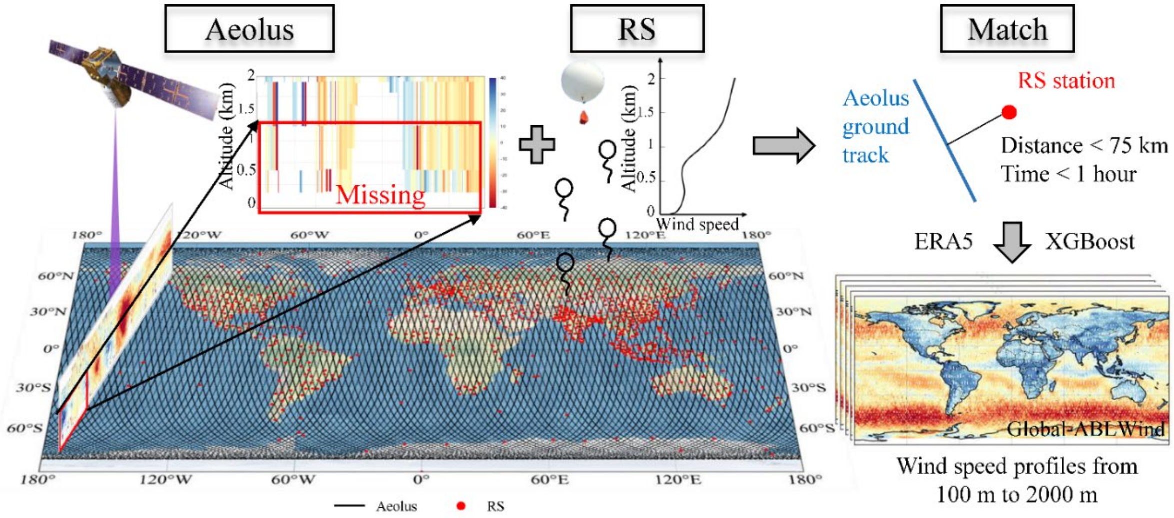


Figure 1. The data collected in this manuscript and their matching diagram. The map below represents the spatial location of RS stations and Aeolus ground tracks. The black lines represent the Aeolus ground tracks, and the red dots represent RS stations.

2.2 Radiosonde measurements

The RS can provide a fine resolution profile of atmospheric parameters. It is deployed on the balloon and measures atmospheric variables such as air pressure, temperature, humidity, wind speed, and wind direction during ascent (Durre et al., 2006). This study uses the Integrated Global Radiosonde Archive (IGRA) data from National Centers for Environmental Information (NCEI), which can be downloaded for free from the NCEI data portal (https://www.ncei.noaa.gov/products/weather-balloon/integrated-global-radiosonde-archive, last access: 07 Jan 2026). These RS data not only come from ground-based observations, but also from ships and mobile platforms, totaling over 2800 globally distributed observation stations (Durre et al., 2018). All data are filtered through a series of quality assurance algorithms, including basic plausibility checks, site elevation rationality and temporal consistency checks, internal consistency checks, repeatability checks, climate-based checks, and data-completeness checks (Durre et al., 2018). In all datasets, approximately half of the observation stations launch two detections per day, with once per day and two or more observations per day accounting for about one-fourth each. The most frequent launch times are 0000 and 1200 UTC.

The RS data from all stations worldwide are collected within the same time range as Aeolus L2C data from July 2020 to April 2023. There are a total of 960 RS stations with valid wind data and geographic information within this time span. Figure 1 displays the geographical locations of these RS stations. Following the established matching principles (Zhang et al., 2016; Guo et al., 2021), the geographic matching threshold is set to 75 km and the time matching threshold is set to 1 hour. After matching, a total of 391 stations have data paired with Aeolus. In addition, the RS observations with missing wind speed data need to be removed. And considering that some stations only provide vertical height information at fixed pressure levels, the height at variable pressure levels is interpolated based on the height at fixed pressure levels and their pressure data.

It is worth noting that the balloons may drift away from the launch site during ascent. Seidel et al. (2011) studied data from 419 sounding stations over a period of 2 years and found that the average drift distance in the middle troposphere was about 5 km. However, this study focuses on within 2 km above the surface so that the drift error can be approximately ignored.

2.3 Aeolus L2C

Aeolus satellite operates in an approximately 320 km sun-synchronous orbit of 90-minute orbital period with a revisit cycle of 7 days (Witschas et al., 2020). The L2C product is a superset of L2B level products, consisting of L2B level data and assimilation related information of L2B wind, providing horizontal wind observations from the ground to the stratosphere at around 30 km (Banyard et al., 2024). ALADIN has two receivers that simultaneously capture the Doppler shift from both the Rayleigh (molecular) and Mie (aerosol/cloud) channels. Due to the difference in the strength of the echo signals, the horizontal resolutions of the Rayleigh channel and Mie channel are 86.4 km and 10 km, respectively(Šavli et al., 2019). Correspondingly, on the vertical scale, the resolution of 24 vertical bins ranges from 0.25 km near the surface to 2 km. By integrating the two inversion channels and two observation types, a total of four wind observations were obtained: Rayleigh-clear, Rayleigh-cloudy, Mie-clear, and Mie-cloudy. Among them, the optimal quality of Rayleigh and Mie measurements is achieved under clear-sky and cloudy conditions, respectively (Baars et al., 2020; Martin et al., 2021). Therefore, Rayleigh-clear and Mie-cloudy observations are the focus of this study. In order to express concisely, Ray wind is used to refer to Rayleigh-clear wind and Mie wind is used to refer to Mie-cloudy wind in the following text. All L2C level data can be accessed and downloaded for free on the ESA's Aeolus portal (https://aeolus-ds.eo.esa.int/oads/access/collection/L2C_Wind_Products, last access: 07 Jan 2026). The validity flags are used to screen L2C wind speeds and filter out wind data that fail assimilation. Only those L2C wind speeds with an assimilation validity flag of 1 are included in this study.

The L2C product itself has the potential to provide wind profiles within the ABL on a global scale. However, it suffers from significant missing and low vertical resolution within the ABL (Guo et al., 2021; Abril-Gago et al., 2023). Statistical analysis reveals that the average missing rate of Ray wind at an altitude of 0-2 km worldwide is 16.4%. Figure 2 shows the missing rate of time and space for Ray wind product. From Fig. 2 (a), it can be seen that the missing rate starts to increase as the satellite operation time increases, which may be due to the decrease in signal-to-noise ratio of the signals caused by the decay in laser energy. Figure 2 (b) illustrates the missing rate of Ray wind product globally. The global missing rate is lower over land than over oceanic regions. The missing rate of the Ray product is relatively high in the prevailing westerly belt areas in high-latitude regions, especially in the oceanic regions. This may be due to frequent bottom clouds or ice crystals interfering with the Rayleigh molecular signals in high-latitude oceanic regions (Baars et al., 2020). In addition, there are areas with abnormally high missing rates on all the west coast of South America, Africa, Mexico, and Australia, which coincide with common cold current regions such as the Peru Current, Benguela Current, Canary Current, California Current, and Western Australia Current. The altitude-longitude and altitude-latitude missing rate profiles are given in Fig. 2 (c) and (d), respectively. It can be seen that the missing rate below 300 m is relatively high, close to 0.3, and the near surface missing is more severe in high-latitude areas. It is evident from the altitude-latitude profile that different height bin settings are used in different latitude intervals. The closer to the ground, the more serious the missing situation, this is because the closer to the ground, the signal is more interfered with by aerosols, resulting in a weaker scattered signal and a lower signal-to-noise ratio (Rennie et al., 2021). The missing rate of time and space for Mie wind product is shown in Fig. S1. The missing rate of the Mie product is more severe than the Ray product, with an average missing rate of 54.9%. This is because the inversion conditions of the Mie channel are more stringent, and the observation signals are strong enough only in thin clouds and thick cloud tops. Overall, Ray and Mie products generally have high missing rates near the surface at 0-2 km.

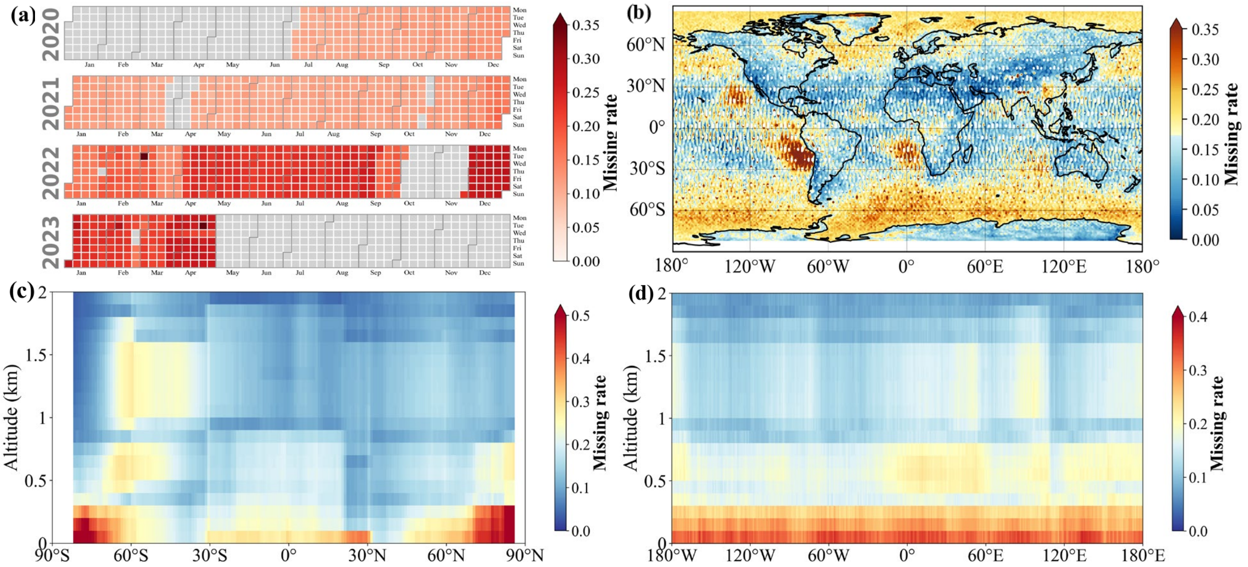


Figure 2. The missing rate of Ray wind below 2 km on a time plane **(a)**, altitude–longitude plane **(b)**, altitude-latitude plane **(c)**, and altitude-longitude plane **(d)** from July 2020 to April 2023.

From the perspective of accuracy, the comparison results between Ray and RS wind are shown in Fig. 3. Overall, the precision of Ray wind is good, with correlation coefficient (R), MAE and RMSE of 0.88, 1.64 m s^{-1} and 2.42 m s^{-1}, respectively. The bias of Ray wind shows a Gaussian distribution, which is concentrated around 0, with an average value of 0.13 ± 2.41 m s^{-1}. This indicates that the difference between the Ray wind and the RS wind is not significant. In order to analyze the performance of Ray winds at different altitudes, the differences between Ray winds and RS winds with central heights falling within each 100 m interval are calculated separately (Fig. 3 (c), (d) and (e)). It can be seen that the Ray winds at the bottom layer of 0-0.3 km are significantly higher than the RS winds, with a maximum bias of up to 1 m s^{-1}. And the RMSE and MAE are also higher in the near-surface layer compared to the high altitude. This all suggests that Ray winds perform poorly near the surface and have significant overestimation errors. This may be due to the interference of strong aerosols near the surface on molecular scattering and the signal attenuation during propagation. Similarly, the accuracy performance of the Mie wind at heights of 0-2 km above the surface is also evaluated (Fig. S2). The R of the Mie wind is

0.86, slightly weaker than the Ray winds. In summary, both Ray winds and Mie winds exhibit significant positive deviations at altitudes below 500 m near the surface. These findings are also similar to many other researchers' results on L2B product (Guo et al., 2021; Borne et al., 2024).

