

Response to the reviewers

Dear Editor and Reviewers,

We thank the reviewers for the careful and constructive evaluation of our manuscript. We have addressed every comment point by point below and have revised the manuscript accordingly. The main revisions are:

- Clarifications of the data product, methodology, and statistical procedures. We clarified the two-stage use of radiosonde (RS) and ERA5 10 m wind speed, the along-track spatial and temporal sampling of Global-LTWind, the relationship between the Rayleigh- and Mie-based realizations, the bias-correction binning and fitting procedure, and the definition of the missing rate.
- Additional analyses and experiments. We added a direct evaluation of the effect of replacing RS 10 m wind speed with ERA5 10 m wind speed. We also added conditionally sampled case studies of low-level jets (LLJs) and strong low-level wind shear to demonstrate the ability of Global-LTWind to reproduce important lower-tropospheric wind structures.
- Revision of the boundary-layer terminology. We replaced “boundary-layer winds” with “lower-tropospheric winds” where appropriate because the reconstructed 100–2000 m range cannot be assumed to lie entirely within the atmospheric boundary layer at all locations and times.
- Improvement of concision and presentation. We removed or condensed repetitive passages, reduced the repetition of numerical values already provided in figures and tables, and shortened detailed descriptions of graphical features. The revised Results section now emphasizes the main findings, quantitative comparisons, and physical interpretation rather than restating figure contents point by point.
- Manuscript-wide terminology and consistency checks. We systematically checked and corrected product names, model names, figure and panel labels, table parameter names, captions, abbreviations, and typographical errors throughout the manuscript and Supplementary Material.

Because one figure was added within Sects. 4.3, the figures of the discussion paper have been renumbered. In this letter, figure numbers refer to the revised manuscript unless stated otherwise.

The line numbers cited in our replies refer to the revised manuscript with Track Changes enabled. In this response letter, reviewer comments are shown in black, our replies are shown in blue, and new or modified manuscript text is shown in bold blue.

On behalf of all co-authors,
Boming Liu and Xin Ma

Response to Referee #1 (RC1):

This manuscript presents Global-ABLWind, a global atmospheric boundary layer wind speed profile dataset derived from Aeolus L2C observations, radiosonde measurements, and ERA5 ancillary variables for July 2020 to April 2023. The stated aim is to reconstruct wind speed profiles at 100 m vertical spacing within the atmospheric boundary layer using an XGBoost-based framework with near-surface bias-correction terms and upper-level Aeolus dynamical constraints. The dataset is potentially valuable because global high-vertical-resolution ABL wind profiles are difficult to obtain from existing observing systems, and the combination of Aeolus wind lidar information with radiosonde and reanalysis constraints addresses a relevant community need. This manuscript can be published after considering the following comments concerning clarifications or extensions:

Thank you for your positive comments, we are pleased that the main messages and importance of the study come across. The comments below have been extremely helpful to refine our work. The line numbers mentioned in our responses refer to the revised version with Track Changes enabled for review.

Comment R1.1

1. 3.2.2 states that 10 m RS wind speed is used as the bottom constraint during model construction because ERA5 10 m wind correlates weakly with 100 m RS wind speed. Sect. 4.3 then states that the global product uses ERA5 10 m wind speed as the alternative input where RS is unavailable. The reported model-validation metrics are derived from a configuration using RS 10 m wind, whereas the released global product depends on ERA5 10 m wind. The rationality of this substitution needs further clarification.

We sincerely thank the reviewer for raising this important concern. We agree that this issue requires clarification because the validation metrics reported for the observation-constrained configuration should not be interpreted as being directly identical to the performance of the final global-product configuration. The use of RS 10 m wind speed during model construction was intentional. This configuration allows the model to learn the vertical wind-speed relationship under an observationally constrained setting and avoids introducing additional uncertainty from the reanalysis into the training relationship. In our matched dataset, the correlation between ERA5 10 m wind speed and RS 100 m wind speed is only approximately 0.47, indicating a relatively weak correspondence between the ERA5 near-surface wind and the observed wind structure at the target height.

For global inference, however, RS observations are spatially sparse and cannot provide continuous input over most of the world. We therefore use ERA5 10 m wind speed as a globally available surrogate. This strategy is consistent with the broader transfer-learning concept adopted in previous wind-profile studies. For example, Yu and Vautard (2022) applied the trained model to large-scale gridded regional climate model outputs. Frech et al. (2025) similarly trained a wind-profile model using limited lidar observations and subsequently applied the trained model to a spatially continuous satellite-derived surface-wind product. Although these studies do not employ exactly the same RS-to-ERA5 substitution as our study, they provide methodological support for using observation-constrained model development followed by large-scale application with spatially continuous datasets.

Importantly, we also directly evaluated the effect of this substitution. As shown in Sect. 4.3, the global product generated using ERA5 10 m wind speed has a mean difference of only -0.58 m s^{-1} at 100 m relative to RS observations. This indicates that the transition from the RS-constrained training configuration to the ERA5-based global inference configuration does not introduce a great systematic bias at 100 m. We further emphasized this change in the corresponding sections of the manuscript, such as Sects. 3.2.2 (Lines 301-313):

“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 choice is intended to provide an observation-constrained lower boundary. In our matched dataset, the correlation between the 10 m ERA5 wind speed and the 100 m RS wind speed is relatively weak ($R = 0.47$), which may introduce additional uncertainty into the near-surface-to-hub-height relationship. However, for global product generation, RS observations are spatially sparse and cannot provide continuous lower-boundary inputs over most regions. Therefore, the trained model is subsequently applied using the globally available ERA5 10 m wind speed as an alternative input where RS observations are unavailable. This represents a two-stage strategy in which high-quality observations are used to constrain model development, while spatially continuous gridded data are used for large-scale application. The effect of this input substitution is further evaluated using the global-product configuration in Sect. 4.3.

Reference:

- 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.
- 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.

Comment R1.2

- 4.3 states that Global-ABLWind maintains the same spatial and temporal resolution as Aeolus L2C. The product is therefore along Aeolus observation profiles rather than a globally gridded continuous field. Several figures use global map shading that visually resembles gridded continuous coverage. The author should provide a clearer explanation in the manuscript.

Thanks for your advice. We sincerely thank the reviewer for pointing out this important issue regarding the spatial representation of Global-LTWind (Renamed from Global-ABLWind). Global-LTWind is constructed from Aeolus L2C observations and therefore retains the original spatial and temporal sampling characteristics of the Aeolus measurements. Thus, the global maps should be interpreted as a visualization of the spatial distribution of along-track profile observations, rather than as a globally continuous gridded wind field. To avoid this potential misunderstanding, we have revised Section 4.3 to explicitly clarify the spatial sampling characteristics of Global-LTWind (Lines 443-445):

“The generated Global-LTWind product maintains the same spatial and temporal

sampling characteristics as L2C. Specifically, the product consists of wind profiles associated with individual Aeolus observation locations along the satellite ground tracks.”

