Title: Validation of Aeolus Level 2B wind products using wind profilers, ground-based Doppler wind lidars, and radiosondes in Japan

Author(s): Hironori Iwai et al.

MS No.: amt-2021-243

MS type: Research article

Special Issue: Aeolus data and their application (AMT/ACP/WCD inter-journal SI)

Responses to referee #2

First, we would like to thank the referee #2 for your constructive comments that helped us improve our manuscript. In the present document, we provide our responses to the comments. The comments of the referee are reported in black font, our responses and the corresponding modifications in the manuscript in blue font, and the changes to the original manuscript in red font.

Review Summary

The paper manuscript "Validation of Aeolus Level 2B wind products using wind profilers, ground-based Doppler wind lidars, and radiosondes in Japan" H. Iwai et al. deals with the validation of Aeolus horizontal line-of-sight (HLOS) winds during different time periods (Oct to Dec 2018, Jun to Dec 2019 and Apr to Oct 2020) based on three different reference data sources coming from windprofilers, two coherent Doppler wind lidars as well as from radiosondes launched from Japan. Hence, these comparisons validate data obtained with the two different Aeolus laser transmitters FM-A and FM-B as well as data based on two different processor versions 2B02 and 2B10. The paper manuscript is clearly structured, all used methodologies are well explained and all figures are clearly visible. The content of the paper is well suited for AMT and especially for the Aeolus special issue. It is recommended to accept the paper manuscript after addressing the minor points that are mentioned below.

We are grateful for the referee's appropriate and positive comments on our manuscript. We revised the manuscript according to your comments and the details are shown as follows.

- General:

The entire analysis uses Aeolus data that is processed with two different processor versions 2B02 and 2B10. Could you somehow address the main differences that were implemented in the processor between version 2B02 and 2B10. This would help the reader to understand why one would expect really differences especially in the systematic error due to the implementation of the M1 mirror temperature correction or the hot pixel correction that was implemented.

Thanks for the suggestion. We modified and added the sentences in Sect. 2 as follows:

In this study, we used three different periods during the processor baseline 2B02 and 2B10 periods to assess L2B data products: 1 October 2018 to 15 May 2019 (2B02), 28 June to 31 December 2019 (2B10) and 20 April to 8 October 2020 (2B10). The first period with baseline 2B02 was within the commissioning phase, which was from launch to end of January 2019. The L2B data products with the 2B10 baseline include a bias correction for ALADIN's telescope primary (M1) mirror temperature variation (Rennie and Isaksen, 2020; Weiler et al. 2021b) and have been available for new observations since April 2020. A hot pixel correction has also been improved in the 2B10 baseline processor version. The L2B winds from 28 June to 31 December 2019 are a homogeneous reprocessed dataset using also the 2B10 processor version. We mainly discuss the measurement performance of Aeolus for Rayleigh-clear and Mie-cloudy winds during the baseline 2B02 and 2B10 periods. The baseline 2B10 period is composed of the M1 mirror and hot pixel bias corrected observations and the reprocessed data set.

We added the sentence in Sect. 5.1.1 as follows:

The reduced bias of the baseline 2B10 period compared to the baseline 2B02 is most likely due to the M1 mirror bias correction (Rennie and Isaksen, 2020; Weiler et al. 2021b) and the improvement of the hot pixel correction.

We added Weiler et al. (2021b) to the reference.

Weiler, F., Rennie, M., Kanitz, T., Isaksen, L., Checa, E., de Kloe, J., Okunde, N., and Reitebuch, O.: Correction of wind bias for the lidar on-board Aeolus using telescope temperatures, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2021-171, in review, 2021b.

You compare your determined results with different studies (GPS-RS results with Martin et al.; CDWL results with Witschas et al., WPR results with Guo et al.). But actually, you can use the data of all three instruments as reference. Thus, I am missing a detailed comparison of your own results from the different measurement data of the WPR, the CDWL and the GPS-RS. This should also be available in the summary, maybe with a table that summarizes to retrieved systematic errors based on the respective reference data sets.