In addition, the L2C products adopt different vertical height bin settings in different latitude intervals. Meanwhile, the vertical resolution from the surface to the upper levels also varies. These result in L2C products having different vertical resolutions at different locations and heights. The distribution of the vertical resolution of the L2C products below 2 km is statistically plotted in Fig. 3 (b) and Fig. S2 (b). The bars in the figure indicate the histograms and the curves represent the densities. The vertical resolution of the Ray winds is mainly distributed in the interval of 250-550 m, and also in the regions of 900 m and 1250 m. This indicates that the vertical resolution of Ray winds below 2 km is generally lower than 250 m, with a relatively coarse resolution. The Mie winds exhibit a similar distribution of vertical resolution, which is concentrated around 400 m, with sporadic distributions at 900 m and 1250 m. Overall, the vertical resolution of L2C products below 2 km is mainly around 400 m. The relatively coarse resolution limits the application of L2C products in the fields of wind energy and extreme weather. In summary, L2C products suffer from high missing rates, low vertical resolution, and limited accuracy from surface to 2 km. These defects indicate that Aeolus has the non-negligible observation gaps in the near surface.

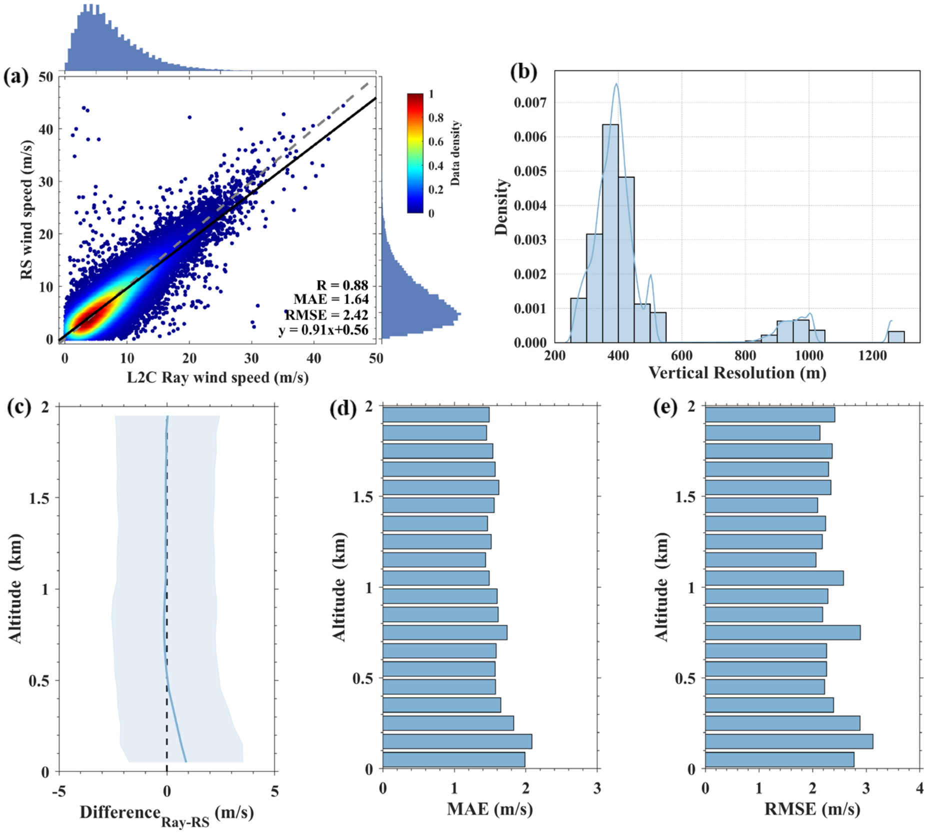


Figure 3. The **(a)** scatterplot and vertical distributions of **(c)** bias, **(d)** MAE, and **(e)** RMSE between Ray wind and RS wind below 2 km. **(b)** Vertical resolution histogram below 2 km for Ray product and the solid lines are the density curves.

2.4 ERA5

ERA5 is an atmospheric reanalysis dataset produced by the ECMWF, which simultaneously combines models and observational data (Soci et al., 2024). It provides hourly estimates of multiple atmospheric, terrestrial, and oceanic climate variables using a latitude and longitude grid of 0.25°× 0.25°. There are a total of 37 pressure levels for atmospheric variables in the vertical direction from the surface to the top (Hersbach et al., 2020). This study uses the " ERA5 hourly data on single levels from 1940 to present", which provides comprehensive surface parameters such as wind, radiation, temperature, and air pressure. Here, a total of 22 surface parameters are collected, including dewpoint temperature (d2m), temperature (t2m), skin temperature (skt), surface pressure (sp), surface net solar radiation (ssr), surface sensible heat flux (sshf), surface latent heat flux (slhf), Charnock, forecast surface roughness (fsr), friction velocity (zust), slope of sub-gridscale orography (slor), standard deviation of orography (sdor), angle of sub-gridscale orography (anor), anisotropy of sub-gridscale orography (isor), boundary layer height (blh), convective available potential energy (cape), convective inhibition (cin), instantaneous eastward turbulent surface stress (iews), instantaneous northward turbulent surface stress (inss), boundary layer dissipation (bld), gravity wave dissipation (gwd), and total precipitation (tp). Among them, d2m, t2m, skt, sp, and tp are meteorological parameters related to wind speed. The ssr, sshf, and slhf represent the level of solar radiation and directly affect the generation of wind. The Charnock, fsr, and zust characterize the surface roughness, determining the magnitude of the surface deceleration effect. The slor, sdor, anor, and isor are four terrain-related covariates. And blh, cape, cin, iews, inws, bld, and gwd are related to atmospheric turbulence and affect the vertical distribution of wind profiles.

The ERA5 variables were assigned to the corresponding grids of L2C wind profiles and RS stations based on their latitude, longitude and time information. In addition, wind profiles at different pressure levels from the “ERA5 hourly data on pressure levels from 1940 to present” dataset are also used to compare with the wind speed profiles generated in this study. They are interpolated to heights at 100 m intervals from surface to 2 km based on the surface geopotential and the geopotential of each pressure level.

3 Method

3.1 Power law method

The PLM describes the wind profile as following a power-law variation with height and has been extensively applied to vertical extrapolation and wind power generation estimation (Hellmann, 1914; Fu and Ren, 2026). It only requires surface wind speed as input and uses the following equation to calculate the upper-level wind speed:

$$v = v_0 \times (h/h_0)^{\alpha} \tag{1}$$

where v denotes the wind speed at height h, v_0 and h_0 are the wind speed and height of the surface observation, α indicates the dynamic power-law exponent whose value depends on time, height, and terrain conditions. (Watson, 2023). Typically, α is set to a constant of 0.14 in the absence of field observations. However, many researchers have pointed out that wind profiles extrapolated using constant α values exhibit significant errors (Jung and Schindler, 2021; Yang et al., 2024).

3.2 XBG-Wind method

The XGB-Wind uses satellite observations as the upper-level dynamical constraints, ERA5 meteorological variables and bias mechanisms as bottom-level constraints to generate global wind profile product within the whole ABL. The specific flowchart is shown in Fig. 4. Considering the lower missing rate and higher correlation coefficient of Ray product, the remaining part of this paper conducts experiments based on Ray product. The results based on the Mie product are included in the supplementary document (Fig. S3-Fig. S11).

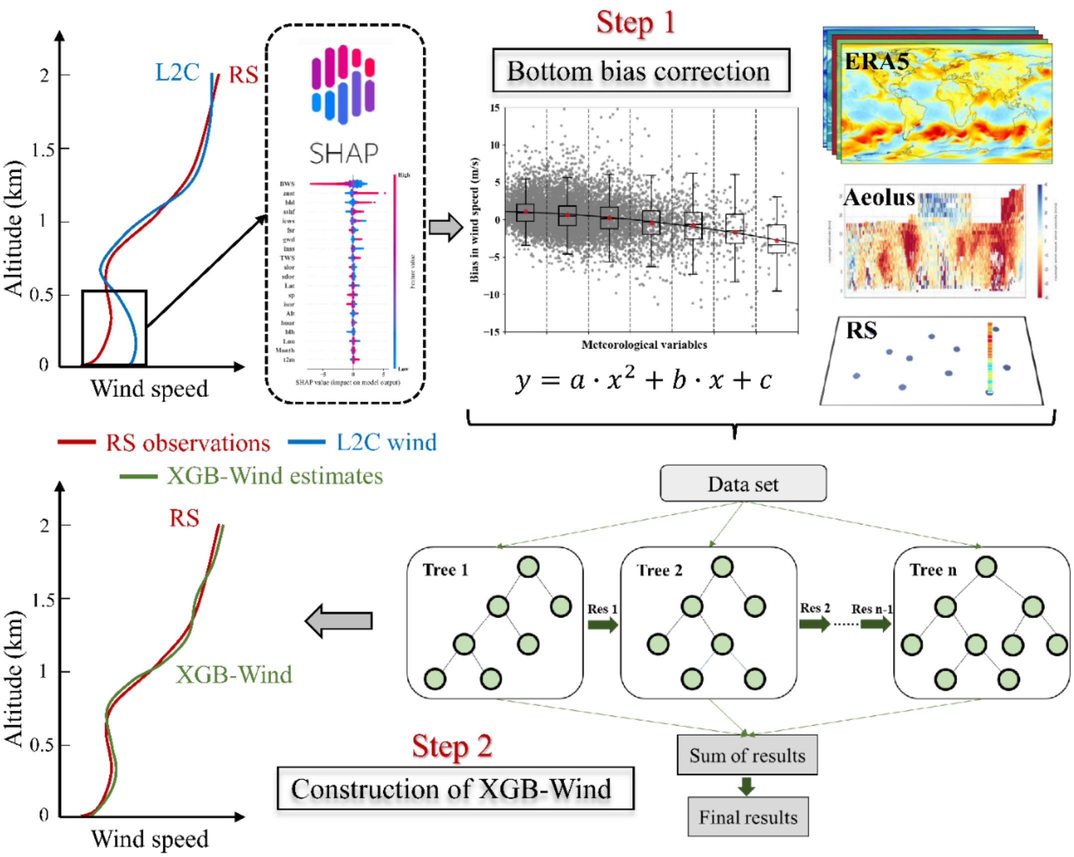


Figure 4. The flowchart of XGB-Wind method.

3.2.1 Correction of bottom wind speed bias

As mentioned in Sect. 2.3, there is a great positive bias in the L2C product at the bottom layer. In order to correct this bias, we constructed a model between the 0-500 m bottom wind speed bias and various parameters, and interpreted it using SHapley Additive exPlanations (SHAP) to explore the underlying reasons for the bias (Lundberg and Lee, 2017). As can be seen in Fig. 5 (a), the biases of Ray winds are strongly related to the bottom wind speed, friction velocity and sensible heat flux. The bottom wind speed is negatively correlated with the bias, and the lower the bottom wind speed, the more significant the positive bias of the Ray product. This may be because the laser signal is weakened by atmospheric attenuation over long distances, and the small echo frequency shifts at low wind speeds are easily masked by instrumental noise. These result in overestimated inversion wind speeds. The friction velocity and the surface sensible heat flux are both positively correlated with the bottom wind speed. The larger the friction velocity or surface sensible heat flux, the more the Ray product overestimates the actual wind speed. This may be due to the large wind shear caused by strong turbulence in the boundary layer when the friction velocity is high and the surface sensible heat flux is strong. ALADIN tends to focus more on the upper part of the bin with stronger signal strength during bin interval measurement. And the relatively high wind speed in this part leads to an overestimation in the Ray product near the surface. In addition, high sensible heat flux often occurs during the daytime, when strong solar background noise from sunlight will affect the detection accuracy of ALADIN (Borne et al., 2024). The boundary layer dissipation also contributes significantly to the positive bias of the bottom wind speed in the Ray channel measurements. The bld indicates the intensity of dissipation into thermal energy. High bld generally occurs in the strongly convective boundary layer. And the rapidly changing vertical wind profile in this situation affects the average measurement of Rayleigh channels.

