Reply to reviewers

Italics below indicate new text which is included in our revised manuscript.

We thank the reviewers for their careful attention to the manuscript and for the suggestions for improvements.

General comments :
Both reviewers ask for more comparison with other studies and more discussion of the contribution of radar measurement errors and other effects on the total random errors in the comparison between Aeolus winds. We have added more information and discussion as detailed below. One reviewer also asks for the evolution of the random errors over the 2-year period to be considered so we have added new figures dividing the data set into quarters and commented on this in the conclusions.

Specific comments by reviewer 1.

1. Further references to Aeolus validation studies have been added

Validation exercises soon after the start of the mission found that the quality of retrieved winds in part depended on the satellite's geolocation and on orbit orientation (see e.g. Guo et al, 2021; Lux et al., 2021). This could be traced back to unexpected instrumental effects, most prominently the influence of temperature on the performance of the primary telescope mirror of the instrument (Witschas et al., 2020; Lux et al., 2021; Weiler et al., 2021). The subsequent changes to the data-processing gave substantial improvement of the biases from more than 5m/s (Martin et al., 2021; Rennie and Isaksen, 2020) to less than 2 m/s (e.g. Iwai et al., 2021; Baars et al., 2022). However, Baars et al. (2022) noted that those improvements were partly masked by worsening instrument performance (e.g. decrease in laser output energy) that led to an increase of the random error. Nevertheless, Aeolus winds have been shown to make a positive contribution to global weather forecasting (Reitebuch et al,. 2020; Rennie et al., 2021; Weiler at al., 2021). A good number of validation comparisons of the corrected data processing after 2020 against a variety of other data sources have been reported, such as radiosondes (e.g. Martin et al. 2021; Rani et al., 2022; Chou et al., 2022), wind profiling radar (e.g. Guo et al, 2021; Kottayil et al., 2022; Chou et al., 2022), Doppler wind lidars (e.g. Chen et al., 2022, Witschas et al., 2022), numerical weather prediction models (e.g. Lux et al., 2022, Rani et al., 2022), and other satellites (Lukens et al., 2022). Overviews of recent validation comparisons were summarised by e.g. Wu et al (2022) and Ratynski et al. (2023), which mostly indicate possible biases less than 1 m/s and random errors 4-7 m/s for Rayleigh winds, 2-4 m/s for Mie winds. At the same time, the biases and random errors seem to vary more than might be expected between the different measurement techniques and locations used in the validations. Lux et al.(2022) have looked in detail at the non-random nature of differences between Aeolus winds and reference winds and suggest that the exact details of quality control applied in validation studies can significantly affect the results. They found that the bias and random error estimates can be affected by small numbers of outliers, particularly for Mie winds where large errors outside a Gaussian distri-

bution (gross errors) can be caused by misinterpretation of noise as signal. This can lead to predominantly positively-biased gross errors.

We have also updated the reference list accordingly

2. In line 53 we add *(Moveable Atmospheric Radar for Antarctica)* after MARA , *(ES-range atmospheric RADar)* after ESRAD. We change the reference in line 55 to Belova et al. (2021a,b) and add :

Each radar measures profiles of vertical and horizontal wind components in the vertical direction above the radar site. They switch automatically every 1-2 minutes between different modes with vertical resolution of 75 m, 150 m and 600 m (MARA)/ 900 m (ESRAD). The radars sample a cone of the atmosphere with a width of about 5° for ESRAD, 10° for MARA, so the horizontal diameter of the radar beams in the lowest 10 km of the atmosphere is less than 2 km at MARA, 1 km at ESRAD. Random errors (standard deviation of all 1 or 2-minute estimates in the 1-hour averages) are typically 2-3 ms^{-1} for both radars (Belova et al., 2021b). Comparison with radiosondes (Belova et al., 2021b). has shown no significant bias (<0.25 $m\,s^{-1}$) for winds at MARA but systematic biases at ESRAD of 8% for zonal winds, 25% for meridional winds (ESRAD underestimates wind components). These are thought to be due to the geometry of the radar antenna field and a high level of local radio noise. The ESRAD wind estimates are corrected for these biases before being compared with Aeolus winds. For the comparison with Aeolus (as in Belova et al., 2021a), we use 1-hour averaged winds, also averaged over the height intervals corresponding to the Aeolus Rayleigh wind averages. We use only radar measurements where the 95% confidence limit of the 1-h mean is less than 2 $m\,s^{-1}$ (this is calculated from the standard deviation and the number of the samples in the 1-h average, using students t-test).