We have also revised the relevant figure captions (including Figure 11 and 14). In particular, the colored shadows in the global map are values aggregated from all products within each grid. We appreciate the reviewer for highlighting this issue, as the clarification improves the transparency of the product's spatial characteristics and prevents the global maps from being misinterpreted as a spatially continuous wind field.

Comment R1.3

3. The bias-correction method relies on quadratic relationships between binned Ray/Mie wind biases and selected variables. The statistical basis of these relationships is unclear because the manuscript does not specify how bins were defined, how sample sizes vary by bin, whether the R^2 values are based on raw samples or binned means, or whether these fitted corrections were estimated on independent data from the validation folds.

Thank you for this helpful comment. We agree that the previous manuscript did not sufficiently describe how the relationships between the Rayleigh/Mie wind bias and the selected variables were constructed. We have therefore clarified the binning procedure, the criterion for retaining bins, the basis of the reported R^2 values, and the purpose of these quadratic relationships. For each of the five variables, its normal observed range in the matched dataset was divided into 10 equally spaced intervals. Within each interval, we calculated the sample number and the distribution of the corresponding wind-speed bias, including the mean and quartiles. To avoid drawing conclusions from sparsely sampled intervals, only intervals containing more than 30 samples were retained for the subsequent statistical analysis. The quadratic relationships were then fitted to the mean bias of the retained intervals. The matched dataset contains a very large number of individual samples, for which substantial and case-to-case variability is present. Directly fitting all individual samples would therefore emphasize high-frequency sample-level variability and could obscure the systematic dependence of the bias on the explanatory variable. Aggregating the samples within predefined intervals provides a robust representation of the mean bias response and allows the underlying overall relationship to be identified more clearly. Thus, the fitted quadratic curves are intended to characterize the systematic trend of the product bias rather than to predict the bias of individual observations. Accordingly, the R^2 values reported in Figure. 5(b–e) are calculated from the retained interval-mean biases and the corresponding values predicted by the quadratic functions. They therefore quantify how well the quadratic functions describe the variation in the mean bias across the sampled intervals, rather than the variability of the individual observations.

Regarding the independence from the validation folds, we clarify that these quadratic relationships were derived from the complete matched dataset because their purpose is to diagnose and characterize the systematic bias structure of the Aeolus L2C product over the full observation period, rather than to constitute an independently trained machine-learning predictor whose generalization performance is evaluated through cross-validation. The quadratic relationships are therefore treated as empirical error-characterization functions describing the systematic behavior of the L2C bottom-layer bias. They were not selected or optimized according to the validation performance of the subsequent machine-learning model. We have clarified this distinction in the revised manuscript (Sect. 3.2.1). In particular, we now

explicitly state that the binning and quadratic fitting are used to characterize the overall systematic bias pattern identified by SHAP.

Comment R1.4

4. The term “missing rate” appears throughout the manuscript, but it is not always clear whether it refers to missing height levels within an observed profile, missing Aeolus profiles, missing valid retrievals, missing model outputs, or missing map bins after aggregation.

Thank you for pointing out this ambiguity. We have clarified the definition and calculation of the missing rate in the revised manuscript. In this study, the missing rate is calculated based on the coverage of predefined height levels within each valid Rayleigh wind profile, rather than the number of missing profiles, invalid retrievals, model outputs, or spatial grid cells. Specifically, the analysis domain is divided into 100 m height levels from 100 to 2000 m (100, 200, 300, ..., 2000 m). For each valid Rayleigh wind profile, all valid Rayleigh wind speed bins are examined. A given 100 m height level is considered covered if it is overlapped by at least one valid wind-speed bin; otherwise, that height level is considered missing. The missing rate is then calculated as the total number of missing height levels across all valid profiles divided by the corresponding total number of expected height levels. We have added this definition to the manuscript to make the statistical basis of the reported missing rates explicit, such as Lines 143-145:

“For each valid Rayleigh wind profile, and a given height level is considered covered if it is overlapped by at least one valid wind speed bin; otherwise, it is considered missing. The missing rate is then calculated as the number of missing height levels divided by the total number of expected height levels.”

Comment R1.5

5. The relationship between the Ray-based and Mie-based products is not sufficiently explicit. Whether they use the same processing procedure is not clearly stated. The main text is based on Ray wind, while the supplement presents Mie-derived results. However, the manuscript does not make sufficiently clear whether the released dataset should be regarded as one primary Ray-based product with a secondary Mie-based companion, two parallel products, or two alternative realizations of the same product. This distinction is important for users deciding which files to use under different atmospheric or cloud/aerosol conditions.

Thank you for pointing out that the relationship between the Rayleigh- and Mie-based products was not sufficiently explicit. We have clarified their relationship, data sources, and respective characteristics in the revised manuscript. The Rayleigh- and Mie-based Global-LTWind products are two parallel realizations generated using the same XGB-Wind reconstruction framework. For each Aeolus L2C profile, the model combines the corresponding surface wind speed and meteorological and terrain information with an upper-level Aeolus wind constraint to reconstruct the wind profile from 100 to 2000 m. The only difference in the Aeolus constraint between the two realizations is the source of the 2100 m wind speed: the Rayleigh-based realization uses the wind from Rayleigh channel, whereas the Mie-based realization uses the Mie wind. Thus, they represent two realizations of the same Global-LTWind reconstruction

framework rather than a primary product and a secondary companion product or two independent modeling approaches.

The two realizations nevertheless retain different observational characteristics from the underlying Aeolus Rayleigh and Mie retrievals. Rayleigh retrievals are based on molecular backscatter, whereas Mie retrievals exploit particulate backscatter from clouds and aerosols. The stronger particulate backscatter generally allows Mie retrievals to use shorter averaging lengths and therefore provides denser horizontal sampling and a larger number of profiles. Although Mie inversion is often limited by more stringent conditions of particle backscattering, resulting in fewer effective wind observations in the profile, this drawback is addressed by the proposed XGB-Wind framework. In contrast, Rayleigh retrievals generally provide coarser horizontal resolution because of the weaker molecular backscatter signal, but has a higher valid rate. These differences in the source observations are consequently reflected in the spatial sampling and profile availability of the corresponding Global-LTWind realizations.

The two realizations provide complementary choices for different applications. The Rayleigh-based product is used as the main realization in this study because the Rayleigh upper-level wind generally provides a more accurate wind constraint. The Mie-based product is particularly useful when the denser spatial sampling associated with Mie retrievals is advantageous and valid Mie observations are available under suitable cloud and aerosol conditions. The availability and quality of Mie retrievals are therefore more dependent on the presence and characteristics of particulate backscatter, whereas Rayleigh retrievals have broader applicability but coarser spatial sampling. We have clarified these relationships and characteristics in the revised manuscript to make the distinction between the two Global-LTWind realizations clear to potential users (Lines 237-241):

“Both realizations use the same XGB-Wind reconstruction framework. The Rayleigh-based realization uses the Rayleigh wind speed at 2100 m as the upper-level constraint, whereas the Mie-based realization uses the corresponding Mie wind speed. Therefore, the two realizations differ primarily in the source of the upper-level Aeolus constraint and the associated observational sampling characteristics. The Mie-based results generally provide denser spatial sampling, whereas Rayleigh-based results have broader applicability but relatively sparser sampling.”