Thanks for the suggestion. We added the sentence in first paragraph of Sect. 6 as follows:

Overall, the systematic errors of the comparisons with the three reference data sets showed consistent tendency. During the baseline 2B02, both Rayleigh-clear and Mie-cloudy winds exhibited positive systematic errors in the ranges of 0.5 to 1.7 m s^{-1} and 1.6 to 2.4 m s^{-1}, respectively. The statistical comparisons for the baseline 2B10 period showed smaller biases, –0.8 to 0.5 m s^{-1} for the Ryleigh-clear and –0.7 to 0.2 m s^{-1} for the Mie-cloudy winds. This suggests that the derived systematic errors are due to Aeolus Rayleigh-clear and Mie-cloudy wind systematic errors and not the reference data sets. The reduced bias of the 2B10 period compared to 2B02 is most likely due to the M1 mirror bias correction and the improvement of the hot pixel correction.

Sometimes there is a confusion with the "-" sign which sometimes denotes a minus and sometimes denotes a "to". Thus, it is suggested to replace the "-" sign with the word "to" and only use "-" as a minus sign.

Thanks for the suggestion. We replaced the "–" sign with the word "to" and only use "–" as a minus sign.

If possible, I would not spread units over two lines.

Since we don't know where the line break positions will be after the final proofreading, we think that this work is meaningless.

- **Fig.1, caption:** bule → blue

We corrected.

- **Line 92:** “two interferometers”…actually, as you also write later in the text, there are three interferometers: one Fizeau interferometer, and two Fabry-Perot interferometers all of them illuminated sequentially.

We changed “two interferometers” to “three interferometers”.

- **Line 98:** “Fizeau and Fabry Perot as narrowband filters” Somehow this is true, but the bandwidth is still different by a factor of 5 (Fizeau interferometer about 2 GHz, Fabry-Perot interferometers about 11 GHz).

We removed this sentence.

- **Line 99:** “After passing through some optics”. You could skip that part of the sentence as it provides no information.

Thanks for the suggestion. We modified the phrase as follows:

after passing through relay optics

- **Line 106:** “Commissioning phase” which was from x.x to x.x?.

We added the phrase as follows:

commissioning phase, which was from the launch of Aeolus to the end of January 2019

- **Line 110:** “Several different technical, instrumental, and retrieving checks account for this flag.” → Could you give a few examples which checks are performed and considered for deriving the validity flag?

We added the phrase as follows:

for example checking for signal and background radiation levels

- **Line 126:** “There is no significant difference between wind profiler winds and radiosonde winds in the biases and root mean square errors. → What do you mean with “significant”? Is there a difference? If yes, it would be better to quantify here. How do you determine the bias of WPR measurements or rather radiosondes?

We removed this sentence and added the sentences and a reference as follows:

The wind measurement accuracy of the WPRs was evaluated by comparisons with winds forecasted by the NWP model and radiosondes (Tada 2001). From the comparisons, the wind measurement accuracy of the WPRs was comparable to that from radiosonde observations.

Tada, H.: Use of wind profiler data in Japan. Technical Report on Numerical Weather Prediction (in Japanease), 34, 55–58, Numerical Prediction Division, Japan Meteorological Agency, 2001.

- **Line 126:** “294m” → 294 m (space)

We corrected.

- **Line 148:** “Doppler beam swinging (DBS) technique “. Is this a well-known technique? Could you add a reference here?

Thanks for the suggestion. We added Röttger and Larsen (1990) to the reference.

- Line 149: "The Doppler velocity spectra for all range bins of each beam were obtained 10,000 times on average. Since the PRF was 10 kHz, the accumulation time of each beam was 10 s". If I understand it correctly, the averaging time is one second, or you accumulate 100000 spectra. Or do I misunderstand something here?

You are right. We replaced '10,000' to '100,000'.

- Line 152: "Maximum likelihood estimator" Can you give a reference here?

Thanks for the suggestion. We added Levin (1965) to the reference.

Levin, M. J.: Power spectrum parameter estimation, IEEE Trans. Inform. Theory, 11, 100–107, https://doi.org/10.1109/TIT.1965.1053714, 1965.

