Answer to Referee #1

We greatly appreciate the insightful feedback provided by Referee #1, which we received on June 02, 2023. The comments from the referee that have been addressed in the manuscript are indicated in green, and the responses from the authors to the referee are highlighted in red.

The authors present statistics of the validation of Aeolus winds against independent ECMWF model fields and radiosondes. This is important work to gain knowledge on the errors of Aeolus winds. The region used for validation is limited to the tropics, which on the other hand is a very interesting region because of the challenging weather conditions with dust events and convective clouds and because of limited other Aeolus related Cal/Val campaigns in this region.

- Major comments

  ==============

  - G1) At many places in the paper, the authors compare MADI against EEtot. This is a fundamental mistake as MADI is not a metric related to standard deviation such as EEtot (and SMAD). The authors can confirm this by taking a sample of random numbers, with normal (Gaussian) distribution, and compare the MADI value with the input standard deviation value.
    - Thank you for your comment. We acknowledge the fundamental error in our manuscript. We have therefore thoroughly removed all comparisons between MADI and EE throughout the document.
  - G2) The authors should be more strong on their main conclusion in the abstract, by ending e.g. with: "Based on the data used in this study Aeolus Rayleigh winds do not meet the mission random error requirement and Mie winds do most likely not meet the mission bias requirement."
    - We acknowledge the need for a stronger conclusion to emphasize our findings. As you pointed out, the Mie bias doesn't meet the mission's random error requirement. However, when we account for the standard error of the bias, the Mie winds statistically align with the recommended value. Therefore, we have included the following sentence in the abstract of the manuscript: "*It is therefore concluded that Rayleigh-clear winds do not satisfy the random error requirement of the mission, whereas Mie-cloudy winds do so, when considering the standard error.*"
  - G3) The classes discussed in lines 171 to 174 are very unclear. For instance, what is meant with "below 3 km (very high, high, mid-level, low, very low and fractional cloud types)"? How can you have very high clouds below 3 km? Also, why not using the useful signal at measurement level to identify clouds within the profile?
    - We recognize that the previous description of our classification method has caused confusion. The primary objective of this classification is to categorize each observation based on the presence of clouds along the satellite track.

The confusion stemmed from our earlier approach of listing cloud types both within and above each altitude range. To address this, we have only included cloud types within the specific altitude range. Moreover, we have added the following sentence to succinctly explain the classification's purpose: "According to this classification, an observation bin is considered as cloudy if it is situated within or below a cloud." We trust that this clarification provides a more straightforward and comprehensible explanation. In response to your second question, while using signal intensity can help us detect clouds in the atmosphere, it may not be accurate enough on its own to tell the difference between clouds and other particles, like dust.

- G4) line 195. Did you check this statement, e.g., using spectra following Skamarock (2008). They show that the area below the kinetic energy spectrum (which is actually the atmospheric variability over the integrated scales) can be quite substantial when starting at 340 km (or truncation wavenumber 60).
    - Indeed, the kinetic energy spectrum, as presented by Skamarock, exhibits significant fluctuations at scales smaller than 340 kilometers. Nevertheless, in the free troposphere, it is well-established that African Easterly Waves (AEWs) with scales of 2000 to 4000 kilometers are a majorsource of variability over West Africa. Consequently, we consider it appropriate to adopt a more flexible criteria for the colocation radii. Our findings confirm this hypothesis, as we observe no error dependency with respect to the colocation radii.

- General comments

  ================

    - line 11; measurements -> observations (Note that for Aeolus an observation is the result of accumulated measurements; mixing these terms in the text is confusing. Please correct everywhere in the text accordingly)
        - revised accordingly
    - line 15; the orbital-dependent bias of up to 2.5 m/s applies to only some parts of the atmosphere. This nuance should be made here.
        - revised accordingly
    - line 33; "..... along the LOS of the instrument, which is directed perpendicular to the direction of satellite propagation. Please add the last part.
        - revised accordingly
    - line 54: Replace: "..... that still needs to be explored ..... potentially affecting ....." => .... that needs further exploration .... which impact ....
        - revised accordingly