Comment R1.6

6. The inclusion of R and RMSE in a hyperparameter table is ambiguous because they are performance metrics rather than hyperparameters. The RMSE units are not shown in the table.

We agree that R and RMSE are performance metrics rather than hyperparameters, and that the unit of RMSE should be explicitly indicated. The hyperparameter columns describe the configuration used for each model, while R and RMSE are now explicitly identified as test-set evaluation metrics. The unit of RMSE (m s^{-1}) has also been added to the table. The table title has been revised accordingly to avoid referring to all listed variables as hyperparameters. Similar modifications are also applied to the supplement.

Comment R1.7

7. Terminology is inconsistent in several places. The most conspicuous example is Fig. 12 and

Fig. S10, where the legend reads “Global-PBLWind” while the manuscript and data record describe “Global-ABLWind”. The title of section 3.2 is “XBG-Wind”, while elsewhere in the manuscript it is “XGB-Wind”. Fig. 8 also has panel-lettering inconsistencies between the visual layout and caption. Tables 1 and S1 contain the misspelling “learing_rate.” The title of the subgraph in Figure 2 does not match the content. Please carefully review the entire manuscript.

Thank you for carefully identifying these inconsistencies. We have carefully reviewed the entire manuscript, including the main text, figures, figure captions, tables, and supplement, and corrected the inconsistencies. Specifically, “Global-PBLWind” in Figs. 12 and S10 has been corrected to “Global-LTWind” (This terminology more accurately reflects the actual vertical coverage of our product without implying that all retrieved levels are necessarily within the ABL). “XBG-Wind” in Section 3.2 has been corrected to “XGB-Wind” to ensure consistency throughout the manuscript. The panel lettering in Fig. 8 has been checked and corrected to match the figure layout and caption. The misspelling “learing_rate” in Tables 1 and S1 has been corrected to “learning_rate” and the title of the subgraph in Figure. 2 has been corrected to accurately reflect its content.

In addition, we conducted a systematic consistency check of model names, dataset names, figure and panel labels, table terminology, parameter names, abbreviations, and figure captions throughout the manuscript and Supplementary Material to identify and correct other similar errors.

The revised Figure 12 (Lines 514-518):

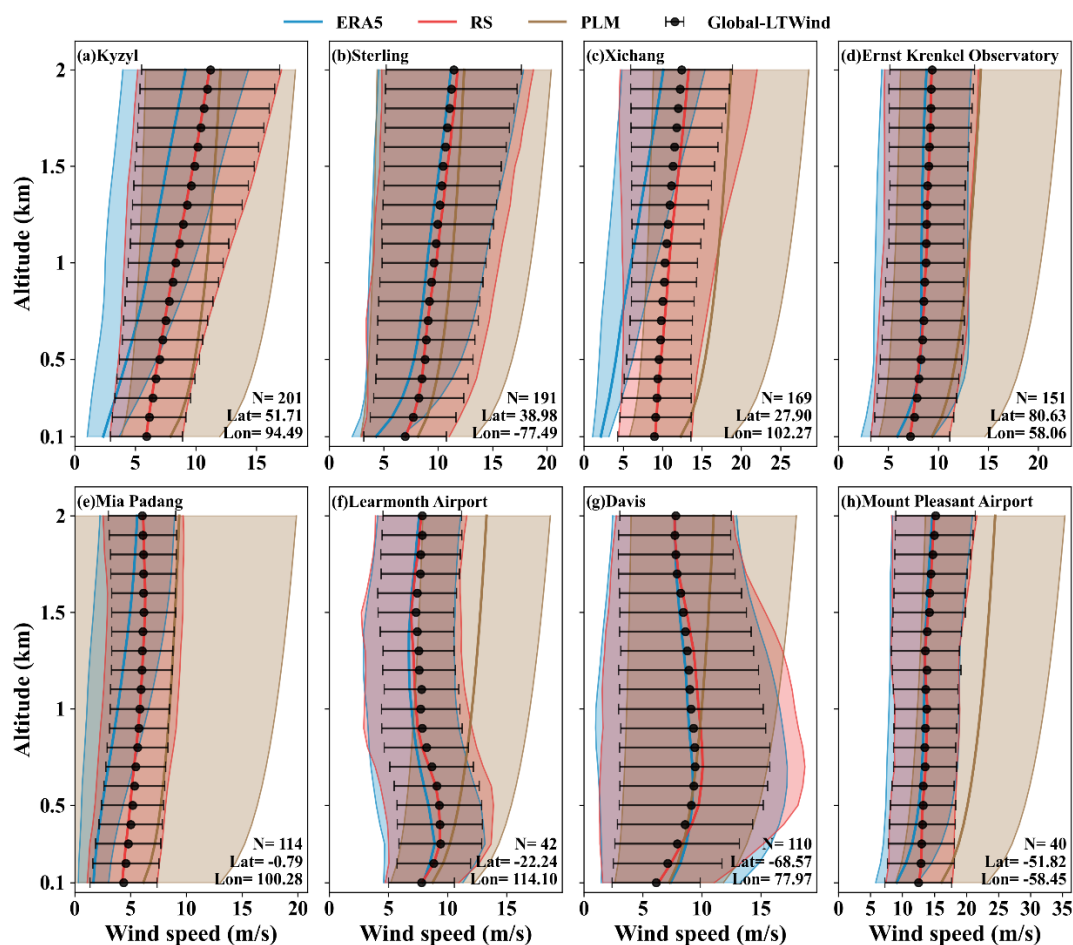


Figure 12. Vertical distributions of average wind speed as derived from multiple methods at (a) Kyzyl, (b) Sterling, (c) Xichang, (d) Ernst Krenkel Observatory, (e) Mia Padang, (f) Learmonth Airport, (g) Davis, and (h) Mount Pleasant Airport. Blue, red, brown, and black represent ERA5, RS, PLM, and Global-LTWind, respectively. The colored shading or error bars denote the standard deviation. The specific geographical location of the sites can refer to the red-edge dots in Fig. 11.

Response to Referee #2 (RC2):

Summary statement:

This study proposes a valuable addition to globally available airflow data for many potential applications. It produces and documents a dataset of horizontal wind speed at 100m vertical increments from the surface to 2000m asl with global coverage from the Aeolus mission's active wind-lidar observational system for the period July 2020 to April 2023. Machine learning techniques, more specifically extreme gradient boosting (XGB,) is used to improve currently available products in the lowest layers near the surface by learning from the space and time matched global radiosonde dataset (IGRA) regarding data availability and precision. In addition to describing their methodology and reporting the statistics of their 'Global-ABLWind' product, an attempt is made to describing the time-averaged circulations/ phenomena found in the lower troposphere for the observational period.