- Line 153: "Bias was estimated at 0.02 m s^{-1} using measurements from a stationary hard target." → I guess, this value is only true for single LOS measurements, isn't it? If yes, this should be clarified here.

Thanks for the suggestion. We added the phrase "for single LOS measurements".

- Line 159: "As mentioned earlier, we averaged Doppler velocity spectra for all range bins of each beam from 30 min before to 30 min after the passage of Aeolus, and then the vertical profiles of horizontal wind speed and wind direction were acquired by the DBS technique." → Is it true that you averaged all spectra for 60 minutes and then you determine the mean wind speed? Using such an approach, I would expect the center peak in your power spectrum to be rather broad. Wouldn't it be better to calculate the wind speed on a e.g. one-minute average, and then calculate the mean over the 60 data points? Furthermore, I do not understand why you calculate and average for your CDWL comparison. Couldn't you actually just use the profile measured directly during the Aeolus overpass?

You are right. We averaged all spectra for 60 minutes and then estimated the LOS wind speeds of each beam.

We compared the Doppler velocity widths estimated from 60-min-averaged Doppler velocity spectra (σ_{60min}) with those estimated from 10-s-averaged Doppler velocity spectra during the Aeolus overpass (σ_{10s}). At Kobe, σ_{10s} = 0.51±0.18 m s^{-1} and σ_{60min} = 0.78±0.31 m s^{-1}. At Okinawa, σ_{10s} = 0.59±0.18 m s^{-1} and σ_{60min} = 0.79±0.19 m s^{-1}. We calculated the mean differences (BIAS) and the standard deviation (STD) of the difference between σ_{10s} and σ_{60min}. The BIAS is 0.26 (0.19) m s^{-1} and the STD is 0.27 (0.20) m s^{-1} at Kobe (Okinawa). Since the Doppler velocity resolution of the CDWLs is 0.75 m s^{-1}, the longer spectral averaging time (60 min) have a small impact on the Doppler velocity estimate.

The horizontal resolutions of the Aeolus wind observations are about 90 km for the Rayleigh channel and 90 (or 10) km for the Mie channel, so there is no sense in comparing instantaneous measurements obtained from the CDWLs during the Aeolus overpass. Although the longer spectral averaging time (60 min) yield broader power spectra, that is needed for the higher altitude where the signal-to-noise ratio (SNR) is much lower. Even if it is difficult to detect the spectral peak and estimate the Doppler velocity due to the low SNR for the accumulation time of 10 s, the probability of the peak detection and Doppler velocity estimation gets higher for the longer spectral averaging time.

- Line 229: "Note that this estimation of σ includes the representativeness error due to the spatial and temporal mismatch between Aeolus and reference instruments' measurements."

→ The representativeness error is likely to be different for the different reference instrument measurements. For instance, for the CDWL measurements you could decrease the representativeness error by decreasing the temporal averaging time to only a few minutes. Have you tried if this changes the random error for the comparison?
If the temporal averaging time decreases to only a few minutes, the numbers of data pairs for the comparison decrease and the derived random errors are not reliable.

- Line 248: "Furthermore, the range-gate settings of Aeolus were changed on 26 February 2019, which also increased the number of available data points during the baseline 2B10 period." → What was changed in particular? More range gates in the troposphere, less in the stratosphere? Or just increasing the range bin size? Here it would be helpful to get more details.
Thanks for the suggestion. We modified this sentence as follows:
The range-bin settings of Aeolus were changed on several occasions (Rennie and Isaksen, 2020). The number and resolution of the bins in the lower troposphere increased after 21 October 2019. Therefore, the number of available Rayleigh-clear and Mie-cloudy winds for the comparison increased during the baseline 2B10 period.

- Line 252: "slightly positive". As the bias is 1.6 to 1.8 m/s which is a factor of more than two larger the originally specified, I would skip the word "slightly" here.
We removed the word "slightly" as suggested.