- line 117: replace ".... some SRs ....." by ".... small SR values, which are dominated by instrument noise, ..."
    - revised accordingly
- line 120; I do not understand what you mean with ".... and distances between the instruments and the height bins"? Please explain or rephrase.
    - We replaced "the instruments" by "Aeolus" to make this sentence clearer.
- line 123: ".... especially in the case of strong Mie returns, which are not detected by the classification procedure, ..." The addition is important because in principle measurements with strong returns should be classified as "cloudy" and not enter the Rayleigh-clear wind.
    - revised accordingly
- line 138-139; vertical resolution is not in m/s. Probably you mean that the balloon ascending speed is 5 m/s, then measuring every 2 seconds gives a vertical resolution of 10 m. Please correct.
    - revised accordingly
- line 216; this a surrogate for the standard error, Right? Please use this more well-known terminology in statistics, rather than "uncertainty of the mean bias".
    - Yes, we made the necessary corrections to the text by using the term "standard error".
- line 227; I guess the representativeness error is different for Mie and Rayleigh winds as they sample the atmosphere along different length scales, i.e., about 10-15 km for Mie and and about 90 km for Rayleigh, along the satellite track? Can the authors please comment on this?
    - Since identical co-location radii are used for Rayleigh and Mie, the observations from both channels are averaged using the same length scale. Although this averaging involves more data points from Mie compared to Rayleigh due to the differing integration lengths, we consider it reasonable to apply the same representativeness error range for both channels. We have included the following sentence at line 231: "*Note that despite the integration lengths differing, we average Rayleigh and Mie observations over the same co-location area, leading to the application of a consistent representativeness error range for both channels.*"
- line 239; with EE you mean EE_Aeolus as in Eq. (9), right? Please be consistent in the text
    - We did indeed mean EE_Aeolus. We checked the entire text for consistency and changed it accordingly.
- line 243; what do the authors mean with: "noise related to atmospheric temperature and pressure"? Do errors in these parameters lead to wind random errors or biases?
    - Rayleigh-clear winds are measured based on the double-edge technique, where the Doppler shift of the broadband molecular scattered light is measured by means of two Fabry-Perot interferometers that are spectrally shifted by several GHz. The ratio of the intensity measured behind these

Fabry-Perot interferometers is proportional to the wind speed. To retrieve the actual wind speed, calibration procedures have to be performed as discussed by Dabas et al. 2008 (Correcting winds measured with a Rayleigh Doppler lidar from pressure and temperature effects, Tellus A, 60, 206–215, 2008). As the shape of the spectrum of molecular scattered light (Rayleigh-Brillouin spectrum) changes with temperature and pressure, the calibration of the Aeolus Rayleigh channel depends on the accurate knowledge of these two parameters. Hence, any uncertainties in temperature and pressure contribute to the random error of the retrieved wind speeds. However, this uncertainty can be considered to be significantly smaller compared to the contribution of the SNR. For further clarification of this topic, we adapted the respective sentence in the paper manuscript according to: "*Future baseline versions are foreseen to also include contributions to the EE caused by uncertainties of NWP temperature and pressure used in the processor for instrument calibration procedures as well as the one caused by an insufficient correction of the narrowband particulate return that is transmitted to the Rayleigh channel (Dabas 2008)*".

- Figure 1. red and orange are hard to discriminate. Please use a different color for orange (Rayleigh-cloudy).
    - Thanks for your suggestion. We have switched the orange color to green.
- Caption of figure 1, please mention explicitly that you used model equivalents from the model background (which did not (yet) use the radiosonde), see also line 129. This is important, obviously and good to mention again.
    - Thank you for your comment. We have included this information in the caption.
- line 275 mentions a STD of 2.1 m/s for sqrt(<HLOS_ECMWF-HLOS_RS>^2) at Rayleigh-clear locations. The same metric shows a value of 2.93 m/s at Mie-cloudy locations in line 279. That is quite a large difference for parameters with quite consistent and well-known error characteristics. Assuming that the quality of radiosonde observations is rather constant for the complete profile this suggests that ECMWF performs substantially worse at locations where Mie winds are found (lower troposphere) than at locations of Rayleigh-clear winds (upper troposphere, lower stratosphere). Or is this discrepancy simply a statistical effect due to the limited data set? Can the authors please comment?
    - Thank you for raising this point. In simple terms, the ECMWF model equivalents perform less accurately in cloudy areas compared to clear-sky conditions. Since Mie-cloudy observations are mostly present in cloudy regions, it's expected that the model equivalents for these situations would show lower performance compared to the model equivalents from Rayleigh-clear observations, which occur in clear sky conditions. However, this result may also be influenced by the small statistical sample size of Mie-cloudy. We mentioned this point in the text, with the following sentence: "*Please note*

that Mie-cloudy model equivalents, present in cloudy conditions, are anticipated to demonstrate lower performance compared to Rayleigh-clear model equivalents occurring in clear sky conditions."