The methodology is well described and thought-out, I enjoyed reading most sections of the manuscript and believe that the product has a substantial enough merit to warrant publication in ESSD. I only have a few scientific, and some editorial comments and suggestions. The question of how and if to address the latter is at the discretion of the authors, but I believe it would benefit the publication.

Overall, I recommend acceptance of the manuscript following minor revisions addressing the comments listed below.

We thank the reviewer for taking the time to carefully review our manuscript and for their positive summary. We have amended the revised manuscript to include the suggestions provided by the reviewer and detail our responses to each comment below.

Comment R2.1

General scientific comments:

1. Uncertainty in radiosonde reference data: Judging from Figs. 3c and 8b/ 8d and the accompanying text I am convinced that the XGB method significantly improves the match between the L2C-Rayleigh winds and the radiosondes (RS). However, the authors neither investigate nor mention the uncertainty of the RS profiles in the lowest hundreds of meters above ground. Often, RS profiles even do not contain data near the surface and/or observations are discarded in post-processing because of potential artifacts due to observer effects (when the observer releases the balloon), local obstruction/ deflection of the airflow by buildings and topography (such as hills, mountain chains etc), and delays in receiving the radiocommunication data from the ascending balloon. In other words: the RS data itself, which the authors here consider as the reference or 'truth', suffer from the largest observational error in the exactly the lowest heights approx. < 200 to 300m. I realize that the authors carefully screened the RS profiles and discarded potentially problematic data, but even if RS wind speeds are reported and used, the uncertainty near the surface is substantial. I would like to see some discussion of this i) measurements uncertainty by e.g. a comparison with tower locations (if existent in the vicinity) or ground-based wind lidars. I realize that these comparison can only be done for a limited time and spatial extent of the Global-ABLwinds, and ii) how this uncertainty compares to the improvement from the

LC2-Ray to XGB winds.

Thank you for prompting us to reconsider this point carefully. We agree that radiosonde (RS) measurements in the lowest few hundred meters can have additional uncertainties associated with balloon release, local flow disturbance, balloon motion, spatial representativeness, and the temporal delays. To explicitly assess the reliability of the RS reference data used in this study, we conducted an additional comparison between the RS profiles and long-term ground-based Doppler lidar observations at three ARM sites for which sufficiently long lidar records were available. The corresponding radiosonde and lidar sites and their spatial separations are summarized in Table R1. Two of the sites have close spatial correspondence, whereas the ENA lidar and radiosonde are separated by more than 80 km, providing an additional indication of the influence of spatial representativeness.

Table R1. Three ARM lidar sites and their nearest matching RS sites, along with their distances.

Lidar station	Geographic coordinates	The nearest RS station	Geographic coordinates of the RS station	Distance
NSA (North Slope of Alaska)	71.32°N 156.62°W	USM00070026	71.28°N 156.78°W	7.10 km
SGP (Southern Great Plains)	36.61°N 97.49°W	USM00074646	36.60°N 97.50°W	1.37 km
ENA (Eastern North Atlantic)	39.09°N 28.03°W	POM00008508	38.77°N 27.09°W	88.93 km

The profile results of comparisons is shown in Fig. R1. At the two closely located sites, the RS and lidar observations show generally good consistency within the lowest 300 m, with high correlations and relatively small mean biases. The agreement at NSA and SGP indicates that, despite some discrepancies near the surface, the RS observations do not exhibit a pronounced degradation within the lowest 300 m when compared with an independent wind-profiling instrument at closely collocated sites. In contrast, the ENA comparison shows slightly larger discrepancies. However, the ENA lidar and radiosonde sites are separated by more than 80 km, so the larger differences cannot be attributed solely to radiosonde measurement uncertainty and are likely also affected by spatial representativeness differences between the two observing systems. Overall, these comparisons support the use of the quality-controlled RS profiles as reference observations.

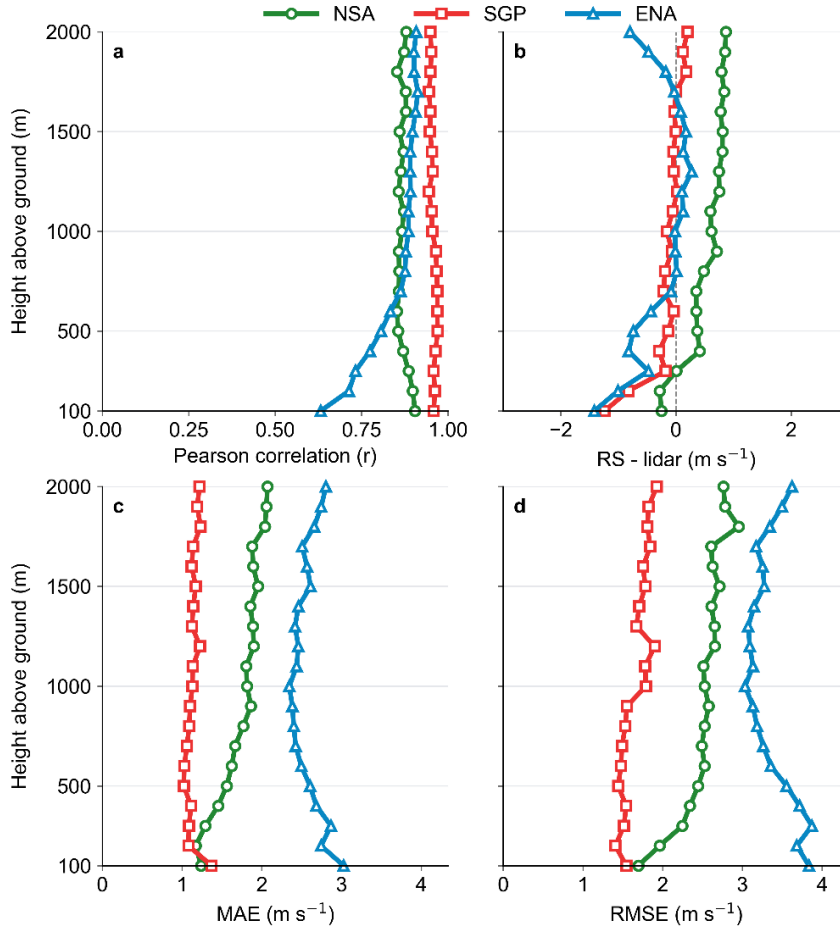


Figure R1. Comparison of wind speed profiles between three lidar stations and their corresponding RS stations.

We further examined the XGB-Wind reconstruction against the ground-based lidar observations. After matching Rayleigh and RS, only SGP sites have enough XGB-Wind reconstructed profiles that can be compared with lidar (Fig. R2). XGB-Wind shows good consistency with the independent lidar observations throughout the 100–2000 m layer, with correlations generally exceeding 0.8 and reaching approximately 0.95 at several levels. The biases remain relatively small compared with the overall wind-speed variability, although the reconstruction tends to underestimate lidar wind speeds at most levels. This independent comparison is important because the lidar was not used as the reference for training or the primary evaluation of XGB-Wind. It therefore provides an additional check that the good performance of XGB-Wind is not solely a consequence of fitting to the RS reference.