For the purpose of quantifying the influence of these variables on the bias, we investigated the relationship between wind speed bias below 500 m and the top four important parameters and found that they can be represented by a quadratic function:

$$y = a \cdot x^2 + b \cdot x + c \tag{2}$$

where y represents the wind speed bias, x indicates the influencing variable, a, b and c are the parameters of the fitting function. As shown in Fig. 5 (b), (c), (d) and (e), the scattered dots denote the sampling points and the black curves denote the quadratic fit curves. Each box and red dot represent the quantile statistics and mean values within each interval, respectively. The determination coefficients (R^2) and equations of the fit are also listed in the figure. The asterisk in the upper right corner of R^2 marks results that are statistically significant. The influence of surface wind speed, friction velocity, sensible heat flux, and boundary layer dissipation on Ray wind bias can all be described by quadratic functions. The R^2 are 0.99, 0.97, 0.89 and 0.75, respectively. The R^2 of all the parameters are above 0.75 and have passed the significance test. This demonstrates a significant correlation between the Ray wind bias and these parameters. Accordingly, the bias correction can be calculated based on these mechanisms and applied to the subsequent machine learning model to enhance the accuracy of the near-surface wind speed inversion.

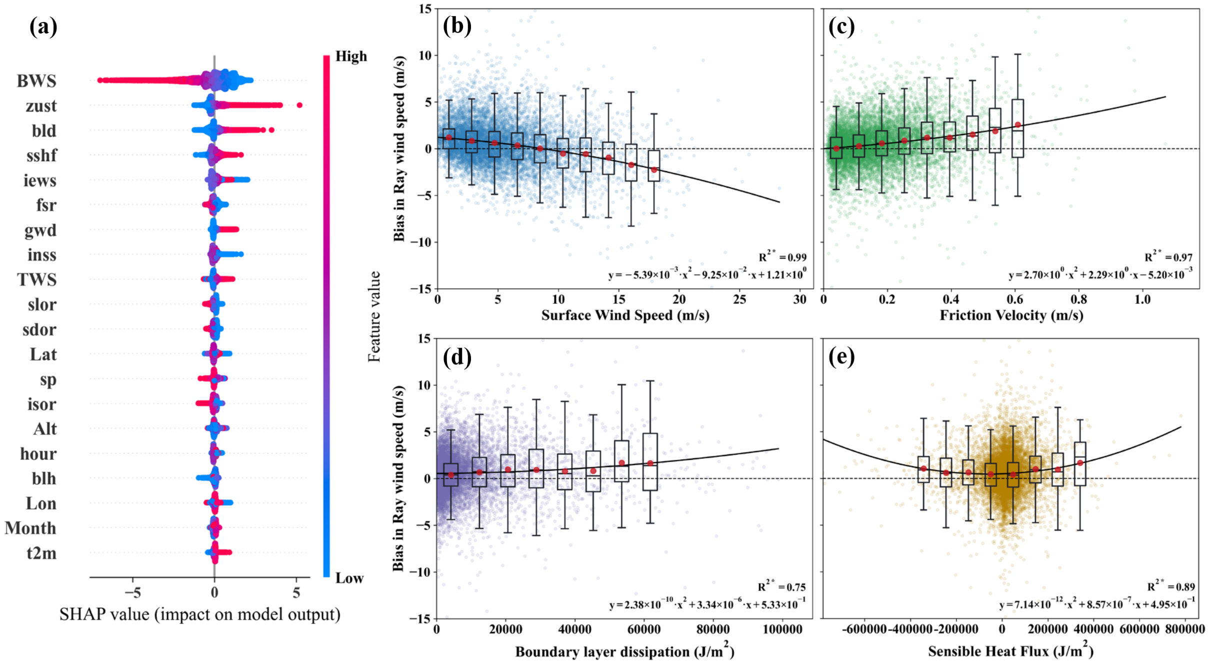


Figure 5. (a) SHAP analysis of the reasons for bottom wind speed bias for Ray wind. BWS: bottom wind speed; zust: friction velocity; bld: boundary layer dissipation; slor: slope of sub-gridscale orography; sshf: surface sensible heat flux; iews: instantaneous eastward turbulent surface stress; fsr: forecast surface roughness; gwd: gravity wave dissipation; inss: instantaneous northward turbulent surface stress; TWS: top wind speed; sdor: standard deviation of orography; Lat: latitude; sp: surface pressure; isor: anisotropy of sub-gridscale orography; Alt: altitude; blh: boundary layer height; Lon: longtitude; t2m: temperature; d2m: dewpoint temperature; anor: angle of sub-gridscale orography. The Ray wind speed bias below 500 m as a function of **(b)** surface wind speed, **(c)** friction velocity, **(d)** boundary layer dissipation, and **(e)** sensible heat flux. Red dots and black lines denote the mean values of each interval and the fitting curves, respectively.

3.2.2 Model construction

Under the preceding matching of the L2C and RS profiles, RS wind profiles that can be interpolated into 10 m and 100-2000 m (100 m intervals) are selected from the matching profiles to unify the vertical resolution for subsequent model construction and training. Among them, the 10 m RS wind is used as model input to constrain the near-surface wind speed. The 100-2000 m RS wind is used as labels to train the model. Similarly, the L2C profiles that can be interpolated to the 2100 m wind speed as the top-level dynamical constraints are screened out. It is worth noting that we use a 10 m RS wind speed instead of a 10 m ERA5 wind speed as bottom constraint. This is because the correlation between the 10 m ERA5 wind speed and the 100 m RS wind speed is very weak, with only 0.47, which directly affects the accuracy of the model in inverting the bottom wind speed. In addition, the correlation coefficients between the Ray wind at 2100 m and the RS wind speed at 2000 m are 0.90. This indicates that it is reliable to adopt the L2C 2100 m wind speed as the top dynamical constraint. A total of 15,255 Ray wind profiles are used for model training and validation after the screening process.

https://doi.org/10.5194/essd-2026-73
Preprint. Discussion started: 23 March 2026

The XGBoost model improves nonlinear prediction performance by combining multiple decision trees (weak learners) and has been applied to wind speed extrapolation and prediction (Xiong et al., 2022; Yu and Vautard, 2022). Based on the XGBoost model, we constructed an XGB-Wind model for inverting wind profiles within the ABL by combining physical constraints with meteorological parameters. Specifically, the XGB-Wind model accepts 22 ERA5 meteorological variables, month, hour, longitude (Lon), latitude (Lat), altitude (Alt), 10 m RS wind speed (BWS), 2100 m L2C wind speed (TWS), and four bias corrections (f (BWS), f (sshf), and f (zust) and f (bld)) as inputs. The reference labels are RS wind speeds from 100 m to 2000 m at 100 m vertical resolution. The XGB-Wind model adopts Bayesian optimization for hyperparameter selection and 10-fold cross-validation as the evaluation method. Bayesian optimization uses a probabilistic surrogate model to approximate the distribution of the objective function and determines the next most promising point by maximizing the acquisition function, thus effectively finding the optimal solution in the hyperparameter space (Snoek et al., 2012). Bayesian optimization is considered an effective hyperparameter optimization method and has been proven to have good performance in various machine learning training tasks (Wang et al., 2023; Lai, 2024). When the hyperparameters are set to Table 1, the XGB-Wind model reaches its optimum, with R (RMSE) values of 0.89 (2.32 m s^{-1}).

Table 1. Optimized hyperparameters of XGB-Wind model.

Hyperparameter	XGB-Wind
n_estimators	1500
max_depth	6
learing_rate	0.017
min_child_weight	2.2
subsample	0.5
colsample_bytree	0.6
gamma	2.7
reg_alpha	0.1
reg_lambda	1
R	0.89
RMSE	2.32

In order to understand the physical meaning of each parameter in the model and interpret the “black box”, we used SHAP to explain and conduct an importance analysis of the inputs, as shown in Fig. 6. At different heights, XGB-Wind focuses on different features. In the near surface, the model mainly focuses on BWS and the added correction terms f (BWS), zust, sshf, sdor, and Alt. BWS is the bottom wind speed. The zust and sshf are the friction velocity and surface sensible heat flux, respectively. The sdor and Alt are the standard deviation of surface height and the altitude, respectively. This indicates that the XGB-Wind model calculates wind profiles near the surface by non-linear fitting of surface wind speed, meteorological parameters, and terrain. In contrast, at the upper level, TWS, bld, and iews are the main relevant features. The bld stands for boundary layer dissipation, indicating the degree of kinetic energy loss. The iews is the instantaneous eastward turbulent surface stress, representing the friction exerted by the atmosphere on the surface in the east-west direction. This demonstrates that the XGB-Wind model is dominated by top wind speed and meteorological parameters at upper level. Overall, the XGB-Wind model can determine the dominant factors of wind speed according to the altitude. At the bottom level, XGB-Wind mainly focuses on surface meteorological parameters, wind speed, and terrain features, and explores the laws of wind variation from the surface vertically upwards. At the upper level, top dynamical constraints and parameters related to turbulence are more closely noticed by XGB-Wind, which takes into account the mechanism of atmospheric stability on wind speed at upper levels. Therefore, XGB-Wind can achieve accurate characterization and reasonable physical interpretation of wind profiles through this layered perception and feature mining.

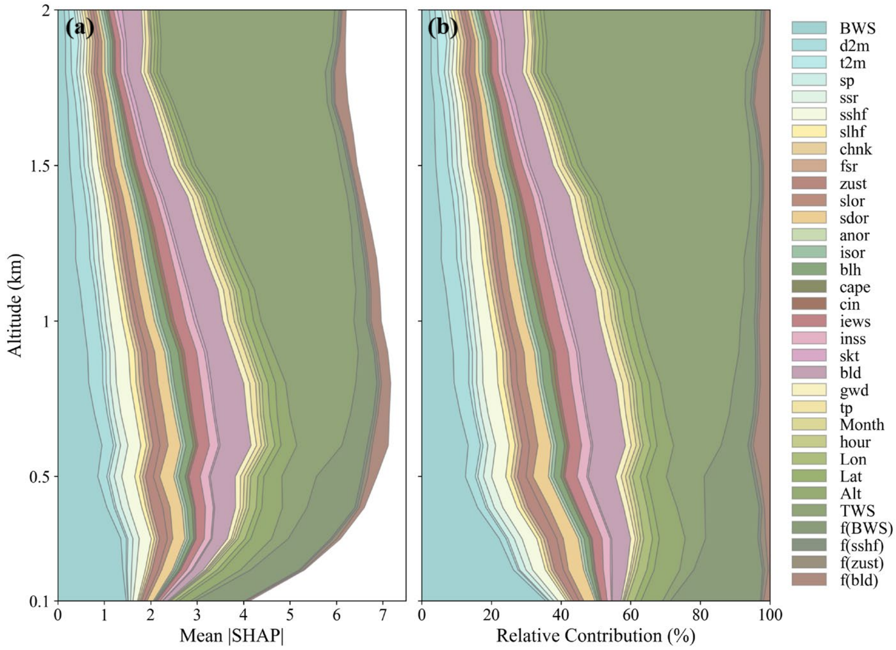


Figure 6. Vertical distribution of **(a)** mean |SHAP| and **(b)** relative contribution of |SHAP| of each feature on XGB-Wind. BWS: bottom wind speed; d2m: dewpoint temperature; t2m: temperature; sp: surface pressure; ssr: surface net solar radiation; sshf: surface sensible heat flux; slhf: surface latent heat flux; chnk: Charnock; frs: forecast surface roughness; zust: friction velocity; slor: slope of sub-gridscale orography; sdor: standard deviation of orography; anor: angle of sub-gridscale orography; isor: anisotropy of sub-gridscale orography; blh: boundary layer height; cape: convective available potential energy; cin: convective inhibition; iews: instantaneous eastward turbulent surface stress; inss: instantaneous northward turbulent surface stress; skt: skin temperature; bld: boundary layer dissipation; gwd: gravity wave dissipation; tp: total precipitation; Lon: longitude; Lat: latitude; Alt: altitude; TWS: top wind speed; f(BWS), f(sshf), f(zust), and f(bld) are the bias corrections calculated by the corresponding variables.