- line 283; "as most of the systematic and random errors seem to be specific to the Aeolus Rayleigh-clear winds". But in the text above you show that Mie-cloudy biases are larger than for Rayleigh-clear. Please correct.
  - You are correct in pointing out that Mie has a more pronounced average bias, even though we have shown in section 4.1.3 that Rayleigh-clear exhibits an orbital-dependent bias. Therefore, we have removed the term "systematic" from the sentence in question.
- line 284: "This stresses the need to identify the underlying potential error sources of Rayleigh clear observations with respect to the presence of clouds and dust aerosols ......"
  - revised accordingly
- Given the larger systematic errors in Mie-cloudy I would think that these are more sensitive to clouds and aerosols. The fact that random errors are larger for Rayleigh than Mie is pretty clear. Please comment.
  - Please refer to our response to your earlier comment. We have addressed the issue by removing the systematic error in the sentence. You are correct in noting that when it comes to systematic errors, Mie-cloudy might be more influenced by clouds and aerosols.
- Table 2. sigma_mu is not defined in section 3.2. Please do.
  - We made an error by replacing "sigma_mu" with "epsilon_mu" without ensuring consistency throughout the text
- line 304; "For Mie-cloudy, the systematic difference indicates a bias of 0.9±0.3 ms1, which is within the uncertainty range of the ESA's specification ..."
  No, it is not, see major comment G2. Please correct.
  - Considering the standard error, the Mie error range does overlap with the recommended value of 0.7 m/s. This suggests that the Mie bias, accounting for the uncertainty represented by the standard error, is statistically consistent with the recommended range, when the standard error of the bias is taken into account. We have reformulated the text as follows, in line 309: "*For Mie-cloudy, the systematic difference reveals a bias of -0.9 ± 0.3 m/s, falling within ESA's specified uncertainty range when considering the standard error of the bias. This bias remains relatively consistent across regions and orbital nodes, with a slightly larger bias observed in the descending orbits and over Sal.*"
- line 306; how do you arrive at 1.1-2.3 m/s? Following Eq.8 with sigma_rep = 1.5-2.5, sigma_RS=0.7 and sigma_tot=2.9, I end up with sigma_Aeolus in the range 1.3-2.4. Where do I go wrong? See also table 3.
  - The results are not the same because you looked at information from Table 2, which includes data from all altitudes, while Table 3 only focuses on

heights between 2 and 16 kilometers for Mie-cloudy conditions. That's why the results are different. In the caption for Table 2, we have now specified "for all altitude ranges."

- line 315. I think AVATAR-T carries a 2 micron lidar, so measuring particles only. How can you compare these with Rayleigh-clear, measured in clean air conditions?
    - It is true that, besides the ALADIN airborne demonstrator (A2D), a 2-µm heterodyne detection wind lidar was flown onboard the DLR Falcon research aircraft during the AVATAR-T campaign. Due to the heterodyne measurement principle, the system indeed depends on the narrow-band particulate backscatter signal and will not provide winds from aerosol/particle-free atmospheric conditions. However, the detection scheme is much more sensitive compared to the direct-detection measurement principle used by ALADIN. Hence, the 2-µm wind lidar provides winds even at very low scattering ratios that are classified as Rayleigh-clear winds. This fact is for instance demonstrated by Fig. 3 in *Witschas et al., 2022*, which shows a flight example where the 2-µm wind lidar has almost full data coverage, and Aeolus measures almost only Rayleigh-clear winds. Hence, the 2-µm wind lidar is also well-suited to validate the quality of Rayleigh-clear winds. To clarify this fact, we added the following sentence to the paper: "*In this analysis, comparisons are only made with statistics derived from airborne wind lidar measurements acquired during the AVATAR-T campaign, which was also part of JATAC (Witschas et al., 2022; Lux et al., 2022b). In particular, the statistics derived from a heterodyne detection wind lidar (2-µm DWL) flown onboard the DLR Falcon research aircraft are used for comparison. Due to the high sensitivity of the heterodyne detection principle, the 2-µm DWL provides accurate wind speed data even in a clear atmosphere where Aeolus only provides Rayleigh-clear winds. Hence it is a well-suited reference instrument for the validation of both, Rayleigh-clear and Mie-cloudy winds.*"
- Figure 2. "Differences (dots) and average differences (lines)"
  I cannot conclude from the plot that the line is the average value. For instance in the left panel at 17500 m, all blue dots are on the right hand side of the line. Similar issues appear at all altitudes.
    - Thank you for your comment. We realized that we omitted mentioning in the caption that we smoothed the lines vertically using a 3-value moving average to reduce variability. We have made the necessary update to the caption, which now includes the following text: "*The lines were smoothed vertically using a three-value moving average.*"
- In the caption of Figure 2, mention Aeolus Rayleigh-clear winds.
- line 351; How is it possible to have '+'s with EE values > 5 m/s in figure 3a?

- In the caption, we mentioned that '+' symbols are defined for EE_Aeolus < 5 m/s. However, in Figure 3a, only EE_tot is shown, which explains values above 5 m/s.