However, a direct height-by-height comparison between the original L2C-Rayleigh product and the lidar at the same 100 m vertical resolution is not possible because the original L2C-Rayleigh product has much coarser and incomplete vertical sampling in the lowest 2 km. In the same dataset of SGP, only approximately three valid Rayleigh bins are typically available below 2 km, centered approximately around 100–200, 500–700, and 1100–1400 m. Therefore, the lidar comparison cannot provide a like-for-like 100 m resolution evaluation of the original L2C-Rayleigh product. According to the comparison results against RS on the original bin in the manuscript, the correlation increases from 0.90 to 0.92 and the RMSE decreases from 2.23

to 1.94 m s^{-1} , corresponding to a reduction of 0.29 m s^{-1} (about 13%). We agree that this improvement should be interpreted in the context of RS uncertainty. In particular, the observed RS–lidar differences at the closely collocated sites are larger than the improvement. However, these differences should not be directly regarded as RS measurement uncertainty because they also include lidar uncertainty and spatial and temporal representativeness differences. More importantly, the improvement is not supported only by the RS-based evaluation. The independent XGB-Wind and lidar comparison shows good agreement over the vertical range, including the lowest 300 m. Thus, while the exact magnitude of the improvement is inevitably affected by uncertainty in the RS reference, the independent lidar comparison provides additional evidence that the improvement reflects a genuine improvement in the reconstructed wind profiles rather than an artifact of the radiosonde observations.

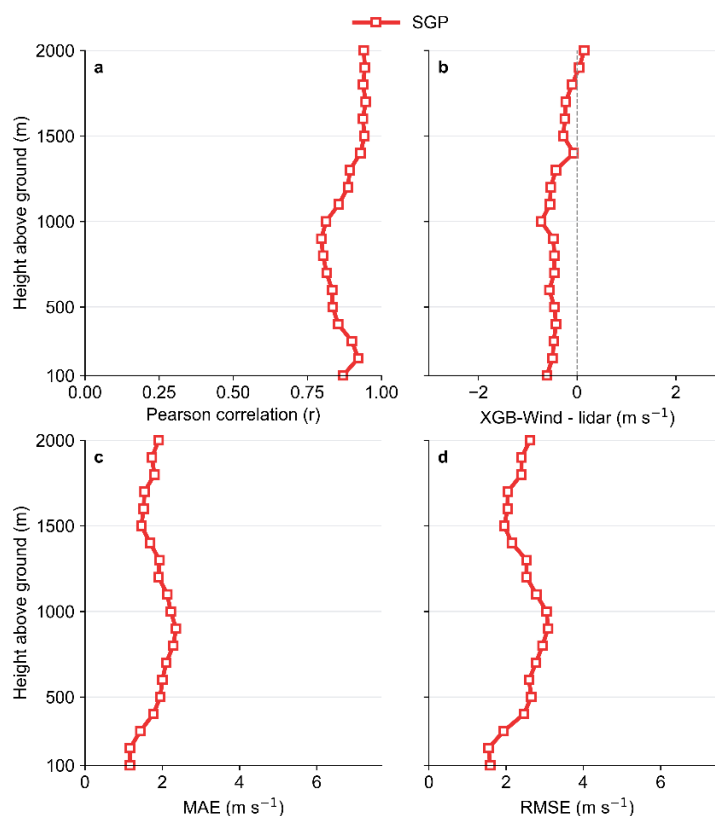


Figure R2. Comparison of wind speed profiles from SGP lidar and corresponding XGB-Wind. Only SGP have XGB-Wind reconstructed profiles after collocation between Rayleigh and RS.

Comment R2.2

2. Wind speed vs wind direction: While the focus of the study is on wind speed, the direction of the airflow and its vertical directional shear is also important for many potential applications including those in the wind energy sector. Observing height-dependent wind direction in the lowest troposphere is much more difficult than solely the magnitude of the wind vector, but I would appreciate if the authors discussed/ showcased the usefulness of their method to also compute improved wind directions having in mind applications for which directional shear is important (cyclogenesis, orographic effects, ocean-air interaction, wind energy, aviation meteorology etc).

We acknowledge this valuable suggestion. The height-dependent wind direction and directional shear (wind veer or backing) are important for applications such as wind-energy assessment, aviation meteorology, and the analysis of atmospheric and orographic processes. In particular, vertical directional shear across the rotor can affect wind-turbine inflow and power production (Mata et al., 2024)

Reconstructing wind direction is substantially more challenging than simply extending a wind-speed reconstruction framework. Previous studies have identified several factors underlying this difficulty. First, wind direction is a circular variable with a 0° – 360° periodicity, and its high temporal and spatial variability makes it more difficult to model using conventional regression approaches. A systematic review of machine-learning studies found that wind-direction study remains much less developed than wind-speed study, with the vast majority of reviewed studies focusing exclusively on wind speed (Alves et al., 2023). Second, for vertical-profile applications, wind direction involves the additional problem of reconstructing wind veer/backing with height. Observational studies have shown that vertical wind-direction changes are strongly influenced by atmospheric stability, Coriolis effects, low-level jets, and other atmospheric and surface processes, making their vertical structure distinct from that of wind-speed shear (Kelly and van der Laan, 2023; Xue et al., 2026). All of this proves the great challenge of wind direction reconstruction.

In response to the reviewer's suggestion, we conducted an additional experiment to extend our framework to the joint reconstruction of the horizontal wind components (u and v). The resulting u and v predictions achieved R values of 0.89 and 0.82 and RMSEs of 2.37 and 2.45 m s^{-1} , respectively. However, when converted to wind speed and direction, the reconstructed wind speed just achieved $R = 0.86$ and $\text{RMSE} = 2.23 \text{ m s}^{-1}$, while the wind-direction reconstruction yielded $R = 0.72$ and an RMSE of 46° . Moreover, the joint reconstruction substantially degraded the wind-speed performance compared with the wind-speed-only XGB-Wind model and introduced a systematic wind-speed bias (-1.40 m s^{-1}).

These results indicate that directly extending the current wind-speed reconstruction framework to wind direction is not yet sufficiently robust for producing a reliable global directional-wind product. We therefore do not include the experimental wind-direction reconstruction in the released Global-LTWind dataset. Instead, we regard height-dependent wind-direction and veer reconstruction as an important direction for future development. Future work will investigate dedicated approaches that explicitly account for the circular nature of wind direction and the physical mechanisms governing vertical veering and backing, with the aim of providing reliable wind vector profiles for applications in wind energy and atmospheric and aviation meteorology.