- Line 261: "Comparison to A2D data". → You should be careful when comparing to A2D data as they compare LOS winds and not HLOS winds.
Thanks for the suggestion. We modified the sentences in the second paragraph of Sect. 5.1.1 as follows:
Lux et al. (2020b) compared the Rayleigh-clear winds measured along the Aeolus LOS with LOS winds measured with the ALADIN Airborne Demonstrator (A2D) during the WindVal III validation campaign carried out in central Europe from 17 November to 5 December 2018 (i.e., during the baseline 2B02 period). They reported a bias of 2.56 m s^{-1} with a scaled MAD of 3.57 m s^{-1}, corresponding to HLOS values of 4.25 and 5.93 m s^{-1}, respectively.
And we removed the sentence "They also reported that the slope of the linear regression line and the correlation coefficient of Rayleigh-clear versus A2D winds were 0.83 and 0.80, respectively."

- Line 289: It is worth mentioning here that only ascending orbits were underflown during the WindVal III campaign.
Thanks for the suggestion. We added the following sentence:
It is note that the WindVal III flights were conducted for probing the ascending orbit.

- Line 311: "The main reason for not yet achieving the mission requirement for random errors is probably related to the large representativeness error due to the large sampling volume of the WPR." → Also the Aeolus laser pulse energy is remarkably less than specified. With the representativeness error you would argue that the actual random error of Aeolus L2B data could meet the requirements. However, this is not true.
Thanks for the suggestion. We modified the sentence as follows:

The main reason for not yet achieving the mission requirement for random errors is the lower laser energy compared to the anticipated 80 mJ (Reitebuch et al. 2020a and 2020b). Additionally, the large representativeness error due to the large sampling volume of the WPR is probably related to the larger Aeolus random error.

We modified the related sentence in Abstract as follows:

The main reason for the large Aeolus random errors is the lower laser energy compared to the anticipated 80 mJ. Additionally, the large representativeness error of the WPR is probably related to the larger Aeolus random error.

We modified the related sentences in Sect. 6 as follows:

The main reason for the large Aeolus random errors is the lower laser energy compared to the target of 80 mJ. Additionally, the large representativeness error due to the large sampling volume of the WPR is probably related to the larger Aeolus random error.

- Line 320: "…there are relatively many paired data points for comparison (Fig. 6a)." → What does relatively many mean? Can you please quantify. It would also be very helpful to have this information plotted in Fig.6/Fig.7. This would give the possibility to understand how likely the shown bias trend is. For instance, is the negative Mie bias for descending orbits in 10 km altitude real, or just a result of very few data points and thus not reliable.

Thanks for the suggestion. Since it is difficult to plot the number of compared data points in Figs. 6 and 7, we plotted the vertical profile of the number of compared data points as shown in Figs. S1 and S2. We removed the phrase "where there are relatively many paired data points for comparison". We added the sentence as follows:

