

Title: Validation of Aeolus winds using ground-based radars in Antarctica and in northern Sweden  
Author(s): Evgenia Belova et al.  
MS No.: amt-2021-54  
MS type: Research article  
Special Issue: Aeolus data and their application (AMT/ACP/WCD inter-journal SI)

We thank the referee for the comments that help us to correct and improve our paper. The referee comments are in **black**, our reply is in blue and changes in the manuscript are in **magenta**.

The main addition we will make to the paper following the reviewers comments (particularly reviewer 3) is to include two new figures summarising the mean differences between Aeolus and radar winds, also showing a comparison with the ERA5 model.

In preparing these figures we realised that one of the quality checks for the radar wind data (the requirement that 95% confidence limit for the time/height average should be  $< 2$  m/s) was not applied correctly. Correcting this leads to somewhat fewer comparison points (about 23% less for Rayleigh winds, about 13% for Mie winds) and to changes in the exact numbers for intercepts/biases/standard deviation etc in the Tables. Standard deviations are generally slightly less, biases changed by less than 1 m/s and the changes are within the confidence limits given in the original tables. Corresponding changes will be made in the text.

Anonymous Referee #3

#### General comments

In the manuscript the reprocessed data set from 1 July 2019 until 31 December 2019 is validated. First of all, this is an important information which should be mentioned in the abstract.

We will add the dates to the first sentence in the abstract to say:

Winds measured by lidar from the Aeolus satellite are compared with winds measured by two ground-based radars, MARA in Antarctica (70.77° S, 11.73° E) and ESRAD (67.88° N, 21.10° E) in Arctic Sweden, for the period 1 July - 31 December 2019.

The second point is, did you also analyze the operational data set from this time period in order to estimate the improvements of the new processor versions especially at the locations near the poles, where this information could be of interest for the processor developers?

We did not really look at the operational datasets (baselines 2B06/2B07) since our understanding was that fairly large biases were a general problem until the corrections for mirror-temperature effects were developed (starting with baseline 2B10). Following this query we have looked briefly at monthly average biases for 2B06/2B07 but we found they were highly variable and sometimes in excess of 10 m/s so we don't think a comparison is useful.

Aeolus wind observations are filtered for an error estimate threshold of 8 m/s for Rayleigh-clear and 4.5 m/s for Mie-cloudy winds. On which facts is this QC-criteria based? Did you try other values and how this affects the number of available data points and the determined random error of Aeolus wind observations?

The rejection thresholds of 4.5 m/s for Mie and 8 m/s Rayleigh errors were recommended by ESA/DISC (Aeolus Data Science and Innovation Cluster) (Reitebuch et al., 2019; Stoffelen et al., 2019; Rennie and Isaksen, 2020) for early period of Aeolus observations (laser FM-A) for the Aeolus CAL/VAL teams. For later period (laser FM-B) the limit for Mie was slightly changed for 5 m/s. These thresholds are chosen subjectively based on the compromise between the number of observations that pass quality control and the overall quality of the data set (Rennie and Isaksen, 2020). We followed the 1<sup>st</sup> recommendation and did not try other QCs. However, since we anyway needed to correct the application of our quality test for the radar winds, we have now redone all of the analysis with a rejection threshold of 5 m/s for Mie winds.

We will add to the text (P 6 l 150):

The rejection thresholds for Mie and Rayleigh winds were chosen as recommended by Rennie and Isaksen (2020). Those authors found appropriate thresholds subjectively based on a compromise between the number of observations that pass quality control and the overall quality of the data set.

Rennie, M. P. and Isaksen, L.: The NWP impact of Aeolus Level 2B winds at ECMWF, ECMWF technical memo, 864, doi: [10.21957/alift7mhr](https://doi.org/10.21957/alift7mhr), 2020.

As horizontal collocation criteria a radius of 100 km around the radar wind profiler sites was chosen. The location of the two RWP sites close to the poles provide the opportunity to cover many Aeolus orbits at this latitudes where the satellite tracks are closer together. Did you try to increase the radius to a larger value (120 or 130 km) to see if this could improve the statistics by including more data points although the representativeness error would increase?