Reference:

- Alves, D., Mendonça, F., Mostafa, S. S., and Morgado-Dias, F.: The potential of machine learning for wind speed and direction short-term forecasting: a systematic review, *Computers*, 12, 206, <https://doi.org/10.3390/computers12100206>, 2023.
- Kelly, M. and van der Laan, M. P.: From shear to veer: theory, statistics, and practical application, *Wind Energy Science*, 8, 975–998, <https://doi.org/10.5194/wes-8-975-2023>, 2023.
- Mata, S. A., Pena Martínez, J. J., Bas Quesada, J., Palou Larrañaga, F., Yadav, N., Chawla, J. S., Sivaram, V., and Howland, M. F.: Modeling the effect of wind speed and direction shear

on utility-scale wind turbine power production, *Wind Energy*, 27, 873–899, <https://doi.org/10.1002/we.2917>, 2024.

Xue, Y., He, J., Chan, P., and Li, Q.: Multi-source observational study of wind shear and veer at Hong Kong international airport, *Urban Climate*, 65, 102758, <https://doi.org/10.1016/j.uclim.2025.102758>, 2026.

Comment R2.3

3. Statistics of Global-ABLWind product: While I enjoyed reading almost all sections, I found Fig. 12 and section 4.4 the weakest parts of the manuscript and wonder if they can be replaced by more meaningful analyses. For Fig. 12 and accompanying text in Section 4.3, it is entirely unclear where these stations are located, why they were selected, and how representative these annual (?, averaging interval never described) are at a global scale. Winds change diurnally, synoptically (3 to 5 days in mid latitudes), and seasonally, so presenting a crude comparison of averaged profiles is not helpful in estimating the merit of the Global-ABLWind product. Only when RS observations deviate substantially from those of Ray L2C and other products for the lowest heights near the surface, the improvements by XGR should be largest. So, instead of a bulk analysis, I suggest replacing the bulk statistics with conditionally sampled ensemble averages across methods when i) speed and directional shear in the lowest heights are most important to the atmospheric science community (examples see comment B), or ii) the physical representation and interpretation of the found circulations/ phenomena (e.g. monsoons) have changed before and post application of the XGR method. In other words, I suggest the authors showcase the improvements in their data product to the reader in a more convincing manner using concrete case studies (with limited space and time domains), instead of global maps and profiles.

Thank you for this constructive suggestion. We agree that simple averaged wind profiles cannot fully demonstrate the advantages of Global-LTWind under specific atmospheric conditions. Therefore, we have added several condition-specific case studies focusing on low-level jets, strong vertical wind shear, monsoon circulation, and typhoon-related winds. These new analyses are designed to directly demonstrate the ability of Global-LTWind to represent physically important and highly variable lower-tropospheric wind structures, and thus provide a more convincing evaluation of its scientific value beyond bulk statistics.

The added content in the revised manuscript is located on Lines 519-548:

“To further evaluate the ability of Global-LTWind to represent dynamically important lower-tropospheric wind structures, we conducted conditionally sampled case studies of low-level jets (LLJs) and strong vertical wind shear in the 100-500m layer (Fig. 13). LLJs were identified following the criteria of Lampert et al. (2016), in which the wind-speed maximum must be at least 2 m s^{-1} or 25% higher than the subsequent minimum, while small secondary minima associated with increases of less than 1 m s^{-1} are disregarded. Strong-shear profiles were selected from the top 20% of the radiosonde wind-speed gradients between 100 and 500 m, after excluding profiles classified as LLJs. These products have substantially different vertical sampling in the 100–2000 m layer. Global-LTWind and PLM provides values at 20 levels spaced 100 m apart, whereas the Aeolus profiles generally provide only three to four valid bins and ERA5 provides

approximately eight levels. For the LLJ case (Fig. 13a), Global LTWind accurately reproduced the non-monotonic structure of the local maximum wind speed in the lower troposphere displayed by the RS profile, including the location and magnitude of the low-level wind-speed maximum. The coarser sampling of Aeolus and ERA5 provides less detail around this localized feature and the illustrated Aeolus profile has only one valid bin. Both of them show larger deviations from the RS observations. The PLM cannot reproduce the observed LLJ structure and instead produces a monotonically increasing profile, resulting in substantially larger errors. Across all 1846 LLJ profiles (Fig. 13b), Global-LTWind has the lowest MAE and RMSE in the 100–500 m layer. These results demonstrate that Global-LTWind can reproduce local peak structures that are not adequately captured by the other evaluated methods. The strong-shear case (Fig. 13c) provides a complementary test under a different vertical-wind regime. The RS profile shows a marked increase in wind speed within the 100–500 m layer, indicating enhanced vertical shear near the surface. Global-LTWind follows this enhanced vertical gradient closely, whereas the original Aeolus and ERA5 profiles exhibit larger departures from the observed wind-speed increase. The PLM method substantially underestimates the wind speed at upper levels and fails to reproduce the great magnitude of the observed shear. The statistical metrics consistently indicate lower errors for Global-LTWind than for the alternative products. Together, the LLJ and strong-shear cases highlight the superior performance of Global-LTWind over the other methods in reproducing complex lower-tropospheric wind structures at finer vertical spacing.

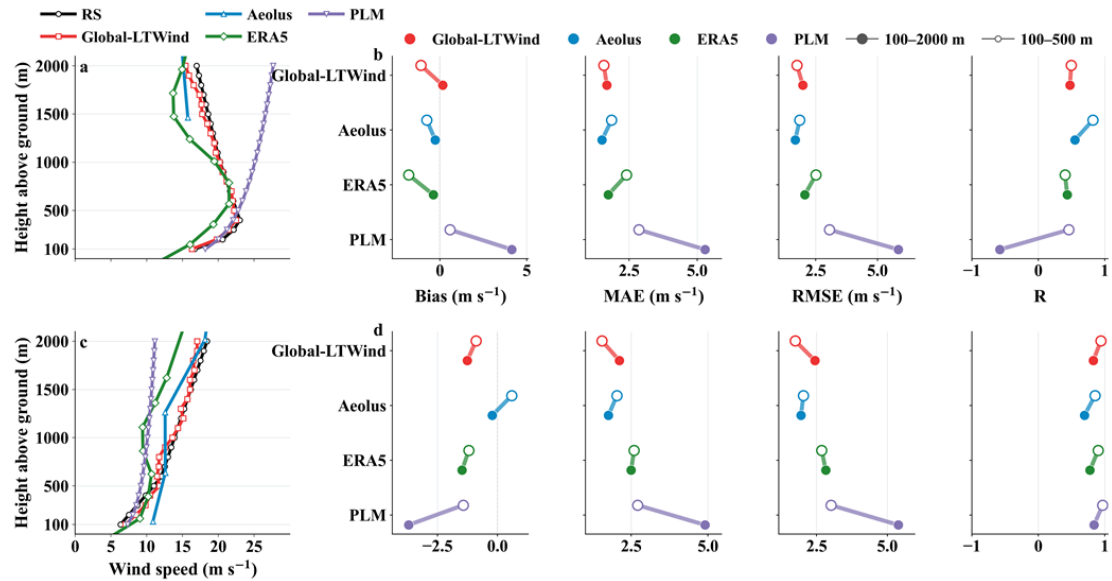


Figure 13. Performance of Global-LTWind under LLJ and strong low-level wind-shear conditions. (a) and (c) Vertical wind-speed profiles from RS, Global-LTWind, Aeolus, ERA5, and PLM for a LLJ case and a strong-shear case, respectively. (b) and (d) The bias, MAE, RMSE, and R across all LLJ and strong-shear profiles, respectively. Filled and open markers represent statistics over 100–2000 m and 100–500 m, respectively. Global-LTWind, RS and PLM are shown at 100 m intervals while Aeolus and ERA5 are shown at their native vertical sampling levels.”