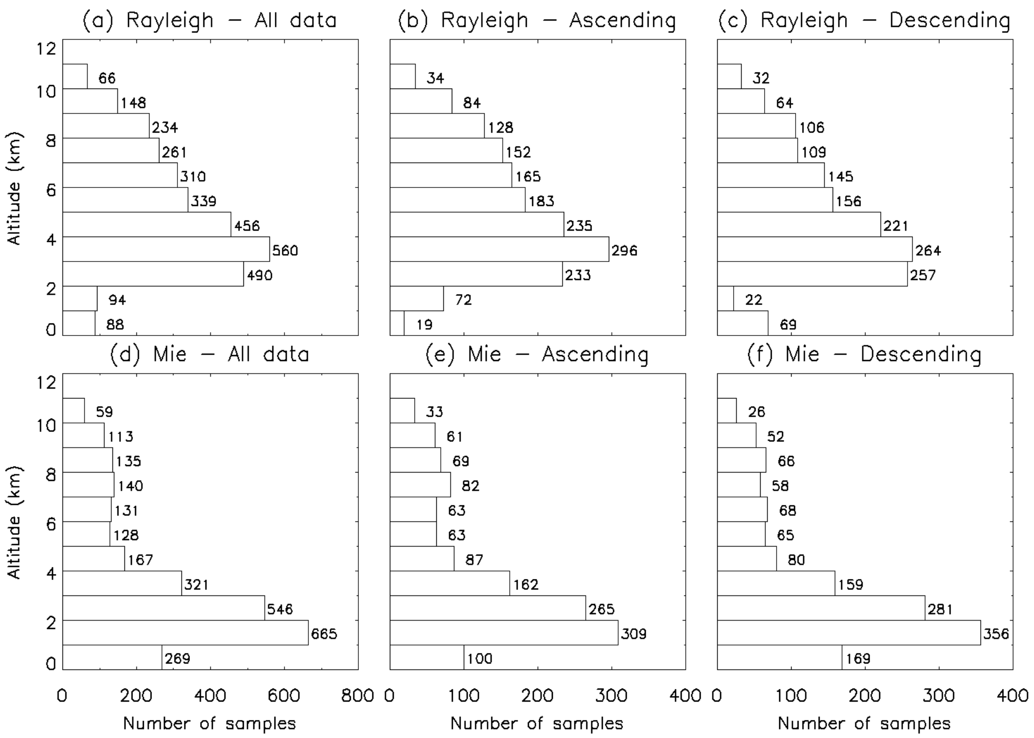

Figure S1. Vertical profiles in 1 km bins of the number of compared data points between the Aeolus and WPR HLOS winds for (a, b, c) Rayleigh-clear winds and (d, e, f) Mie-cloudy winds for (a, d) all data and (b, e) ascending and (c, f) descending orbits for baseline 2B02.

But there are very few paired data points in 10 km altitude (Figs. S1e and S1f) and thus the biases in 10 km altitude are not reliable.

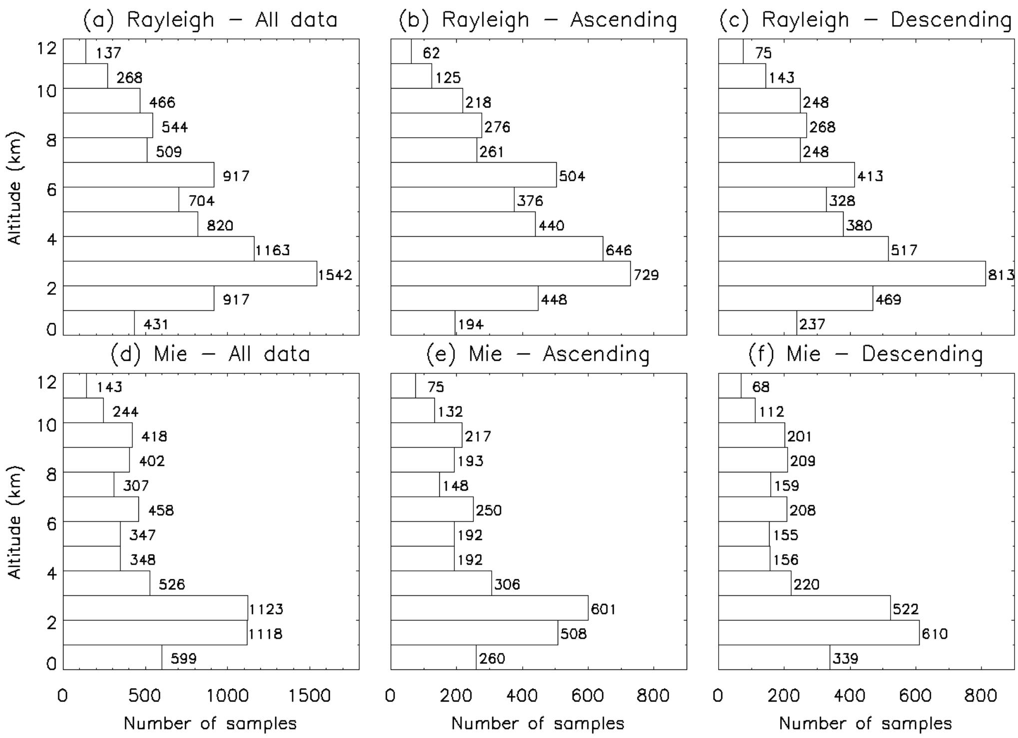

Figure S2. Same as Fig. S1 but for baseline 2B10.