For an extended discussion of the radar / spatial variability contribution to the random errors see point 1 in the reply to reviewer 2.

We add:

Belova et al, 2021b : Belova, E., Voelger, P., Kirkwood, S., Hagelin, S., Lindskog, M., Körnich, H., Chatterjee, S., and Satheesan, K.: Validation of wind measurements of two mesosphere–stratosphere–troposphere radars in northern Sweden and in Antarctica, Atmos. Meas. Tech., 14, 2813–2825, https://doi.org/10.5194/amt-14-2813-2021, 2021b.

and change Belova et al., 2021 to Belova et al., 2021a everywhere

3. See new paragraph added under 2.

4. We add the total number of data points to the title of each panel in the figures so they become :

Aeolus Rayleigh - MARA N=907
Aeolus Mie - MARA N=489
Aeolus Rayleigh - ESRAD N=2093
Aeolus Mie - ESRAD N=1052

and add to the Figure captions :

N in the panel titles is the number of samples corresponding to n=100%.

We also add the following to the description of Figure 1 in the text :

The fraction of total comparison points left after applying both the EE and ModZ$_i$ *rejection criteria (*n_z*) increases sharply for EE < 5 ms-1 and more slowly after that to just over 90% for Rayleigh (corresponding to ~1800 points) and to about 70% for Mie winds (~700 points).*

And to the description of Figure 2 in the text :

The fraction of total comparison points left after applying both the EE and ModZ$_i$ *rejection criteria (*n_z*) increases sharply for EE < 5 ms-1 and more slowly after that to just over 90% for Rayleigh (corresponding to ~1800 points) and to about 70% for Mie winds (~700 points).*

and add as a second last sentence in section 2:

This results in 80-90% of Rayleigh wind comparison points and about 60% of Mie wind points being available for analysis, sufficient numbers for further division according to summer/winter and ascending/descending orbits.

5. The text in lines 110 and 138 is correct. The caption for Table 1 (MARA) will be corrected to:

(summer, 23 September - 22 March and winter 23 March - 22 September). The caption for Table 3 (ESRAD) should read (winter, 23 September - 22 March and summer 23 March - 22 September)

6. We can only guess at possible explanations for why the distribution is skewed - it's only for ESRAD Mie not Rayleigh, and only for ESRAD not MARA. It might, for example be to do with local meteorology . We add the following to the end of section 5:

There is no obvious reason why the distribution is skewed, only for Mie winds, and only at ESRAD. One possibility might be local meteorology as the ESRAD area is often *covered by mountain-lee-wave clouds which might affect Mie (cloudy) measurements differently to Rayleigh (clear) ones.*

7. More discussion is now included in the conclusions. See our response under point 1 to reviewer 2.

8. We have divided the data into quarter-years and looked for changes over time (there are not enough points in the last quarter at MARA to include this). These are included as new Figures 9 an 10 (see below) and addressed in the final paragraph of the conclusions:

Figures 9 and 10 show how $bias_z$ and its confidence limits, SD_z and $ScMAD_z$ vary over the two years of the present study. These show an overall increase in confidence limits for $bias_z$ for all cases, and in SD_z and $ScMAD_z$ for Rayleigh winds. These are in line with the increase in estimated random errors for Aeolus winds between June 2019 and June 2021 (2B11 baseline) shown by Lux et al. (2022) , which is due to degradation in power of the Aeolus lidar. There is no clear increase in SD_z / ScMAD for the Mie wind comparison which could be due to the bigger influence of spatial variability on those values.

Technical corrections

We agree to all the technical corrections. For the missing units in Figures 3 and 5, we add this information to the caption to avoid cluttering the figures/tables.

Specific comments by reviewer 2.