3.2.3 Sensitivity analysis

Figure 7 shows the relationship between several primary features and the bias of wind from Ray and XGB-Wind. Blue and red represent Ray and XGB-Wind, respectively. It should be noted that the resolution of XGB-Wind wind is reduced to be consistent with L2C bin to facilitate a fair comparison. The wind bias of XGB-Wind is significantly lower than Ray under different conditions. The bias of Ray wind varies significantly with surface wind speed, friction velocity, boundary layer dissipation, and sensible heat flux, showing a clear positive deviation overall. In contrast, the bias of the XGB-Wind model remains relatively stable around 0 and does not vary with these parameters. This indicates that the XGB-Wind model exhibits good generalization ability under different surface wind speeds, friction velocities, boundary layer dissipation, and sensible heat flux. The XGB-Wind model takes into account the friction velocity related to terrain, making it applicable to different regions around the world. In addition, the XGB-Wind incorporates various meteorological variables such as surface wind speed, boundary layer dissipation, and sensible heat flux, which characterize the atmospheric state under different conditions. This makes it have good generalization under different atmospheric conditions. All the above results prove that XGB-Wind maintains good accuracy and robustness in various situations.

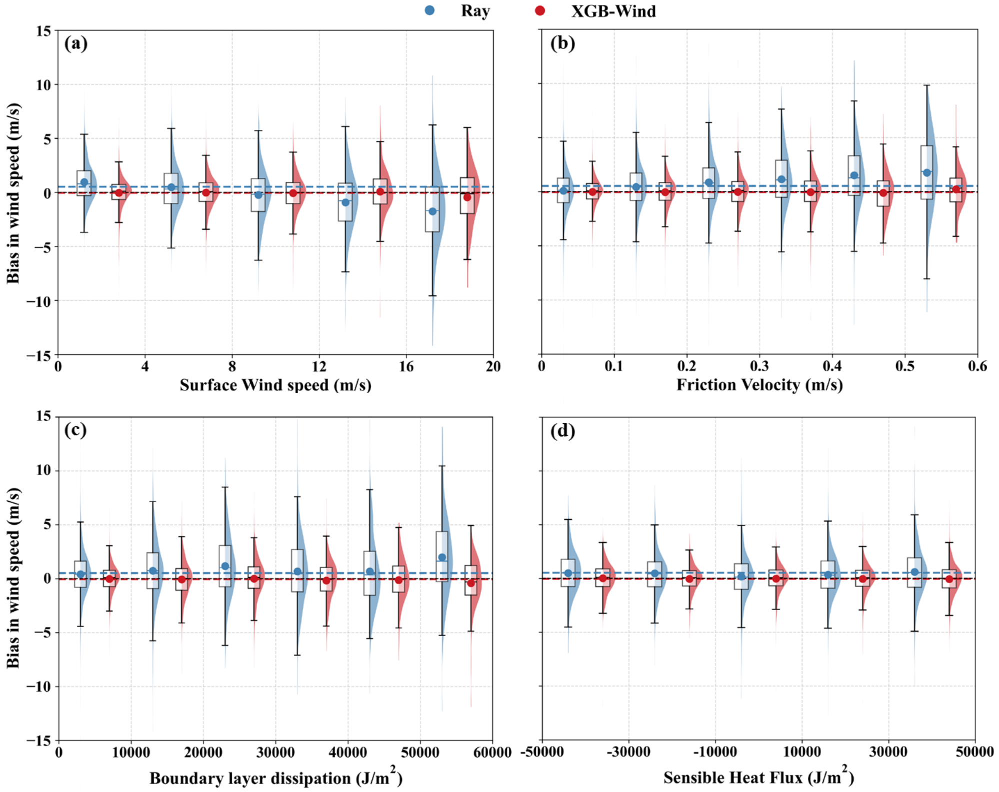


Figure 7. The boxplot of the bias in Ray (blue) and XGB-Wind (red) below 500 m under varying **(a)** surface wind speed, **(b)** friction velocity, **(c)** boundary layer dissipation, and **(d)** sensible heat flux. Colored shading denotes the density and dots denote average values.

4 Results and discussion

4.1 Performance comparison between XGB-Wind and Ray wind

The wind from the XGB-Wind and Ray product was compared with RS wind observations. The scatter plot is shown in Fig. 8 (a) and (b). The color indicates the data density, the solid line represents the linear fitting line, and the dashed line represents the y=x reference line. The R, bias, MAE, RMSE, and linear fitting equation are also displayed in each graph. The R (RMSE) from Ray and XGB-Wind are 0.90 (2.23 m s^{-1}) and 0.92 (1.94 m s^{-1}), respectively. The accuracy of XGB-Wind is better than Ray. The Ray product is overestimated at low wind speeds and underestimated at high wind speeds. This may be related to the significant positive bias of Ray at the bottom layer caused by attenuation of long-distance propagation signals and complex atmospheric conditions near the surface, such as turbulent surface friction (Abril-Gago et al., 2023). In contrast, XGB-Wind performs well under different wind speeds, and the fitted line is very close to the y=x standard line. XGB-Wind takes more environmental factors into account during the inversion process, such as zust, sshf, and bld. Moreover, XGB-Wind has also added physical constraints for bias correction. These features enhance the accuracy of the model as it learns the complex nonlinear relationship between them and the wind profile. In summary, the wind derived from XGB-Wind is highly correlated with RS wind and has higher accuracy compared to Ray wind.

The performance of Ray and XGB-Wind in vertical altitude is plotted in Fig. 8 (c) and (d) to demonstrate the ability of XGB-Wind to improve the bias of near-surface wind speed. Blue and red colors denote Ray and XGB-Wind, respectively. And the shaded area represents the standard deviation of wind speed bias. At the altitude of 0-500 m, the Ray wind has an obvious positive bias. In contrast, this positive bias is not found in the near surface of the wind derived from XGB-Wind. The MAE and RMSE of XGB-Wind wind at the bottom layer are about 1 m s^{-1} and 1.5 m s^{-1}, respectively, much lower than the 2 m s^{-1} and 2.5 m s^{-1} of Ray wind. From the whole profile, XGB-Wind maintains a small bias from 0 to 2 km, with MAE and RMSE stable at small values. Although XGB-Wind presents slightly higher MAE and RMSE than Ray in the range of 1.0-1.5 km, the higher degree is limited. This is because the wind speed inversion in the middle layer is less constrained by the top and bottom wind speeds, resulting in relatively great uncertainty. In general, XGB-Wind takes into account the influence of surface friction, radiation flux and turbulence on the wind profile by adding the bottom bias mechanisms and several surface parameters, such as zust and sshf. Compared with Ray product, XGB-Wind overcomes the significant positive bias of the bottom layer and maintains good accuracy throughout the entire 0-2 km altitude.

https://doi.org/10.5194/essd-2026-73
Preprint. Discussion started: 23 March 2026

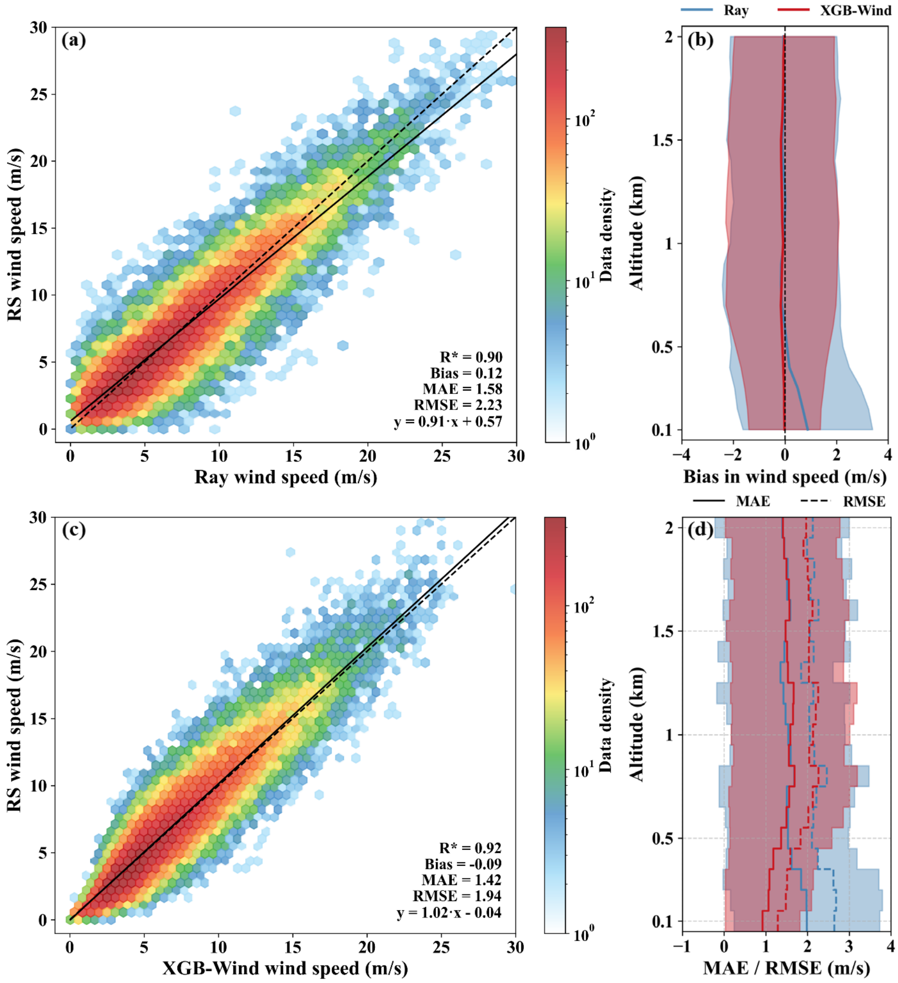


Figure 8. Scatter plots comparing RS-observed wind speeds with estimates from the **(a)** Ray and **(b)** XGB-Wind from 0 to 2 km. The dashed line is the 1:1 reference line, the solid lines denote the linear regression fits, and the color bar indicates data density. The asterisk (*) marks R that are statistically significant ($p < 0.05$). Vertical distributions of the **(c)** bias, **(d)** MAE and RMSE. Blue and red represent Ray and XGB-Wind, respectively. Colored shading represents the standard deviation. The dashed and solid lines in (d) represent RMSE and MAE, respectively.

