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 reviewer 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 #1
The main analysis recommendation is to see if more insight can be gained by digging further into the origin of the differences seen between ESRAD and Aeolus, in particular to what extent biases in ESRA might account for those differences (versus the lack of coincident data, particularly in the Mie channel, which is clearly described). Regarding ESRAD biases, taking a quick look at Belova et al. (2020) AMTD (https://amt.copernicus.org/preprints/amt-2020-405/amt-2020-405.pdf), the differences of ESRAD from radiosonde, HARMONIE-AROME, and ERA5 are complex, and I couldn't find a clear statement in it about recommended bias correction to ESRAD (apologies if I missed it). I was looking for this because the nature of and rationale for the bias correction made to ESRAD in the current paper (e.g. p.7, line 85-87) is not entirely clear. At p.4, ll.14-16, what does it mean to say that there is a 'systematic underestimate of wind speed by about 8% in zonal wind and 25% in meridional wind'. Wind speed is a scalar, so is the issue that ESRAD winds are weaker (lower speed) than the other products? From the other Belova et al. (2020) paper in AMTD, it seems like these statistics refer to separate linear regressions carried out for U and V. Is this suitable to apply when comparing ESRAD to an HLOS product like Aeolus? Again, why separately correct zonal and meridional winds rather than wind speed and direction? Have the authors done a separate analysis of the comparison of ESRAD to Aeolus without the Belova et al. 2020 bias corrections on ESRAD? If so, what does this reveal?

In the published paper by Belova et al. (2021) which is revised version of Belova et al. (2020) we added discussion and explanation why ESRAD underestimates winds and why the underestimates are different for the zonal and meridional components. (Note the underestimates are by a % of the magnitude of each component, they are not a 'bias' which would imply fixed offsets). In Belova et al., 2021 we answered most of your questions. However, we will add more explanation how and why ESRAD winds were corrected for the underestimates also in the present paper:
To be added at the end of section 3.3:
Radar winds are measured from time delays between signals received on different sections of the radar antenna array as the diffraction pattern of the scattered radio waves is advected by the wind. The baseline for determining the zonal component is longer than that for the meridional one and the receivers for the different parts of the array are not equally susceptible to non-random noise. This leads to underestimates of the wind speed which differ between the two components (Belova et al., 2021)

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, 2021.

Also regarding ESRAD, it is interesting that the fcx_aeolus was implemented as a special effort for the calibration/validation effort (p.7,l.78, and Table 1). But apart from the mention on p.7, a separate analysis of this data does not appear. Were systematic differences were found for this mode?

The mode fcx_aeolus was not analysed separately. It was simply designed to provide higher signal-to-noise ratio than fca_900 in the troposphere at the expense of height coverage of the mesosphere, so as to provide more wind estimates in parts of the atmosphere where the signal is weak (e.g. upper troposphere).

The main textual revisions I recommend are to clarify in the abstract and elsewhere that the analysis is based on a single six-month period of 1 July-31 December 2019, to clarify what the nature of the 'winter' and 'summer' seasons are here, and to discuss the implications of the use of a single season to characterize these errors. The reader might assume from the abstract that the analysis would take place over the entire Aeolus period instead of just when the homogenized and reprocessed 2B10 data was available. The authors could expand on their justification of only including this data in their analysis at p.3, line 84.

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.

We will add more justification at the end of section 2.1 (p3, line 84):
The move to baseline 2B10 and higher has been found to make considerable improvements to biases generally (Martin et al., 2020, Rennie and Isaksen, 2020) so it is most relevant to compare with these baselines. Because of long lead times for data transfer from Antarctica, at the time of writing, the most recent data available from MARA was 31 December 2019, so we focus on the time interval 1 July - 31 December 2019.

We will replace summer and winter by sunlit ('summer') and non-sunlit ('winter') when they are first mentioned in the abstract and the text (section 4)

Abstract :

Results for each radar site are presented separately for Rayleigh (clear) winds, Mie (cloudy) winds, sunlit ('summer') and non-sunlit ('winter'),

Section 4 (p.8, ll.04-05):

All data from 1 July to 31 December 2019 were divided into two seasons: sunlit ('summer' , 1 July-23 September at ESRAD, 24 September-31 December at MARA) with 12-24 hours direct sunlight and non-sunlit ('winter') covering the rest of the time.