1. We include more description of the radar characteristics and the contribution of various factors to the random errors. (see point 2 in the reply to reviewer 1)

To add to the discussion of the contribution of spatial differences, we have added a further column to each of Tables 1-4 which shows the results when we restrict the comparison to Aeolus measurements within 25 km of the radar sites, instead of 100 km. We add to the first paragraph of Section 3:

Tables 1 and 2 also include a further column which shows the results when the comparison is restricted to Aeolus measurements within 25 km of MARA (more precisely, those with the mid-point of the average along the orbit track within 25 km). Because of the geometry of the satellite orbit, these are all on descending passes. Note that since Rayleigh winds are averaged over about 87 km distance along the track, those measurements will still include observations up to 68 km from the radar along the track. For Mie winds, which are averaged over 15 km, observations up to 33 km away can contribute.

and to the first paragraph of Section 4 :

The final column in Tables 3 and 4 shows the results when the comparison is restricted to Aeolus measurements with their mid-points within 25 km of ESRAD. These are all on ascending passes and, due to averaging, will include observations up to 68 km /33 km from the radar along the track for Rayleigh / Mie winds respectively.

At the end of the conclusions we replace lines 187-199 with :

The problem of skewness for Mie winds has also been reported, and addressed in detail, by Lux et al. (2022).
The biases are similar in magnitude to results from other locations (e.g. Wu et al., (2022) Ratynski et al. (2023) and summaries included in those papers). Both Kottayil et al. (2022) and Ratynski et al. (2023) found no differences of the statistical results for ascending and descending passes. However, Martin et al. (2022) noted that biases can depend on latitude. It should be noted that Lukens et al. (2022) found large differences in the standard deviation of atmospheric motion vectors over Antarctica that were derived from Aeolus winds and from geostationary satellites, and indicated that this is due to problems with the correct height assignment. Chou et al. (2022) presented a validation comparison with both radiosondes and radar in northern Canada, i.e. from latitudes similar to ESRAD. They found that Aeolus winds correlate well with their radiosonde data. On the other hand, correlation with radar winds was much less good. The reasons were twofold, firstly the radar only operated for a limited time, leading to only a small number of profiles being available for comparison, secondly the range of the radar was limited, as it was not optimised to measure winds but rather hydrometeors (e.g. rain).

Random differences ($ScMAD_z$) for all data together are 5.9 / 5.3 ms^{-1} for Rayleigh winds at MARA/ESRAD respectively, and 4.9 / 3.9 ms^{-1} for Mie winds. We note that the random errors in the radar measurements should be < 2 ms^{-1} (we have used only radar wind estimates where the 95% confidence limit for the 1-h average is < 2 ms^{-1}) so that this should contribute little to the SD of the differences between radar and Aeolus measurements (less than 0.5 ms^{-1} assuming uncorrelated random errors). Since Aeolus HLOS winds have been sampled within a few 10s of km from the radar sites, we can use the comparison of radar measurements with radiosondes by Belova et al., 2021a to give some indication of the possible combined effects of spatial variability and random errors in the radar measurements (sondes, although launched at the radar sites, can be several 10s of km away by the time they leave the troposphere). Belova et al., 2021b found the standard deviation of differences between winds measured by the radars and sondes to be about 4 ms^{-1} at MARA (covering 291 sondes between February and October 2014), 5 ms^{-1} at ESRAD (28 radiosondes between January 2017 and August 2019). These are comparable the the values of $ScMAD_z$ found in the Aeolus-radar comparison for Mie winds, and slightly less than found for Rayleigh winds. So it is clear that part of $ScMAD_z$ is likely due to spatial variability but it is not possible to accurately quantify this. Assuming that the levels found in the radiosonde comparison are representative, spatial variability could in principle account for all of $ScMAD_z$ (e.g. for the ESRAD-Mie wind comparison), or as little as 25% (e.g. for the MARA-Rayleigh wind comparison).