4.2 Comparison with ERA5 and PLM

At high vertical resolution (100 m), the wind profiles derived from XGB-Wind is compared with existing common wind profile products such as ERA5 and PLM, as shown in Taylor Fig. 9. Figure 9 (a), (b), (c), (d), and (e) show the results at multiple vertical levels (100 m, 500 m, 1000 m, 1500 m, and 2000 m), while Fig. 9 (f) shows the results of the entire profile and each altitude (black represents the entire profile, while colors represent different altitudes). Circles, squares, and triangles represent PLM, ERA5, and XGB-Wind, respectively, and the pentagram represents RS observations. The horizontal and vertical axes in the figure represent the standard deviation, the angle represents the R, and the magenta dashed line denotes the centered RMSE compared with the RS observations. The closer the marker is to the pentagram, the higher the correlation coefficient with the observation, the lower the centered RMSE, and the more similar the standard deviation to the observation. Overall, XGB-Wind has the highest R of about 0.9, ERA5 has a correlation coefficient close to 0.85, and the R of PLM is only approximately 0.6. The centered RMSE of XGB-Wind, ERA5, and PLM are 2.34 m s^{-1}, 2.78 m s^{-1}, and 5.59 m s^{-1}, respectively. This indicates that XGB-Wind exhibits superior performance, followed by ERA5, while the PLM ranks last. From the perspective of altitude, XGB-Wind outperforms the other two methods at all altitudes. The R of both the bottom and top wind speed of XGB-Wind exceeds 0.9, and the result of the middle wind speed is relatively poor (close to 0.85). This is because the mid-level wind speed is at an intermediate altitude and is less affected by the constraints of the input bottom-level meteorological variables and wind speed, as well as the top-level wind speed. The R of ERA5 gradually increased from 0.6 at an altitude of 100 m to 0.9 at 2 km altitude. RS measurements are point observations, while ERA5 provides grid averages. The complex terrain variations near the surface make it difficult for the model to accurately simulate the changes of wind profiles, and the near-surface layer is significantly affected by processes such as surface friction, thermal disturbance, and turbulence. All these reasons contribute to the poor consistency between ERA5 and RS in the near-surface layer (Hu et al., 2023; Pernov et al., 2024). As the altitude increases, the influence of terrain on the wind field weakens, and the dominant mechanism is gradually controlled by large-scale pressure gradients. ERA5 has better simulation ability for large-scale circulation (Soci et al., 2024). PLM shows the opposite pattern to ERA5, with the accuracy rapidly decreasing as altitude increases. The R reaches as high as 0.94 at 100 m, while it decreases to 0.79 at 300 m and as low as 0.44 at 2 km. This is because, beyond the surface layer, the Coriolis force, boundary layer height, wind shear, and low-level jet have significant impacts on the wind field, which collectively render the wind profile more complex (Wiel et al., 2010; Li et al., 2018; Coleman et al., 2021). The traditional PLM (α=0.14) cannot accurately provide wind profiles beyond the surface layer based on experience. From the above results, XGB-Wind exhibits excellent performance throughout the entire ABL.

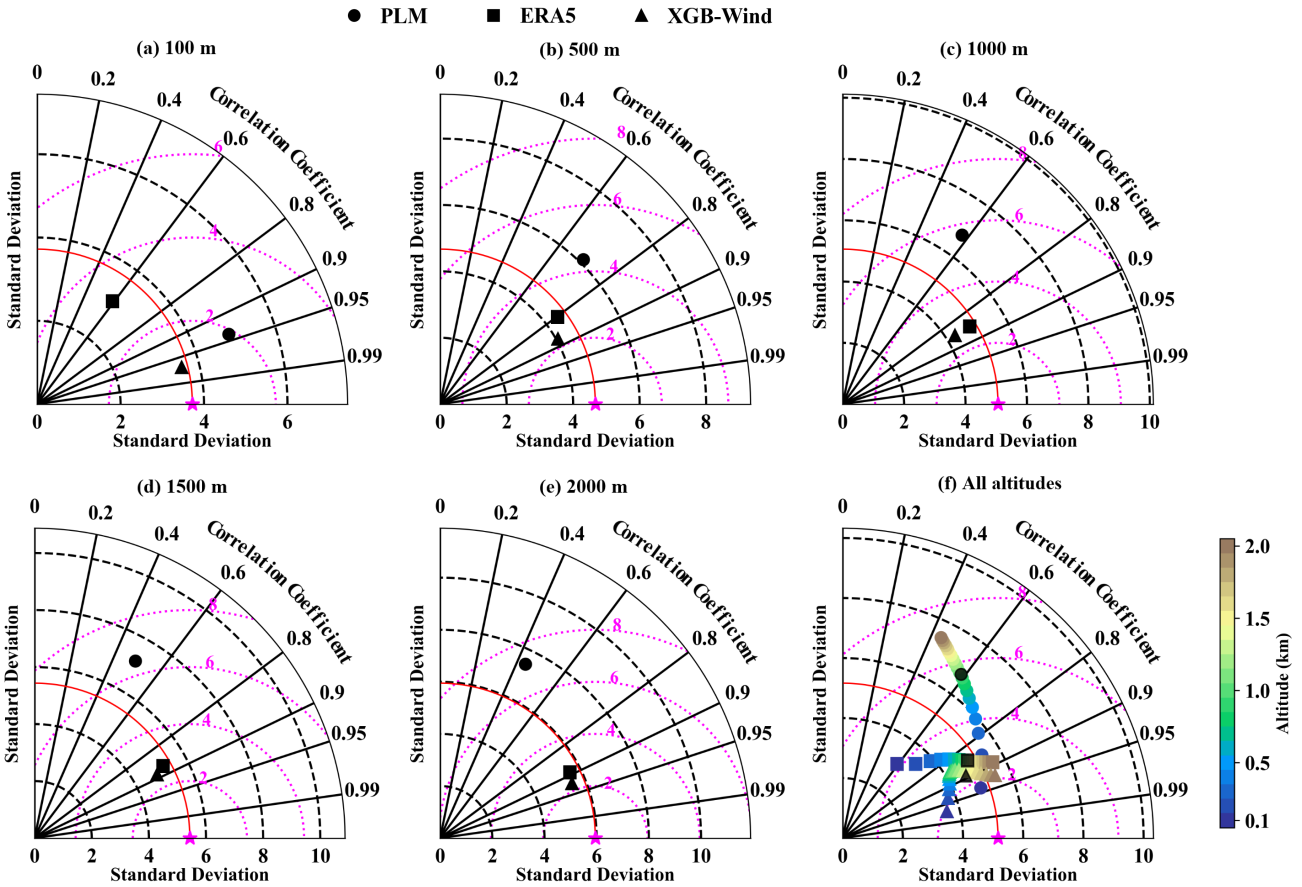


Figure 9. Taylor diagram of the performance of PLM (circles), ERA5 (squares), and XGB-Wind (triangles) at **(a)** 100 m, **(b)** 500 m, **(c)**1000 m, **(d)** 1500 m, **(e)** 2000 m, and **(f)** all altitudes. In subgraph **(f)**, black represents the results of the entire profile, while color represents the results at different altitudes.

4.3 Wind profile evaluation of the Global-ABLWind

The XGB-Wind model is applied to all Ray product throughout the entire data period to generate Global-ABLWind, and we use ERA5 10 m wind speed as an alternative input due to the unavailability of bottom wind speed constraints at locations not covered by RS stations. The generated Global-ABLWind product maintains the same spatial and temporal resolution as L2C, but improves the missing rate, accuracy, and vertical resolution from the surface to 2 km altitude. Statistically, the missing rate of the Global-ABLWind product within the ABL has decreased to 1.9%, significantly lower than the 16.4% of the Ray product. Figure 10 shows the geographical distribution of the missing rate of the Global-ABLWind and its improvement compared to Ray product. The Global-ABLWind product maintains a low missing rate in most parts of the world. Although there is a relatively high missing rate in common cold current areas such as the Peru Current, Benguela Current, Canary Current, California Current, and Western Australian Current, the missing rate is much lower compared to Ray product. And the missing rate of the Global-ABLWind product is significantly lower in both the mainland and marine regions. And unlike Ray product, which has a significantly higher missing rate on ocean than on land, Global-ABLWind product has a significantly lower missing rate on marine areas, basically showing the same low missing rate as on land areas. This phenomenon is particularly evident in mid to high latitude ocean regions, where the missing rate of Global-ABLWind product has generally decreased by approximately 0.2. These results indicate that the missing rate of Global-ABLWind product has significantly reduced compared to Ray product and has a superior global coverage capability. In addition, it should be noted that the vertical resolution of Global-ABLWind is 100 m, which is also much better than Ray product. The above results demonstrate that the Global-ABLWind compensates for the observation gaps of Aeolus with higher accuracy and vertical resolution.

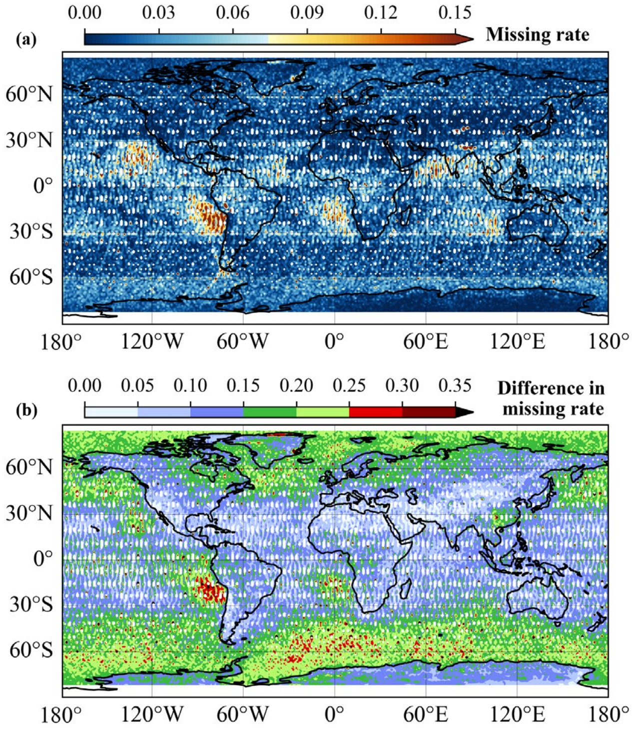


Figure 10. Spatial distribution of **(a)** the missing rate of the Global-ABLWind product and **(b)** its improvement in missing rate compared to Ray product.

Figure 11 illustrates the spatial pattern of average wind speed at an altitude of 100 m for the Global-ABLWind product in 2021 (the year with the largest amount of data). Colored shading represents Global-ABLWind, and colored dots represent average wind speed measured by RS. The RS stations with red boundaries are used for subsequent profile evaluation. Overall, Global-ABLWind can present the distribution of global wind speeds well, similar to the wind speed observed by RS. However, the wind speed derived from Global-ABLWind at some sites in China and northern North America is relatively underestimated. The average latitudinal bias is -0.58 m s^{-1}. This is because the bottom wind speed constraint of global

Global-ABLWind product adopts ERA5 10 m wind, which underestimates the actual wind speed. Regarding spatial distribution, the wind speed is higher over the ocean while lower over the continent. The annual average wind speed on all continents except Antarctica and Greenland is generally below 8 m s^{-1}, while over the oceans, except for a small portion of equatorial ocean areas, it is generally above 8 m s^{-1}. One reason is that the ocean surface is smoother and has lower roughness. Small frictional resistance leads to lower momentum dissipation, resulting in less wind speed loss. A further contributing factor is that the ocean has a high heat capacity, slow temperature changes, and a stable atmospheric structure that is closer to neutral or weakly unstable. Strong turbulent vertical mixing promotes momentum exchange between low and high altitudes. In contrast, the formation of a stable layer at night in land regions with large diurnal temperature differences inhibits the downward transmission of high wind speeds from upper levels. The Southern Hemisphere shows stronger winds than the Northern Hemisphere. The global annual mean wind speed exhibits pronounced latitudinal variations. Overall, In the equatorial belt (approximately 10° S–10° N), wind speeds are the weakest, widely below 7 m s^{-1}, forming a distinct "equatorial low-wind zone." Toward the subtropics (20°–30° N/S), wind speeds slightly increase, while in the mid-latitudes (30°–60° N/S), especially over oceanic regions, wind speeds rise markedly, commonly reaching 10–15 m s^{-1}. The strongest wind regime is found over the mid-latitude Southern Ocean (40°–60° S), where persistent westerlies dominate. In this "Roaring Forties to Screaming Sixties" belt, wind speeds widely exceed 12 m s^{-1}, with local maxima reaching over 15 m s^{-1}, making it the most prominent global high-wind corridor. In contrast, the Northern Hemisphere oceans, such as the North Atlantic and North Pacific mid-latitudes, also feature high-wind belts but with slightly weaker intensity compared to their Southern Hemisphere counterpart. By comparison, tropical and subtropical continental regions (e.g., central Africa, South Asia, Southeast Asia, and the Amazon basin) exhibit relatively weak winds, typically less than 5 m s^{-1}.