- Line328: "The systematic error was less than that of the baseline 2B02." → Why? Here, you could refer to the differences in 2B02 and 2B10 which I suggested to discuss in the previous part of the manuscript.

Thanks for the suggestion. We modified the sentence as follows:

The systematic error is less than that of the baseline 2B02 due to the M1 mirror bias correction (Rennie and Isaksen, 2020; Weiler et al. 2021b) and the improvement of the hot pixel correction (see Sect. 5.1.1).

- Line 332: "However, this result is different from that in the other validation studies conducted during the baseline 2B10 period (Guo et al., 2021)." → What is different? What is the result by Guo et al.? Would be good to write one sentence here such that this information is available without reading Guo et al., 2021.

Thanks for the suggestion. We added the sentence as follows:

Guo et al. (2021) reported a large bias of 3.23 m s^{-1} with a standard deviation of 17 m s^{-1} for the Rayleigh-clear winds in the altitude range of 0 to 1 km.

- Line 339: "(descending) orbit, the minimum (maximum) bias is –1.93 (0.54) m s–1 in the altitude range of 5–6 (4–5) km." → this is confusing. Is the bias of -1.93 m/s corresponding to ascending or descending orbits?

We removed this sentence because this is not true.

- Fig. 8: “Monthly averages” → Why do you calculate monthly averages and do not show a time series on a daily basis?

Since there are only a few data pairs per day, the results from daily averages are statistically unreliable. Therefore, monthly averages are needed to obtain statistically meaningful analysis results.

- Line 358: “because the Mie return signal does not depend on the laser energy (Martin et al., 2021).” → I would not write it that strong. Indeed, compared to Rayleigh returns, Mie signals are much more depending on the atmospheric backscatter. But of course, if the laser pulse energy is too low, one would not be able to measure at all.

You are right. We corrected the sentence as follows:

There is no significant increase in the standard deviations of Mie-cloudy winds with time, because the Mie return signal does not only depend on the laser energy but also on the presence of aerosols or clouds (Martin et al., 2021).

- Line 369: “During the baseline 2B02 period, the bias of Rayleigh-clear and WPR HLOS winds slightly increased as the scattering ratio increased (Fig. 10a).” → Would this be better visible when calculating daily means instead of monthly means?

The results shown in Fig. 10 were calculated using all data during the baseline 2B02 and 2B10 periods. Since there are only a few data pairs per day, the results obtained from daily averages are statistically unreliable.

- Line 391: “result is similar to that in the comparisons of Aeolus and WPR measurements.” → which provides biases of x.x m/s. Would be good to repeat the numbers here.

Thanks for the suggestion. The following phrase was added:

which provides biases of 1.69 m s^{-1} (Rayleigh) and 2.42 m s^{-1} (Mie)

- Line 394: “The values are smaller than the scaled MADs of Rayleigh-clear (Mie-cloudy) versus WPR winds.” → Why do you think is this the case? The representativeness is similar, and the respective measurements errors of the WPR or rather the CDWL is corrected, isn't it?

Thanks for the suggestion. We added the sentence as follows:

The main reason for the difference is probably related to that the random error is larger for the WPR (3 m s^{-1}) than for the CDWL (2 m s^{-1}).

- Line 423: “The Rayleigh-clear winds show good coverage and closely follow the shape of the wind profile at altitudes higher than 2 km.” → Any ideas or hints what causes outliers as e.g. the one at 8 km for the Rayleigh-clear winds? Still aerosol contamination, as the range gates below provides a valid Mie wind? Would be interesting to see if there is an issue with the cross-talk correction of Mie signals in the Rayleigh channel...