Meanwhile, we retain the original multi-station profile comparison because it provides a

complementary assessment that cannot be obtained from individual case studies. The objective of this analysis is not to claim that eight stations constitute a statistically representative global sample, nor to characterize diurnal, synoptic, or seasonal variability. Instead, it serves as a geographically distributed consistency check of the climatological vertical wind structure reconstructed by Global-LTWind. Because Global-LTWind is derived from along-track Aeolus observations, such a comparison provides a complementary assessment of whether the reconstructed vertical profiles remain physically consistent with radiosonde observations across different geographical and climatic environments. We therefore consider this analysis useful as a baseline product-level evaluation.

And we have also clarified the station-selection procedure, the geographical locations, and the averaging period. The stations were first screened according to the number of available RS profiles to ensure sufficient observational sampling. From the stations with adequate sampling, we then selected sites to maximize geographical coverage, while avoiding excessive concentration in a limited longitude or latitude range. The available RS stations are concentrated mainly in two longitudinal regions, approximately 60°W–120°W and 60°E–120°E, with more stations in the latter region. We therefore selected two stations from the former region and six from the latter to provide broader latitudinal coverage. Within these regions, the selected stations were distributed as evenly as possible in latitude. Thus, the eight stations were chosen as geographically distributed examples with sufficient observational sampling, rather than as a statistically representative global sample. Their locations are now explicitly identified in the main text and figure caption, with the corresponding sites marked in Fig. 11. We also modified Fig. 11 to make these eight sites more prominent in the figure. We further clarified the averaging period. The plotted profiles are obtained by averaging all available profiles from the entire temporal coverage of the Global-LTWind dataset at each selected station. Therefore, they represent the mean vertical wind structure over the full study period. The corresponding discussion in Section 4.3 and the caption of Fig. 12 have also been revised accordingly.

Comment R2.4

General editorial comments:

1. Redundant text: While the language is appropriate for an international audience and terminology is correctly used, the current drafts suffers from substantial repetitions, which are a bit annoying. To give concrete examples: i) Section 4.3 starts with a summary of the key methodological steps (Ln 398 to 403), ii) Ln 355-358 repeats information mentioned before; iii) Ln 295 to 302: mere repetition. These are just a few examples. I suggest making the writing more concise, and shorten the manuscript. I encourage the authors to go through the manuscript and reduce the redundant text passages to the absolute minimum.

We sincerely thank the reviewer for this constructive comment. We agree that some parts of the manuscript contained redundant descriptions, particularly where methodological procedures or key findings were restated in multiple sections. We have therefore carefully reviewed the entire manuscript and substantially streamlined the text. Specifically, we removed or condensed repetitive descriptions in Sect. 3 and 4, including the passages highlighted by the reviewer (Lines 295–302, 355–358, and 398–403 in the original manuscript). In particular, the opening summary of the methodological procedure in Section 4.3 has been shortened, and

information that had already been sufficiently described in the Methods section is no longer repeated in the Results section. Similar repetitions throughout the manuscript were also removed or condensed. During this revision, we aimed to retain the information necessary for understanding the methodology, interpreting the results, and reproducing the main workflow, while avoiding repeated explanations of the same procedures or findings.

Comment R2.5

2. Writing style: The authors spend a lot of text on restating the statistics mentioned in the figures/ tables, and describe the figures (e.g. Ln 335-358,...) in unnecessary detail instead of summarizing and interpreting the findings. I suggest moving all this information into the captions (if not already present) to improve the flow of information for the reader, and to make the text more concise.

Thank you for this valuable suggestion. We agree that the previous version devoted excessive space to restating numerical statistics and describing the visual features of the figures, which reduced the conciseness and flow of the Results section. Following the reviewer's suggestion, we carefully revised the presentation of the results throughout the manuscript. In particular, the detailed descriptions of the figures, including the passages around Lines 335–358 in the original manuscript, have been substantially condensed. Numerical information that can be directly obtained from the figures or tables has been reduced in the main text, while the key statistical values necessary for quantitative comparison have been retained. More importantly, the revised Results section now places greater emphasis on summarizing the major findings, comparing the performance of different methods, and interpreting the underlying physical or methodological mechanisms, rather than describing the figures point by point. We have also reviewed the other Results subsections following the same principle to ensure a more concise and consistent writing style throughout the manuscript.

Comment R2.6

Minor comments:

Please use Rayleigh winds instead of 'Ray winds'. While the world is full of abbreviations and acronyms, using the full word improves readability.

We agree that using the full term "Rayleigh winds" improves readability and avoids unnecessary abbreviations. We have therefore replaced "Ray" with "Rayleigh" throughout the manuscript and carefully checked the revised text to ensure consistent terminology. We also revised the Fig. 3 to ensure consistency (Lines 194-196):