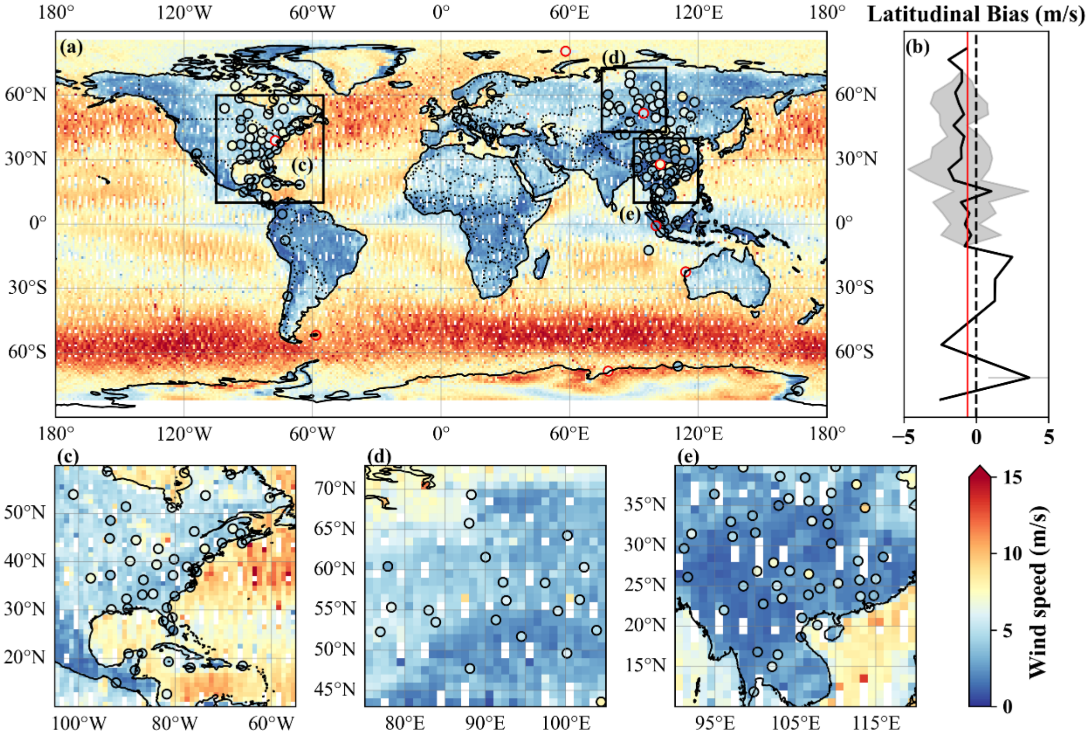


Figure 11. Spatial distributions of the average wind speed at 100 m from the Global-ABLWind product (colored shading) and the RS (colored points) in 2021 in **(a)** global, **(c)** North America, **(d)** Russia and Central Asia, and **(e)** China and Southeast Asia. Red-edge dots represent the 8 stations used for subsequent mean profile display. **(b)** The mean latitudinal bias between Global-ABLWind and RS. The gray shading and red line represent the standard deviation and global mean value, respectively.

Figure 12 shows the average wind profiles of eight globally distributed stations from Global-ABLWind (black), RS (red), ERA5 (blue), and PLM (brown). The specific geographical location of the sites can refer to the red-edge dots in Fig. 11. The station name, latitude, longitude, and number of profiles are also marked in the figure. The wind profiles derived by Global-ABLWind at all eight sites are highly consistent with RS. The upper-level wind profiles of ERA5 at six stations except Kyzyl and Xichang are in good agreement with the observed RS winds. However, there is a clear underestimation of wind speed at the bottom level of six out of eight sites. PLM performs the worst, showing significant overestimation at all sites. Although at some sites, the bottom-level wind profiles obtained by PLM are similar to the trend observed by RS, there are large deviations at upper levels. Especially at Mia Padang, Learmonth Airport, and Davis stations, the wind profiles exhibit a completely different form from the power function. This means that using power function to simulate wind profiles will introduce significant errors. Overall, the wind profile of Global-ABLWind is closest to the observed RS wind. ERA5 performs well at the top and relatively poorly at the bottom. PLM has the worst performance.

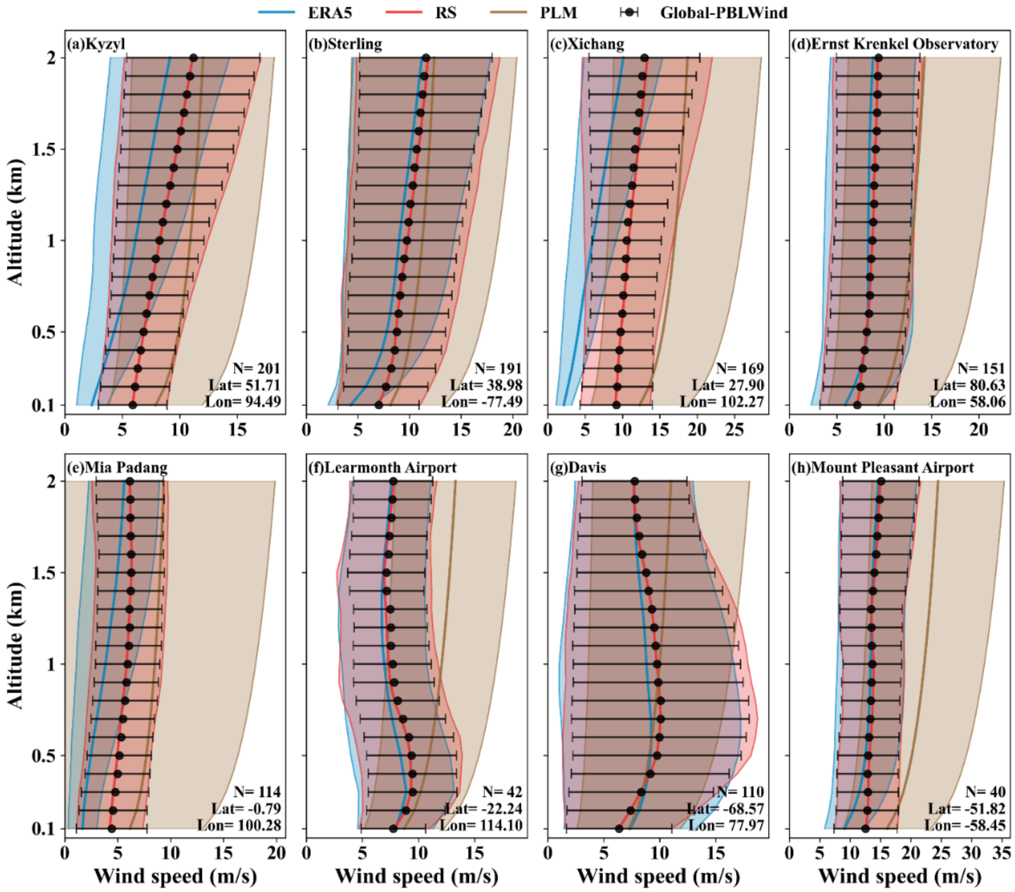
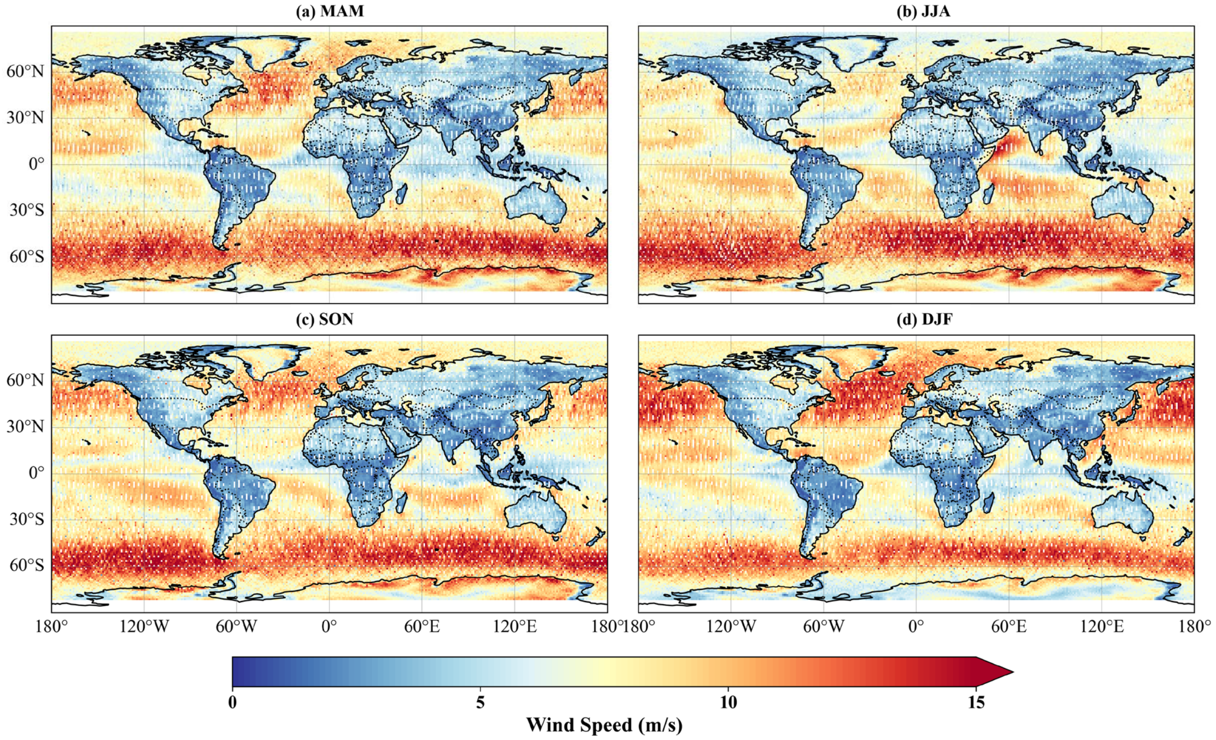


Figure 12. Vertical distributions of average wind speed as derived from multiple methods at **(a)** Kyzyl, **(b)** Sterling, **(c)** Xichang, **(d)** Ernst Krenkel Observatory, **(e)** Mia Padang, **(f)** Learmonth Airport, **(g)** Davis, and **(h)** Mount Pleasant Airport. Blue, red, brown, and black represent ERA5, RS, PLM, and Global-ABLWind, respectively. The colored shading or error bars denote the standard deviation.