An alternative is to consider the effect on $ScMAD_z$ of restricting the comparison to Aeolus wind measurements closer to the radars. These results are shown in the rightmost columns of Tables 1- 4, where only Aeolus measurements with mid-points within 25 km of the radars are included. For Rayleigh winds, there is no improvement in $ScMAD_z$ for the restricted data set, but the long along-track averaging distance of the Aeolus Rayleigh winds means that they still include contributions from up to 68 km away. For the Mie winds, with much shorter along-track averaging distance, there are improvements in $ScMAD_z$ with the restricted data set to 3.3 ms^{-1} at MARA and 3.6 ms^{-1} at ESRAD, well below the values for the other data subsets which are 4.2 - 6.4 ms^{-1} or MARA, 3.9 - 4.1 ms^{-1} for ESRAD. It seems likely that spatial variability is an important contributor to $ScMAD_z$, particularly at MARA. The geometry of the orbit passes at MARA means that 2 passes per week within 100 km are ascending, two descending. Only one (descending) pass per week comes within 25 km. The higher $ScMAD_z$ for ascending compared to descending passes at MARA could be explained by 50% of the descending passes being very close to the radar. Likewise, the much higher $ScMAD_z$ for winter compared to summer at MARA may simply reflect higher spatial variability of the winds in winter, particularly as the comparison is based primarily on measurements from the lower troposphere. At ESRAD, three Aeolus orbits per week pass within 100 km, two descending (one to the East and one to the West) and one ascending (only the latter within 25 km). The only difference between the 25km dataset and the full ascending dataset is the along-orbit distance included in the averaging for the comparison. The small improvement of $ScMAD_z$, from 4.0 to 3.6 ms^{-1}, with the restricted dataset suggests that spatial variability along the orbit path contributes a little at ESRAD. There is no difference no difference in $ScMAD_z$ between ascending and descending passes, suggesting along-orbit and East-West spatial variability are about the same. The slightly higher $ScMAD_z$ for winter (4.1 ms^{-1}) compared to summer (3.9 ms^{-1}) may again be due to slightly higher spatial variability in winter.

The higher values for $ScMAD_z$ for MARA compared to ESRAD (by 0.6 - 1.0 ms^{-1}) could be due to differences in local meteorology leading to differences in spatial variability. The higher $ScMAD_z$ for Rayleigh winds compared to Mie winds (by 1.0-1.4 ms^{-1}) is as could be expected because of different random errors in those wind estimates from Aeolus.

For the MARA and ESRAD data, Belova et al. (2021a) reported SD values for different subsets in the range 4-6 ms^{-1} for Rayleigh winds, and mostly 3-5 ms^{-1} for Mie winds. The present study shows SD_z 5.5 - 6.8 ms^{-1} for Rayleigh winds, 4.0 - 6.6 ms^{-1} for Mie winds, which are somewhat higher. Figures 9 and 10 show how $bias_z$ and its confidence limits, SD_z and $ScMAD_z$ vary over the two years of the present study. These show an overall increase in confidence limits for $bias_z$ for all cases, and in SD_z and $ScMAD_z$ for Rayleigh winds. These are in line with the increase in estimated random errors for Aeolus winds between June 2019 and June 2021 (2B11 baseline) shown by Lux et al. (2022) , which is due to degradation in power of the Aeolus lidar. There is no clear increase in SD_z / ScMAD for the Mie wind comparison which could be due to the bigger influence of spatial variability on those values.

As requested, we add a sentence to the first paragraph in section 3 to make clear that no correction is made for radar random errors - as we argue later these are small and the spatial variability contributes much more as discussed in the new paragraphs in Section 6.

(Note that no correction is made in the Tables for the random uncertainties in radar measurements).

2. We have added the number of available Aeolus measurements and the estimated radar error to the altitude profiles in Figures 4 and 6. We change the last paragraph of section 3 to :

Figure 4 shows height resolved parameters for the Aeolus-MARA comparison. Fig. 4a and 4d shows that low heights between 1-5 km dominate the comparison even though Aeolus wind estimates are available throughout the troposphere (and higher in the case of Rayleigh winds). This is due to the low sensitivity of the MARA radar in the upper troposhere and above. The uncertainty in radar winds is shown by the green line in Fig. 4b and 4c. Each radar wind is estimated from a 1-h average of measurements and the standard error of the mean (SEM) is used as an estimate of the uncertainty. Since we include only averaged radar winds where the 95% confidence interval is < 2ms^{-1} (this is twice the SEM when the number of data points in the average is large) , SEM is low, below 1 ms^{-1} and increases only slightly with height. (The SEM$_{MARA}$ profile is essentially the same for the ascending and descending passes as for all data so, for clarity, it is not included in the plot.) In Figs. 4e and 4f we can see that the negative bias for Rayleigh descending winds, seen in Tables 1 and 2, is seen at almost all heights, although the uncertainties in the bias become very large above 6 km height. It is partly balanced by a positive bias (marginally significant) for the ascending passes so that, for all data together (Figs. 4b and 4e), the mean bias becomes closer to zero. For the Mie winds, with notably more restricted height coverage, there is no significant bias at any height.