- line 356; "discrepancy". This discrepancy is expected, see major comment G1.
  - Please refer to the response provided for the major comment.
- Figure 3. The binning of the stepwise solid lines is not explained. Why does it go up to 8 m/s in 3a, while you have much less than 40 data points at this value.
  - We acknowledge that the caption did not adequately describe the usage of the stepwise lines. Therefore, we have revised it as follows: "*The solid stepwise blue lines indicate the MADI, and the dotted blue lines represent the SMAD of Rayleigh-clear. Each step includes a minimum of 40 data points to ensure significance.*"
- line 407; "Table 4 describes the error dependency of the Rayleigh-clear observations with respect to the presence of clouds and dust"
  This classification is not clear. Do the authors mean presence inside the bin or from bins aloft or both?
  - We acknowledge that the previous classification caused confusion. This classification's purpose is to categorize each observation based on cloud cover along the satellite track, which is discussed mainly in section 4.2.2 "Cloud type and dust." In other words, when a bin falls within the 7 to 16 km range, all clouds within this range and those above 16 km are taken into account to calculate the percentage of cloud cover along the track. The confusion arose from listing cloud types both within and above each altitude range. To address this, we have changed the text to only specify the cloud types within each altitude range. The sentence "*According to this classification, an observation bin is considered as cloudy if it is situated within or below a cloud*" explains how this classification works.
- line 415-417. MADI compared against EE is invalid, see major comment G1. The conclusion that: "EEtot in clear sky conditions is well calibrated, while it is becoming gradually too low with the increasing presence of clouds and dust." is not well explained.
  - You are absolutely correct; we cannot directly compare MADI to EE. Consequently, we have eliminated all comparisons between MADI and EE in the manuscript. We have also revised the sentence as follows: "This underscores the trend where EE_{tot} is slightly overestimated in clear sky conditions and gradually becomes underestimated with the increasing presence of clouds and dust.
- Table 4. In the caption replace 25% by 50%
  - revised accordingly
- Figure 7a. use a different x-axis scaling to better visualize the differences between the curves, e.g., x in [-25,35]. Also for fig 8a and 9a.

        - We chose this scaling to maintain consistency across all three figures (7a, 8a, 9a) while ensuring ample space for a legible legend. It represents the most suitable compromise we could achieve.
    - Figure 7b, how do you arrive at the blue curve? And how the grey curves? Are the latter obtained from Eq.(9) with EE_Aeolus from the L2B product?
        - Thank you for your comment. Indeed, the blue and grey curves are derived from Eq. 9 using EE_Aeolus data from the L2B product. We have clarified this in the caption, specifying that they were obtained from Eq. 9.
    - line 483; "with a minimum of 3.5 m/s ...". This value does not follow from fig 7b. Please correct.
        - Thank you for your comment. The value 3.5 m/s mentioned in the text was initially associated with EE_Aeolus, but we have rectified it to 4.2 m/s to align with EE_tot.
    - Figure 8; the grey lines in b/c/d in look the same as in figure 7. Same for Figure 9. Despite the completely different scenes. Where do these curves come from?
        - It is correct and intended that the grey lines look the same in figure 7, 8, 9, as they represent the lines of all the 20 radiosonde profiles, used as a reference. To avoid confusion, we slightly reformulated this part of the caption: "*(b) Vertical profile of the Rayleigh-clear EE_{tot} (blue line), together with the EE_{tot} of all 20 profiles (grey solid lines) obtained from Eq. \ref{eqn:EEtot} and their average (black solid line).*"

- Minor comments

==============

- line 9; Raleigh -> Rayleigh
    - revised accordingly
- line 11; can be degraded -> are degraded
    - revised accordingly
- line 87; add "off-nadir" and "in the tropics" in "it points at 35 degrees off-nadir with an angle of ~10 degrees from the zonal direction in the tropics".
    - revised accordingly
- line 109; processing chain -> mission
    - revised accordingly
- line 112; L2bP 3.50 -> L2Bp version 3.50 (the rest of the paper uses L2B with all capitals)
    - revised accordingly
- line 113; "should is" - remove "should"
    - revised accordingly

- line 134; "Between the 7 and 28th ...". Correct to either: "Between 7 and 28 ...." or "Between the 7th and 28th of ....".
    - revised accordingly
- line 183; "in the presence of ...". Add "the"
    - revised accordingly
- line 200; remove "bin-to-bin"
    - revised accordingly
- line 209; Eq. (5) misses the index (i) between the brackets. Please correct.
    - revised accordingly
- line 232; "the the". Please correct
    - revised accordingly
- line 244; In contrary -> In contrast or Contrary? Please check.
    - revised accordingly
- line 473; black lines -> black line
    - revised accordingly