4.4 Seasonal variation of global wind speed at 100 m altitude

Figure 13 shows the average wind speed of Global-ABLWind product at an altitude of 100 m in different seasons. MAM represents March, April, and May, JJA represents June, July, and August, SON represents September, October, and November, and DJF represents December, January, and February. Seasonal contrasts are obvious. The wind speed in the Northern Hemisphere is maximized in DJF and minimized during JJA. On the contrary, the Southern Hemisphere shows the opposite pattern, with JJA being the largest and DLF being the smallest. During DJF (Northern Hemisphere winter), the Northern Hemisphere mid-latitude westerlies intensify, producing enhanced wind speeds over the North Atlantic, North Pacific, and adjacent European seas (mean values frequently ranging from 8 m s^{-1} to 15 m s^{-1}), consistent with stronger storm-track activity in winter. Conversely, during JJA (Northern Hemisphere summer), the Northern Hemisphere mid-latitude westerlies weaken and shift poleward, resulting in reduced wind speeds over the North Atlantic and North Pacific relative to winter. Differently, the Southern Ocean high-wind belt remains a dominant feature in all seasons with mean wind speeds commonly exceeding 12 m s^{-1} in the 40°–60° S band. Tropical and monsoon domains exhibit marked seasonal variability at regional scales. The equatorial low-wind zone migrates seasonally with the shifting Hadley circulation and Intertropical Convergence Zone, shifting northward in Northern Hemisphere summer and southward in Northern Hemisphere winter. In JJA, monsoon-affected regions (South and Southeast Asia) display clear land–sea contrasts: continental interiors often show relatively low mean wind speeds, while adjacent seas (Arabian Sea, Bay of Bengal, South China Sea) exhibit locally enhanced low-level flow associated with monsoon circulation. Transitional seasons (MAM and SON) show intermediate patterns between the Northern Hemisphere summer and winter states. Another notable contrast emerges over Antarctica and its surrounding seas. Compared to other continents with distinct sea-land difference, the coastal edge of the Antarctic continent has the same high wind speed areas as the Southern Ocean. But the interior of the Antarctic continent is characterized by weaker winds, typically below 5 m s^{-1}. This suppression is mainly attributable to the high, ice-covered plateau, enhanced surface friction, and the prevalence of a strongly stratified polar boundary layer. However, at the coastal margins and sloping ice-sheet regions, local wind speeds are markedly enhanced due to persistent katabatic flows, creating sharp gradients between the weakly ventilated interior and the windier periphery. This contrast highlights the distinct dynamical regimes of the Antarctic ice sheet and the circumpolar storm track.



490 **Figure 13.** Spatial distributions of the average wind speed at 100 m from the Global-ABLWind product in **(a)** MAM (March, April, and May), **(b)** JJA (June, July, and August), **(c)** SON (September, October, and November), and **(d)** DJF (December, January, and February).

5. Data availability

The global IGRA data are available from https://www.ncei.noaa.gov/products/weather-balloon/integrated-global-radiosonde-archive. The Aeolus L2C dataset can be freely available from https://aeolus-ds.eo.esa.int/oads/access/collection/L2C_Wind_Products. The ERA5 products can be publicly downloaded from https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=overview and https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview. The generated Global-ABLWind products are available at https://doi.org/10.5281/zenodo.18286457 (Tong et al., 2026).

6. Conclusions

The wind speed profiles within the ABL are crucial for both atmospheric dynamics and wind energy assessment. Existing global products capable of providing wind profiles across the entire ABL, such as ERA5 and Aeolus L2C, still suffer from pronounced inconsistency or unignorable observational gaps. These limitations reflect a deeper challenge in representing the physical coupling between surface forcing, boundary-layer turbulence, and upper-level wind structures. To overcome this limitation, this study generates Global-ABLWind, a new global wind speed profile dataset covering the entire ABL from the surface to 2 km altitude (20 height levels) spanning from July 2020 to April 2023. The physics-constrained machine learning framework that integrates physically interpretable bias mechanisms and upper-level dynamical constraints was applied to model wind speed profiles.

1. The dual-constraint design, including surface-driven physical bias mechanisms and upper-level dynamical constraints, enables XGB-Wind to maintain both precision and robustness across different atmospheric conditions, especially near the surface (0-500 m). The wind speed reconstructed by the proposed model exhibits the best consistency with RS data, with R, MAE, and RMSE values of 0.89, 1.71 m s^{-1}, and 2.34 m s^{-1}, respectively.
2. The generated Global-ABLWind fills the near-surface gap caused by signal attenuation and surface backscatter contamination. Its missing rate within ABL significantly decreases from 16.4% of Aeolus L2C product to 1.9%, achieving full ABL coverage. And the dataset exhibits the advantages of enhanced accuracy, finer vertical resolution, and reduced data gaps.
3. Compared to existing reanalysis-based wind profile datasets, Global-ABLWind reveals the geographical distribution and seasonal variations of global wind speed profiles from an observational perspective. Analyses based on the dataset demonstrate pronounced land–sea contrasts and latitudinal dependence, including weak winds in the equatorial belt and the strongest annual mean wind speeds in the “Roaring Forties to Screaming Sixties” belt. Seasonal variations show higher wind speeds in winter and lower wind speeds in summer in both hemispheres, with the equatorial low-wind zone migrating seasonally in association with large-scale atmospheric circulation.

While Global-ABLWind represents a significant improvement over existing global wind profile products, several limitations should be noted. The accuracy of wind speed estimates in the middle portion of the ABL needs to be optimized, resulted by the unavailability of direct highly correlated constraints at these heights. In addition, the quality of upper-level wind information significantly influences the precision of wind profiles in the upper ABL, and the coarse spatial and temporal resolution of the input Aeolus L2C data imposes constraints on the final product. Users are therefore advised to consider these uncertainties when applying the dataset, particularly for high-frequency analyses. Despite these limitations, Global-ABLWind provides a valuable new global dataset for a wide range of applications in boundary-layer meteorology, climate studies, air quality research, and wind energy assessment. Although the Aeolus mission has ended, the methodology underpinning Global-ABLWind demonstrates the potential of integrating spaceborne wind lidar observations with multi-source data to generate comprehensive ABL wind datasets, and may inform the development of future satellite-based wind profile products.

Author contributions

Conceptualization: BL and ZT; Data curation: ZT and HZ; Formal analysis: BL and GH; Funding acquisition: XM; Investigation: ZT and HD; Methodology ZT and BL; Project administration: WG and JG; Resources: YM; Software: ZT and YM; Supervision: XM, HZ, HD, JG, and WG; Validation: GH; Visualization: ZT, BL, and HZ; Writing original draft preparation: ZT; Writing review and editing: BL, XM and ZT.

Competing interests

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

Disclaimer

Acknowledgements

The authors would like to express their gratitude to the National Centers for Environmental Information for providing the Integrated Global Radiosonde Archive and the Copernicus Climate Data Store for providing the ERA5 reanalysis data. Our gratitude also goes to the European Space Agency (ESA) for providing the Aeolus L2C products The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Financial support

This work was supported by the National Natural Science Foundation of China (Grant 42171464), the Fundamental Research Funds for the Central Universities (ZNJC202415 and 413000028), the National Key R&D Program of China (2024YFC3015600), and the Science and Technology Program of Hubei Provincial (2025BEB017).

References

Abril-Gago, J., Ortiz-Amezcua, P., Bermejo-Pantaleón, D., Andújar-Maqueda, J., Bravo-Aranda, J. A., Granados-Muñoz, M. J., Navas-Guzmán, F., Alados-Arboledas, L., Foyo-Moreno, I., and Guerrero-Rascado, J. L.: Validation activities of Aeolus wind products on the southeastern Iberian Peninsula, Atmos. Chem. Phys., 23, 8453–8471, https://doi.org/10.5194/acp-23-8453-2023, 2023.

Baars, H., Herzog, A., Heese, B., Ohneiser, K., Hanbuch, K., Hofer, J., Yin, Z., Engelmann, R., and Wandinger, U.: Validation of aeolus wind products above the atlantic ocean, Atmos. Meas. Tech., 13, 6007–6024, https://doi.org/10.5194/amt-13-6007-2020, 2020.

Baas, P., Bosveld, F. C., Baltink, H. K., and Holtslag, A. A. M.: A climatology of nocturnal low-level jets at cabauw, J. Appl. Meteorol. Climatol., 48, 1627–1642, https://doi.org/10.1175/2009JAMC1965.1, 2009.

Banyard, T. P., Wright, C. J., Osprey, S. M., Hindley, N. P., Halloran, G., Coy, L., Newman, P. A., Butchart, N., Bramberger, M., and Alexander, M. J.: Aeolus wind lidar observations of the 2019/2020 quasi-biennial oscillation disruption with comparison to radiosondes and reanalysis, Atmos. Chem. Phys., 24, 2465–2490, https://doi.org/10.5194/acp-24-2465-2024, 2024.

Borne, M., Knippertz, P., Weissmann, M., Witschas, B., Flamant, C., Rios-Berrios, R., and Veals, P.: Validation of Aeolus L2B products over the tropical Atlantic using radiosondes, Atmos. Meas. Tech., 17, 561–581, https://doi.org/10.5194/amt-17-561-2024, 2024.

Coleman, T. A., Knupp, K. R., and Pangle, P. T.: The Effects of Heterogeneous Surface Roughness on Boundary Layer Kinematics and Wind Shear, E-J. Sev. Storms Meteorol., 16, 1–29, https://doi.org/10.55599/ejssm.v16i3.80, 2021.

Deng, X., He, D., Zhang, G., Zhu, S., Dai, R., Jin, X., Fu, W., Shen, W., Chen, J., Fan, Y., Wei, C., and Li, X.: Comparison of horizontal wind observed by wind profiler radars with ERA5 reanalysis data in Anhui, China, Theor. Appl. Climatol., 150, 1745–1760, https://doi.org/10.1007/s00704-022-04247-6, 2022.

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the integrated global radiosonde archive, J. Clim., 19, 53–68, https://doi.org/10.1175/JCLI3594.1, 2006.

Durre, I., Yin, X., Vose, R. S., Applequist, S., and Arnfield, J.: Enhancing the data coverage in the integrated global radiosonde archive, J. Atmos. Oceanic Technol., 35, 1753–1770, https://doi.org/10.1175/JTECH-D-17-0223.1, 2018.

Frech, J., Saha, K., Lavin, P. D., Zhang, H.-M., Reagan, J., and Fung, B.: A new gridded offshore wind profile product for US coasts using machine learning and satellite observations, Wind Energy Sci., 10, 1077–1099, https://doi.org/10.5194/wes-10-1077-2025, 2025.

Fu, K. and Ren, Z.: Multi-task prediction of wind speed and time-varying wind shear coefficient using dynamic graph interactive neural network, Inf. Fusion, 125, 103478, https://doi.org/10.1016/j.inffus.2025.103478, 2026.

Ghannam, K. and Bou-Zeid, E.: Baroclinicity and directional shear explain departures from the logarithmic wind profile, Q. J. R. Meteorolog. Soc., 147, 443–464, https://doi.org/10.1002/qj.3927, 2021.

Gualtieri, G.: A comprehensive review on wind resource extrapolation models applied in wind energy, Renew. Sustain. Energy Rev., 102, 215–233, https://doi.org/10.1016/j.rser.2018.12.015, 2019.

Guo, J., Liu, B., Gong, W., Shi, L., Zhang, Y., Ma, Y., Zhang, J., Chen, T., Bai, K., Stoffelen, A., de Leeuw, G., and Xu, X.: Technical note: First comparison of wind observations from ESA's satellite mission Aeolus and ground-based radar wind profiler network of China, Atmos. Chem. Phys., 21, 2945–2958, https://doi.org/10.5194/acp-21-2945-2021, 2021.

Hatfield, D., Hasager, C. B., and Karagali, I.: Vertical extrapolation of Advanced Scatterometer (ASCAT) ocean surface winds using machine-learning techniques, Wind Energy Sci., 8, 621–637, https://doi.org/10.5194/wes-8-621-2023, 2023.