Following this suggestion, we tried increasing to 130 km radius. This has a negligible effect on the number of available comparison points for ESRAD (<2% more points). The biases are unchanged (within the 90% confidence intervals) but the standard-deviations and confidence intervals increase slightly (by up to 20% for monthly averages). For MARA, there is a bigger increase in available comparison points (14% for Rayleigh winds, 40% for Mie winds). Again, the biases do not change (within 90% confidence limits). For Rayleigh winds, standard deviations and confidence intervals are essentially unchanged, for Mie winds they are reduced by up to 30%. Since finding signs of uncorrected bias seems to us to be the most important aspect of this comparison, and a distance of 100 km is already rather far given the substantial topographic variation around the radar sites, we do not feel it is worthwhile to change our analysis to include more distant observations.

How does the inclusion of all Rayleigh-clear observations within the horizontal collocation radius affect the validation instead of using only the closest observations? This would be consistent with the approach which is applied for Mie-cloudy wind measurements and would provide more data points.

Rayleigh wind profiles are already averaged along 87 km of track. If we were to include further profiles with their mid-point within 100 km from the radar, this would mean observations up to 143.5 km away were included. We want to include points as close as possible. For Rayleigh winds there are enough of those using only the closest profile each pass. For Mie winds, there are too few comparison points if we take only the closest profiles.

Have the authors included the random errors of the RWP measurements as well as an estimate of the representativeness error in the determination of the Aeolus wind observation error? Otherwise, the determined random error of Aeolus would be a combination of the different errors.

In the paper we estimated standard deviation of Aeolus – radar wind difference which is combination of Aeolus observation error, representativeness error and RWP random error. We will reword the text of the paper to make it more clear.

We will change in 5. Summary ... original p 16 1 329

*The random differences (4 - 7 m/s) are in most cases about what could be expected from the known level of random error for Aeolus winds (4-5 m/s for Rayleigh, 3 m/s for Mie) and spatial / temporal differences between radar and Aeolus measurements.*

**The random differences are a combination of Aeolus observation error, representativeness error and radar wind random error. The values we observe (4 - 7 m s<sup>-1</sup>) are in most cases about what could be expected from the known level of random error for Aeolus winds (4-5 m s<sup>-1</sup> for Rayleigh, 3 m s<sup>-1</sup> for Mie) and spatial / temporal differences between radar and Aeolus measurements.**

An ERA5 model comparison which was performed in addition was mentioned several times in the manuscript. Did the authors consider to include this analysis in this work also in view of the possible imperfect bias correction of the ESRAD system?

We will add 2 new figures (see at the end of the reply) showing monthly average bias estimates for Aeolus vs radar, Aeolus vs ERA5, radar vs ERA5 and include references to this in the text.

P.1 L.20: Please include the data set time period and mention that reprocessed data was used in the analysis.

Will be included in the abstract and made clearer at the end of section 2.1.

P.3 L.65: Please include a short overview of the manuscript structure

We appreciate the suggestion but do not consider such overviews helpful.

P.3 L.76: "20 laser pulses": Until January 2019, 19 laser pulses were accumulated for one measurement. Afterwards it was only 18 pulses. See Weiler et al. 2020 or Lux et al. 2021 ("ALADIN laser frequency stability...")

20 will be changed to 18-20

P.3 L.79: Mie winds were horizontally averaged up to 14 km as of March 2019. Also horizontal averaging lengths smaller than 14 km are possible.

We will add (P.3 L.79)

(Also horizontal averaging lengths smaller than 14 km are possible).

P.4 L.14: How was the ESRAD bias correction done? Is this correction stable over all 6 months and the covered heights? From Belova et al. 2020, it seems not to be a constant offset over all heights.

We add more details in the text about this correction (see reply to RC 1). The final version of Belova et al. (2021) also includes more details. Note that it is not a 'bias', i.e. an offset, it is a systematic fractional underestimate, so higher wind speeds give higher differences.

P.5 Fig.2: What is plotted here? Are these the locations of the single wind observations in the L2B product? How does the number of Aeolus wind observations which are used in the comparison fit to the numbers (N points) from Table 2 and 3? For me it looks like there were much more Mie winds used than Rayleigh winds which is not represented by the numbers in the tables. Please clarify.

We will extend the figure text to clarify this as follows:

There are up to 14 (very sparsely populated) Mie-cloudy profiles within 100 km radius of the radar along each orbit track. These are combined to make a single profile for comparison with the single radar profile, as detailed in the text.

P.6 L.37: Is the QC-criteria based on an error estimate threshold applied before choosing the closest Rayleigh-clear wind observation or afterwards? Please clarify.