Thanks for the suggestion. There are large deviations between Rayleigh-clear and GPS-RS winds at 3 km and 8 km. The scattering ratio is 1.15 and the relative humidity is about 30% at 3 km. Although there is a valid Mie wind at 3 km, the Mie wind is filtered out due to the HLOS error threshold of 5 m s^{-1}. This suggests that there are issues with the atmospheric classification and the cross talk correction of Mie signals in the Rayleigh channel. At 8 km there is no valid Mie wind. The scattering ratio on the Rayleigh channel is 1.13 and the relative humidity obtained from the GPS-RS is about 20% at 8 km. This suggests that there was a small influence of the cross talk of Mie signals to the Rayleigh channel on the large deviation at 8 km. We modified and added the sentences as follows:

The Rayleigh-clear winds show good coverage and closely follow the shape of the wind profile at altitudes higher than 2 km, but there are large deviations between Rayleigh-clear and GPS-RS winds at 3 km and 8 km. The scattering ratio on the Rayleigh channel is 1.15 and the relative humidity obtained from the GPS-RS is about 30% at 3 km. Although there is a valid Mie-cloudy wind at 3 km, it is filtered out due to the HLOS error threshold of 5 m s^{-1}. This suggests that the atmospheric classification in the Rayleigh channel was not working properly and the cross talk of Mie signals in the Rayleigh channel could have led to the large deviation. At 8 km, there is no valid Mie-cloudy wind. The scattering ratio and relative humidity are 1.13 and about 20%, respectively. This suggests that the cross talk has a small influence on the large deviation. The reason for that is unclear. Since the horizontal distance between the Rayleigh-clear measurements and the GPS-RS is about 80 km in this height region, large horizontal wind gradients in this height region potentially have an influence on the deviation.

- Line 435: "…the clouds were partly existent in the Aeolus observational domain" → This means that clouds are not sufficiently corrected or filtered out in the Rayleigh data product? E.g., the Rayleigh-clear wind in between 3-4 km where also clouds were partly present shows a quite large bias. Is there any way to analyze this altitude in more detail, e.g., by analyzing single Aeolus "measurements" instead of "observations", or is this information not contained in the L2B data product?

Thanks for the suggestion. Since the relative humidity obtained from the GPS-RS was about 90% at 3 to 4 km altitude, this supports that the clouds were existent in the altitude range. The Rayleigh-clear wind shows a large bias at 3 to 4 km altitude, but the scattering ratio on the Rayleigh channel was 1.15. This suggests that there was an issue with the cross talk correction of Mie signals in the Rayleigh channel. We modified and added the sentences as follow:

The occurrence of cloud was sporadically detected by the CDWL and the relative humidity obtained from the GPS-RS was about 90% at 3 to 4 km altitude (not shown). It is assumed that the clouds were partly existent in the Aeolus observational domain. The Rayleigh-clear wind shows a large bias at 3 to 4 km altitude, but the scattering ratio on the Rayleigh channel was 1.15. This suggests that there was an issue with the cross talk correction of Mie signals in the Rayleigh channel.

- Line 440: "There is a possibility regarding horizontal wind gradients in this height region." → You have actual wind speed and direction available from the radiosonde, right? If yes, it might be good to show if there was indeed a wind shear in this altitude.

Thanks for the suggestion. The vertical shear of the GPS-RS HLOS wind is about 10 m s^{-1} km^{-1} in the altitude range of 11 to 12 km (Fig. 13b). Since the range-bin resolution in this altitude range is 1 km, it is expected that the difference between the Rayleigh-clear HLOS wind and GPS-RS HLOS wind is negative. Therefore, the positive difference is potentially influenced by large horizontal wind gradients in this altitude range. We modified the sentence as follows:

Potentially, large horizontal wind gradients in this height region have an influence on the differences.

- Line 477: “Martin et al. (2021) estimated the radiosonde representativeness error, and error sources caused by spatial and temporal displacements need to be considered,” → How do they do that? Based on comparison to measurements or with respect to theoretical assumptions? This would be an interesting side note.

Thanks for the suggestion. We modified this sentence as follows:

Martin et al. (2021) estimated the radiosonde representativeness error σ_{r_GPS-RS} by considering spatial and temporal displacements, and the different measurement geometries of the radiosonde and the Aeolus observations.