In addition, while it is ok to characterize 1 July- 24 September (24 September - 31 December) as boreal (austral) 'summer', the complementary periods are shoulder seasons (boreal autumn/austral spring) and not 'winter'. This nomenclature is used when the authors interpret some of the wind biases in terms of winter-versus-summer seasonality (p.12, ll.63-65; p.16, ll13-14). This interpretation should mention that results could be influenced by a small number of weather systems that happened to occur at these sites during the six-month analysis period.
It is clear that the amount of sunlight distinguishes the two periods (p.8, ll.04-05) and the reason to separate the periods in this way is to focus on the role of insolation backscatter in controlling errors. So could the authors call the 'winter' period something like the 'nonsummer' or 'non-sunlit' period?

p.8, ll.04-05) see above
p.16, ll13-14) see below
p.12, ll.63-65; We will add a sentence on the limitations of the short time interval.

However, given the very short time interval which we have analysed, it is possible that this is not a summer/winter effect but just a result of a small number of individual weather systems.

Specific comments:
* p.2, l.43: Clarify what is meant by 'hot pixels'.

Hot pixels are increased dark current rates for specific ALADIN ACCD detector pixels, which can cause large biases in HLOS winds if not corrected for. We will add to the text:

"…corrections have had to be made for 'hot pixels 'which are increased dark current rates for specific ALADIN ACCD (accumulation charge coupled device) detector pixels…"

* p.2, l.55: Is this comment necessary for this paper? Perhaps it would be better placed in the discussion. The description of the 'limitation' of the Aeolus orbit design is distracting. The sun-synchronous orbit presents a challenge for calibration/validation but it is a reasonable strategy for capturing free tropospheric/lower stratospheric winds whose diurnal cycle is relatively weak, especially on the typical horizontal measurement scales of 10-100km achievable by this technology.

We will remove the sentence 'A particular limitation of the Aeolus dawn-dusk orbit is that measured winds are always made at the same local times. '

* p.3, l.78: to about *an* 87 km horizontal

Ok.

* p.3, l.79: for better impact on weather prediction -> to improve the impact of the retrieved Aeolus winds on numerical weather prediction

Ok.

* p.6, l.38: ">8 m s^-1": While it seems reasonable, how was this rejection threshold chosen, and what impact did it have on the results?

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 the early period of Aeolus observations (laser FM-A) for the Aeolus CAL/VAL teams. For the 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 1st recommendation in order to get more Mie wind data for comparison and did not try other QCs. In revising the paper we will change to 5 m/s for Mie winds since this is in better agreement with the more recent recommendation.
We will add to the text (P 6 l 150, section 3.1):

The rejection thresholds for Mie and Rayleigh winds were 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).
Rennie, M. P. and Isaksen, L.: The NWP impact of Aeolus Level 2B winds at ECMWF, ECMWF technical memo, 864, doi: 10.21957/alift7mhr, 2020.

* p.6, starting l.45: Replace "Mie winds" with "Mie wind measurements"

Ok.

* p.7, Table 1: Is there a typo in the end heights (104km, 100 km) - and if not is this consistent with the descriptions in Section 2?

This is correct for the end heights. These experimental modes are designed to observe the returns from mesospheric heights e.g. polar mesosphere summer echoes. In section 2 we described the lower atmosphere measurements which are relevant for Aeolus.

* Figures 3 and 6 show strong negative values for ascending HLOS in Aeolus but not in MARA; can the authors comment on these outliers? These extreme differences might be

worth pointing out. Is it possible that there are ranges of horizontal wind speeds that aren't captured by MARA's retrieval?

There is no reason why MARA cannot measure strong winds - the comparison with radiosondes in Belova et al (2021) shows good agreement out to at least 40 m/s, with discrepancies in both directions (MARA wind > sonde wind, and sonde wind > MARA wind). However, given that ground level varies from close to sea level at MARA to around 2000 m altitude just 100 km to the south, and strong localised katabatic winds occur in some conditions in Antarctica, strong local variations in low-level winds could be expected. There are indications of this in the very high standard deviation at the lowest heights (Fig 4b and 8b). But there are not really enough points to make a detailed study of this.

* p.12, l.63: The wording is confusing here, suggest "... do not vary between the seasons and weater systems are more variable in winter rather than in summer, which would lead to more spatial variability in winter than in summer". But again, as pointed out above, it is speculative to make this kind of generalization when only analyzing a single season, especially since there is a lot of synoptic variability in Antarctica year-round.

See above.

* p.16, l.10: "For winter the random differences are higher than in the comparison with radiosondes." Could this be quantified?

Since we corrected our quality check for ESRAD data, the random differences are no longer different between summer and winter. We will remove the text relating to summer winter differences.