We can add :

The closest Rayleigh-clear profile is chosen before quality criteria are applied.

P.6 L.36: To better understand the statement "24 or 27 km" the authors could mention that the Aeolus range bins are following a digital elevation model (DEM) and within a radius of 100 km around the MARA site a strong change in topography from 0 to around 3000 m is covered.

We will add this.

P.6 L.40: Between mid October and December there was a so called Strateole range bin setting (RBS) with a maximum height of 17 km (17 km or 19-20 km for the MARA site) and an AMV setting with maximum height of 14 km for 2 weeks in between.

We will add this information.

P.6 L.42: The height resolution is only possible as a multiple of 250m. Please check if 1130 m should be changed to 1000 or 1250 m.

It should be 1000 m. We will correct this.

P.6 L.43: The AMV RBS with 14 km maximum height was in operation for 2 weeks.

We will correct this and make clear it applies at both sites.

P.6 L.44: See comment P.3 L.79

We will change 14 km to '14 km or shorter '

P.6 L.44: The Mie wind RBS are not the same as the Rayleigh RBS. For the sake of completeness, the Mie wind RBS could be added.

P.6 L.49: Why did the authors create a single profile out of all Mie wind observations and did not compare single Mie wind observations with RWP wind measurements? Are the obtained Mie wind profiles free of gaps? Since the Mie wind RBS are not the same as the Rayleigh RBS settings, additional errors could be introduced when averaging Mie winds to the vertical Rayleigh wind resolution.

Our understanding is that Mie wind measurements are only possible from cloud layers and these will be at well defined heights within range bins. Examination of the Mie-cloudy height profiles also shows that each profile usually includes valid wind in only zero or one or two height bins. Since our radar measurements are necessarily an average over an extended height interval, it is better to find enough Mie wind measurements to also average over height. The only way to do this is to collect together several Mie-cloudy profiles along the track and average to suitable height bins. Since the Mie-cloudy height bins change along the orbit track we decided just to use the same height bins as the closest Rayleigh profile. Since we have only a single radar profile available for each pass, we would not be able to calculate confidence intervals if we compared with multiple Mie profiles (correlated errors as all would be compared with the same radar profile).

Subsection 3.1: Could the authors please comment if blacklisted data was excluded?

There were no blacklisted data that affected our study. We checked it at <https://docs.google.com/document/d/1xCOqI3jxhSe8T9IpA5jb1BaRzqhyriQU3UpjZvY2Gz0/edit?usp=sharing>

P.6 L.60: As already mentioned in comments above, the Mie and Rayleigh RBS are not the same. Are you also averaging to the Mie wind profiles?

See above.

P.8 L.16: To be consistent with the bias value in table 2 the value here should be changed to -1.9 m/s.

The text will be checked against the Tables and corrected.

Fig.4, 5, 10 and 11: Please increase the font size to the same value as used in Fig.2.

The figures will be replotted according to the comment.

Table 2 and 3: Can the authors please comment on the reason for the big difference in available data points for "Summer" and "Winter"? Are there less valid Aeolus or less valid RWP winds in winter?

Due to technical problems, the fca\_4500 mode at MARA was not available 24 June-2 September, which reduces the available comparison points in the upper troposphere for winter at MARA. There are also mostly fewer valid Aeolus wind measurements for winter compared to summer (10 % fewer Mie winds in winter at ESRAD, 50% fewer at MARA, 30 % fewer Rayleigh wind measurements in winter at MARA, but 30 % more at ERSAD. A detailed analysis of the differences is beyond the scope of this paper, but we will add a comment on the fca\_4500 problems to Table 1.

P.10 L.35: -2 m/s -> -1.9 m/s to be consistent with values in table 2

The text will be checked against the Tables and corrected.

P.13 Fig.10 b): Do you have any explanation for the strong negative bias for descending tracks at around 9 km altitude? Why is the std difference between ascending and descending orbits that large?

There is no obvious explanation for the bias at 9 km - this may just be an artefact of the small number on comparison points. Regarding the std difference - referring to Fig. 2b it is clear that the descending orbits are further away from the radar than the ascending ones which suggests spatial variability as an explanation of the larger std for descending orbits. We will add a comment on this to the text.