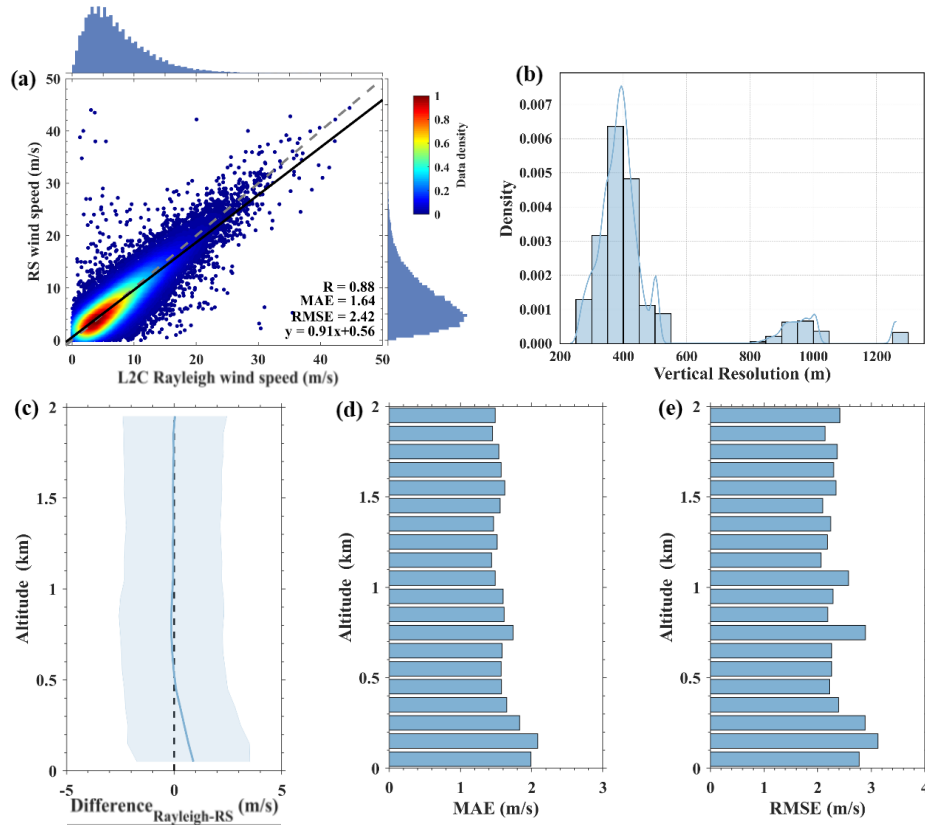


Figure 3. The (a) scatterplot and vertical distributions of (c) bias, (d) MAE, and (e) RMSE between Rayleigh wind and RS wind below 2 km. (b) Vertical resolution histogram below 2 km for Rayleigh product and the solid lines are the density curves.

Comment R2.7

Round statistics to significant digits: Given the uncertainty in both the spaceborn wind lidar and the RS observations and the inherent variability of the turbulent airflows, I am convinced that wind speeds and all their statistics (MAE, STD, etc) are not significant to within 0.01 m/s, which is 1 cm/s. Please assess both the accuracy and precision of the boundary-layer winds are report all statistics (including figures, text, tables) only for significant digits.

Thanks for your valuable comment. We agree that a numerical precision of 0.01 m s^{-1} should not be interpreted as an measurement accuracy of 1 cm s^{-1} . The uncertainties associated with the Aeolus observations, radiosonde measurements, and the spatial-temporal collocation are substantially larger than this level. Therefore, we carefully considered the appropriate reporting precision of the statistics presented in this study. The uncertainty of Aeolus wind observations depends on the observation mode, processing version, and atmospheric conditions. For example, validation against radiosondes and ECMWF model has reported a random error of about $4.1\text{--}4.4 \text{ m s}^{-1}$ for Aeolus Rayleigh winds in an early validation campaign (Martin et al., 2021). In our matched dataset, this value is about 1.8 m s^{-1} . For radiosonde observations, the nominal wind-speed measurement uncertainty of the Vaisala RS41 is approximately 0.15 m s^{-1} . Furthermore, the IGRA documentation explicitly notes that the reporting precision of radiosonde observations can differ from the actual measurement precision, with commonly reported wind-speed precisions of 0.1, 1.0, or 2.0 m s^{-1} depending on the data source and period.

These uncertainties clearly indicate that centimeter-per-second-level differences should not be interpreted as physically meaningful differences in wind-speed accuracy.

We nevertheless retain two decimal places for the main model-performance statistics in the manuscript. This choice is intended as a reporting convention for quantitative comparison, rather than an assertion of corresponding physical measurement precision. The performance statistics are calculated from the matched observations, and retaining two decimal places allows the relative performance of different products or methods to be distinguished more clearly. If these statistics were rounded too aggressively, for example to one decimal place or to integer values, differences between competing methods could become indistinguishable even when their calculated performance metrics are consistently different. This is particularly relevant for RMSE, MAE, bias, and STD, where relatively small but systematic differences between methods are part of the model comparison.

This reporting practice is also consistent with recent wind-profile and machine-learning studies, which commonly report model-performance metrics to two decimal places. For example, Hatfield et al. (2023) reported RMSE, R^2 , MAE, and bias values at the 0.01 m s^{-1} level in the vertical extrapolation of wind using machine learning. Similarly, recent Aeolus validation studies report wind biases, random errors, and error estimates with multiple decimal places to facilitate quantitative comparison, despite observational uncertainties being substantially larger than 0.01 m s^{-1} (Baars et al., 2023; Ratynski et al., 2023). Therefore, we distinguish between numerical reporting precision and physical significance. We retain two decimal places to preserve quantitative discrimination among the evaluated methods, but we do not interpret differences at the 0.01 m s^{-1} level as evidence of physically meaningful improvements in accuracy.

Reference:

- Baars, H., Walchester, J., Basharova, E., Gebauer, H., Radenz, M., Bühl, J., Barja, B., Wandinger, U., and Seifert, P.: Long-term validation of aeolus L2B wind products at punta arenas, chile, and leipzig, germany, *Atmospheric Measurement Techniques*, 16, 3809–3834, <https://doi.org/10.5194/amt-16-3809-2023>, 2023.
- 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.
- 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.
- Ratynski, M., Khaykin, S., Hauchecorne, A., Wing, R., Cammas, J.-P., Hello, Y., and Keckhut, P.: Validation of Aeolus wind profiles using ground-based lidar and radiosonde observations at Réunion island and the Observatoire de Haute-Provence, *Atmospheric Measurement Techniques*, 16, 997–1016, <https://doi.org/10.5194/amt-16-997-2023>, 2023.

Comment R2.8

Use of terminology of ‘boundary-layer winds’: While one expects the lowest part of the troposphere to be/ contain the ABL, the authors never actually evaluate the ABL height or

demonstrate typical ABL features such as nocturnal jets, or speed increase above, or vertical shear, or turbulence intensity/ TKE. Particularly above the oceans/ polar regions the ABL heights should substantially lower than 2000m asl, so I think a demonstration/ analysis should be added to support their claim of producing ABL winds.

Thank you for raising this important point. We agree that the term “boundary-layer winds” was too broad and could imply that the entire 100–2000 m height range is necessarily within the atmospheric boundary layer (ABL). This interpretation is not always appropriate because the ABL height (BLH) exhibits substantial spatial, diurnal, and seasonal variability. In particular, BLH can vary from only a few hundred meters under stable conditions, especially over oceans and polar regions, to more than 1–2 km under strongly convective daytime conditions over land. Moreover, even at the same location, BLH can change substantially between daytime and nighttime because of the development and collapse of the convective boundary layer, and can also vary considerably between seasons owing to changes in surface heating, atmospheric stability, and synoptic conditions. Consequently, a fixed altitude range such as 100–2000 m cannot be assumed to coincide entirely with the ABL at all locations and times. To address this concern, we have revised the terminology throughout the manuscript. Specifically, we have replaced “boundary layer winds” with “lower tropospheric winds” when referring to the reconstructed 100–2000 m wind profiles. We have carefully revised the Abstract, Introduction, Methods, Results, Discussion, Conclusions, figure captions, and other relevant sections to ensure consistent terminology and to avoid overinterpreting the physical extent of the ABL. The revised manuscript therefore presents the product as a reconstruction of lower tropospheric wind profiles, rather than as a profile in which the entire 100–2000 m range is assumed to represent boundary-layer winds. We appreciate the reviewer’s comment, which has helped us clarify the physical scope of the product and make the terminology more scientifically precise.