and the last paragraph of Section 4 to :

Figure 6 provides height-resolved profiles of parameters for the Aeolus-ESRAD comparison. As can be seen in Figs 5a and 5d, in contrast to MARA, the more powerful ESRAD radar provides useful coverage in the upper troposphere as well as the lower troposphere. There are fewer joint ESRAD-Aeolus observations than Aeolus alone although, between 2 and 5 km height almost all Aeolus measurements have corresponding radar ones, about half higher up in the troposphere. The green line in Figs 5b and 5e shows the mean SEM for the ESRAD wind averages, reaching 1 ms^{-1} at in the upper troposphere, lower at lower heights. (The SEM$_{ESRAD}$ profile is essentialy the same for the ascending and descending passes as for all data so, for clarity, it is not included in the plot.) Considering the bias profiles in Figs 4b,c,e,f , above 6 km height, the bias uncertainties are notably lower than at MARA - this is a result of a much larger number of comparison points thanks to the higher power of the ESRAD radar. For Rayleigh winds there is no significant bias at any height, for Mie winds the ~1 m/s positive bias identified in Table 4 is clearly seen at all heights. From Fig. 2 it is clear that a positive bias appears whatever the EE threshold.

3. Reference to Belova et al., 2021a added and HLOS capitalised everywhere

4. ‘for all available data’ added to the figure captions, total number given in panel title and EE limit used in further statistics marked.

5. Reference to Belova et al., 2021a added

6. We have made corresponding plots to Figs 1 and 2 for each data subset separately and these show essentially the same behaviour as Figs. 1 and 2, except for the levels of the bias and SD/ScMad values. These are too many to include in the paper but we add the following text in the last paragraph of Section 2.

We have made similar plots for all of the data subsets which we analyse below and found no reason to choose different thresholds for the different subsets. In all cases, ScMAD is close to $ScMAD_z$ *and their values are constant or changing very slowly between EE values 1 ms-1 above or below the thresholds. Similarly, bias and* $bias_z$ *are close together and insensitive to the EE values around the chosen thresholds, although both the* $bias_z$ *and* $ScMAD_z$ *values can lie at different levels in the different subsets, as shown in Tables 1 and 2 and discussed in the next section.*

7. We have added more discussion of the random errors in the 3rd paragraph of section 6 (see point 1 above)

8. We have added a sentence on seasonal effects, citing Weiler et al.

9. In line 114 we say “For the Rayleigh winds, Table 1 shows that there are no significant differences between summer and winter.” The bias 95% confidence limits for summer are [-0.6 0.6] and for winter [-1.6 0.1]. These have a large overlap which we interpret as no significant difference.

Technical corrections

1. Caption corrected and units added to caption

2. EE thresholds added to captions

3. We have added the text, units and a row showing the range of wind speeds to Tables 1-4.

4. The scale has been expanded

5. X-axis has been relabelled

Revised Tables 1-4

Revised Figures 1,2,4,6,7,8

New figures 9,10

Fig 9. Variation over time of $bias_z$, $ScMAD_z$ and SD_z for the Aeolus-MARA comparison. The shaded area around $bias_z$, indicates 95% confidence limits. Each quarter (all seasons, all orbits) over the 2-year study period was processed separately. Insufficient data is available for the last quarter.

Fig 10. Variation over time of $bias_z$, $ScMAD_z$ and SD_z for the Aeolus-ESRAD comparison. The shaded area around $bias_z$, indicates 95% confidence limits. Each quarter (all seasons, all orbits) over the 2-year study period was processed separately.