Hellmann, G.: Über die Bewegung der Luft in den untersten Schichten der Atmosphäre, Kgl. Akademie der Wissenschaften [G.] Reimer, 1914.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorolog. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Horányi, A., Cardinali, C., Rennie, M., and Isaksen, L.: The assimilation of horizontal line-of-sight wind information into the ECMWF data assimilation and forecasting system. Part I: The assessment of wind impact, Q. J. R. Meteorolog. Soc., 141, 1223–1232, https://doi.org/10.1002/qj.2430, 2015.

Hu, W., Scholz, Y., Yeligeti, M., Bremen, L. von, and Deng, Y.: Downscaling ERA5 wind speed data: a machine learning approach considering topographic influences, Environ. Res. Lett., 18, 094007, https://doi.org/10.1088/1748-9326/aceb0a, 2023.

Iwai, H., Aoki, M., Oshiro, M., and Ishii, S.: Validation of Aeolus Level 2B wind products using wind profilers, ground-based Doppler wind lidars, and radiosondes in Japan, Atmos. Meas. Tech., 14, 7255–7275, https://doi.org/10.5194/amt-14-7255-2021, 2021.

Jung, C. and Schindler, D.: The role of the power law exponent in wind energy assessment: A global analysis, Int. J. Energy Res., 45, 8484–8496, https://doi.org/10.1002/er.6382, 2021.

Kalmus, P., Ao, C. O., Wang, K.-N., Manzi, M. P., and Teixeira, J.: A high-resolution planetary boundary layer height seasonal climatology from GNSS radio occultations, Remote Sens. Environ., 276, 113037, https://doi.org/10.1016/j.rse.2022.113037, 2022.

Kent, C. W., Grimmond, C. S. B., Gatey, D., and Barlow, J. F.: Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds, J. Wind Eng. Ind. Aerodyn., 173, 100–111, https://doi.org/10.1016/j.jweia.2017.09.007, 2018.

Lai, Y.: Application and effectiveness evaluation of bayesian optimization algorithm in hyperparameter tuning of machine learning models, in: 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC), 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC), 351–355, https://doi.org/10.1109/PEEEC63877.2024.00070, 2024.

Li, J., Wang, X., and Yu, X. (Bill): Use of spatio-temporal calibrated wind shear model to improve accuracy of wind resource assessment, Appl. Energy, 213, 469–485, https://doi.org/10.1016/j.apenergy.2018.01.063, 2018.

Li, K., Bai, K., Ma, M., Guo, J., Li, Z., Wang, G., and Chang, N.-B.: Spatially gap free analysis of aerosol type grids in China: first retrieval via satellite remote sensing and big data analytics, ISPRS J. Photogramm. Remote Sens., 193, 45–59, https://doi.org/10.1016/j.isprsjprs.2022.09.001, 2022.

Liu, B., Guo, J., Gong, W., Shi, Y., and Jin, S.: Boundary Layer Height as Estimated from Radar Wind Profilers in Four Cities in China: Relative Contributions from Aerosols and Surface Features, Remote Sens., 12, 1657, https://doi.org/10.3390/rs12101657, 2020.

Liu, B., Ma, X., Guo, J., Wen, R., Li, H., Jin, S., Ma, Y., Guo, X., and Gong, W.: Extending the wind profile beyond the surface layer by combining physical and machine learning approaches, Atmos. Chem. Phys., 24, 4047–4063, https://doi.org/10.5194/acp-24-4047-2024, 2024.

Liu, L. and Stevens, R. J. A. M.: Vertical structure of conventionally neutral atmospheric boundary layers, Proc. Natl. Acad. Sci., 119, e2119369119, https://doi.org/10.1073/pnas.2119369119, 2022.

Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 4768–4777, 2017.

Martin, A., Weissmann, M., Reitebuch, O., Rennie, M., Geiß, A., and Cress, A.: Validation of Aeolus winds using radiosonde observations and numerical weather prediction model equivalents, Atmos. Meas. Tech., 14, 2167–2183, https://doi.org/10.5194/amt-14-2167-2021, 2021.

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

Optis, M., Bodini, N., Debnath, M., and Doubrawa, P.: New methods to improve the vertical extrapolation of near-surface offshore wind speeds, Wind Energy Sci., 6, 935–948, https://doi.org/10.5194/wes-6-935-2021, 2021.

Pelser, T., Weinand, J. M., Kuckertz, P., McKenna, R., Linssen, J., and Stolten, D.: Reviewing accuracy & reproducibility of large-scale wind resource assessments, Adv. Appl. Energy, 13, 100158, https://doi.org/10.1016/j.adapen.2023.100158, 2024.

Peng, Z., Tang, R., Jiang, Y., Liu, M., and Li, Z.-L.: Global estimates of 500 m daily aerodynamic roughness length from MODIS data, ISPRS J. Photogramm. Remote Sens., 183, 336–351, https://doi.org/10.1016/j.isprsjprs.2021.11.015, 2022.

Pernov, J. B., Gros-Daillon, J., and Schmale, J.: Comparison of selected surface level ERA5 variables against in-situ observations in the continental Arctic, Q. J. R. Meteorolog. Soc., 150, 2123–2146, https://doi.org/10.1002/qj.4700, 2024.

Quick, J., Murcia Leon, J. P., Kock, C. W., Servizi, V., Overgaard, N. S., Dimitrov, N., Kelly, M., Réthoré, P.-E., and Kim, T.: Wind speed vertical extrapolation model validation under uncertainty, Renewable Energy, 240, 122028, https://doi.org/10.1016/j.renene.2024.122028, 2025.

Ramon, J., Lledó, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J.: What global reanalysis best represents near-surface winds?, Q. J. R. Meteorolog. Soc., 145, 3236–3251, https://doi.org/10.1002/qj.3616, 2019.

Rennie, M. P., Isaksen, L., Weiler, F., de Kloe, J., Kanitz, T., and Reitebuch, O.: The impact of Aeolus wind retrievals on ECMWF global weather forecasts, Q. J. R. Meteorolog. Soc., 147, 3555–3586, https://doi.org/10.1002/qj.4142, 2021.

Šavli, M., de Kloe, J., Marseille, G.-J., Rennie, M., Žagar, N., and Wedi, N.: The prospects for increasing the horizontal resolution of the Aeolus horizontal line-of-sight wind profiles, Q. J. R. Meteorolog. Soc., 145, 3499–3515, https://doi.org/10.1002/qj.3634, 2019.

Seidel, D. J., Sun, B., Pettey, M., and Reale, A.: Global radiosonde balloon drift statistics, J. Geophys. Res.: Atmos., 116, D07102, https://doi.org/10.1029/2010JD014891, 2011.

Snoek, J., Larochelle, H., and Adams, R. P.: Practical bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, 2012.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 global reanalysis from 1940 to 2022, Q. J. R. Meteorolog. Soc., 150, 4014–4048, https://doi.org/10.1002/qj.4803, 2024.

Stoffelen, A., Pailleux, J., Källén, E., Vaughan, J. M., Isaksen, L., Flamant, P., Wergen, W., Andersson, E., Schyberg, H., Culoma, A., Meynart, R., Endemann, M., and Ingmann, P.: THE ATMOSPHERIC DYNAMICS MISSION FOR GLOBAL WIND FIELD MEASUREMENT, Bull. Am. Meteorol. Soc., 86, 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, Q. J. R. Meteorolog. Soc., 132, 1927–1947, https://doi.org/10.1256/qj.05.83, 2006.

Stoffelen, A., Benedetti, A., Borde, R., Dabas, A., Flamant, P., Forsythe, M., Hardesty, M., Isaksen, L., Källén, E., Körnich, H., Lee, T., Reitebuch, O., Rennie, M., Riishøjgaard, L.-P., Schyberg, H., Straume, A. G., and Vaughan, M.: Wind profile satellite observation requirements and capabilities, Bull. Am. Meteorol. Soc., 101, E2005–E2021, https://doi.org/10.1175/BAMS-D-18-0202.1, 2020.

Tan, Y., Xin, J., Qin, X., Ma, Y., Xu, J., Du, Y., Shang, C., Zhao, L., Tian, Y., Fan, Q., Zhou, X., Ren, Y., Wang, P., Hao, F., Yin, N., Ren, X., Peng, K., Jia, D., Pan, X., Li, J., Wang, Z., and Butt, A. Q.: Differences Among Thermal, Dynamic, and Material Boundary Layers Over Complex Terrain, J. Geophys. Res.: Atmos., 130, e2024JD043043, https://doi.org/10.1029/2024JD043043, 2025.

Tong, Z., Liu, B., Ma, X., Li, H., Wei, R., and Gong, W.: Estimation of boundary layer height from radar wind profiler by deep learning algorithms, IEEE Trans. Geosci. Remote Sens., 62, https://doi.org/10.1109/TGRS.2024.3434403, 2024.

Tong, Z., Liu, B., Ma, X., Zhang, H., Dong, H., Han, G., Ma, Y., and Gong, W.: Global atmospheric boundary layer wind speed profile dataset derived from aeolus observations from July 2020 to April 2023 (V1), Zenodo [dataset], https://doi.org/10.5281/zenodo.18286457, 2026.

Veers, P., Dykes, K., Lantz, E., Barth, S., Bottasso, C. L., Carlson, O., Clifton, A., Green, J., Green, P., Holttinen, H., Laird, D., Lehtomäki, V., Lundquist, J. K., Manwell, J., Marquis, M., Meneveau, C., Moriarty, P., Munduate, X., Muskulus, M., Naughton, J., Pao, L., Paquette, J., Peinke, J., Robertson, A., Sanz Rodrigo, J., Sempreviva, A. M., Smith, J. C., Tuohy, A., and Wiser, R.: Grand challenges in the science of wind energy, Science, 366, eaau2027, https://doi.org/10.1126/science.aau2027, 2019.

Wang, X., Jin, Y., Schmitt, S., and Olhofer, M.: Recent Advances in Bayesian Optimization, ACM Comput, Surv, 55, 287:1-287:36, https://doi.org/10.1145/3582078, 2023.

Watson, S.: Handbook of wind resource assessment, John Wiley & Sons, https://doi.org/10.1002/9781119055402, 2023.

Wiel, B. J. H. V. de, Moene, A. F., Steeneveld, G. J., Baas, P., Bosveld, F. C., and Holtslag, A. a. M.: A Conceptual View on Inertial Oscillations and Nocturnal Low-Level Jets, https://doi.org/10.1175/2010JAS3289.1, 2010.

Witschas, B., Lemmerz, C., Geiß, A., Lux, O., Marksteiner, U., Rahm, S., Reitebuch, O., and Weiler, F.: First validation of Aeolus wind observations by airborne Doppler wind lidar measurements, Atmos. Meas. Tech., 13, 2381–2396, https://doi.org/10.5194/amt-13-2381-2020, 2020.

Xiong, X., Guo, X., Zeng, P., Zou, R., and Wang, X.: A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization, Front. Energy Res., 10, https://doi.org/10.3389/fenrg.2022.905155, 2022.

Yang, X., Jiang, X., Liang, S., Qin, Y., Ye, F., Ye, B., Xu, J., He, X., Wu, J., Dong, T., Cai, X., Xu, R., and Zeng, Z.: Spatiotemporal variation of power law exponent on the use of wind energy, Appl. Energy, 356, 122441, https://doi.org/10.1016/j.apenergy.2023.122441, 2024.

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

Zhang, W., Guo, J., Miao, Y., Liu, H., Zhang, Y., Li, Z., and Zhai, P.: Planetary boundary layer height from CALIOP compared to radiosonde over China, Atmos. Chem. Phys., 16, 9951–9963, https://doi.org/10.5194/acp-16-9951-2016, 2016.