P.16 L.21: As mentioned above: up to 14 km

We will change (14 km) to (14 km or less)

New Figures to be included:

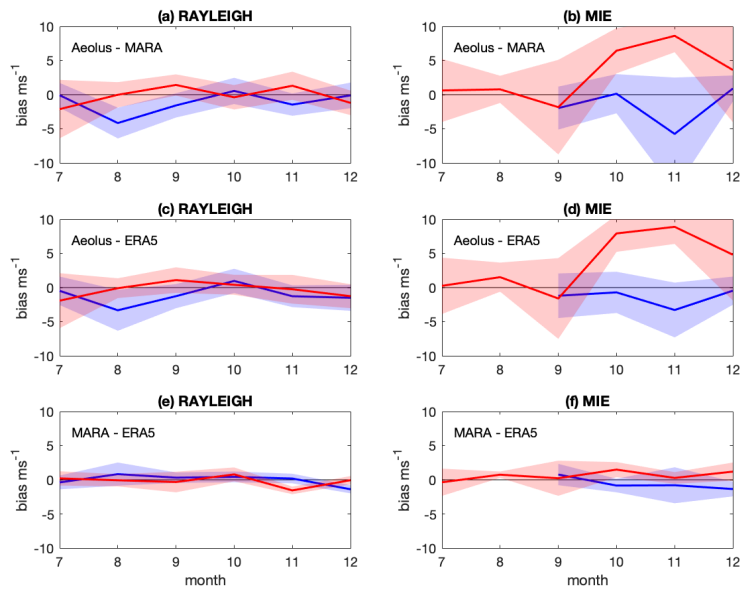


Figure X. Month-by-month mean values of biases in HLOS winds (solid lines) and 90% confidence limits (shaded areas) at MARA. Red for ascending tracks, blue for descending. a) Aeolus Rayleigh minus MARA, b) Aeolus Mie minus MARA, c) Aeolus Rayleigh minus ERA5, d) Aeolus Mie minus ERA5, e) MARA minus ERA5 at available Aeolus Rayleigh comparison times/heights, f) MARA minus ERA5 at available Aeolus Mie comparison times/heights.

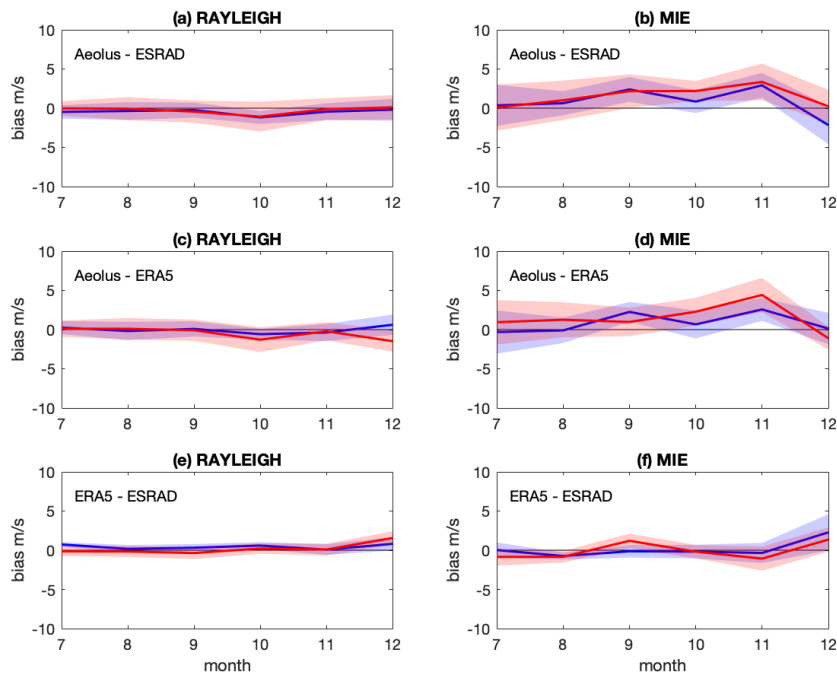


Figure Y: Month-by-month mean values of biases in HLOS winds (solid lines) and 90% confidence limits (shaded areas) at ESRAD. Red for ascending tracks, blue for descending. a) Aeolus Rayleigh minus ESRAD, b) Aeolus Mie minus ESRAD, c) Aeolus Rayleigh minus ERA5, d) Aeolus Mie minus ERA5, e) ESRAD minus ERA5 at available Aeolus Rayleigh comparison times/heights, f) ESRAD minus ERA5 at available Aeolus Mie comparison times/heights.