Authors:

We thank the reviewers for their careful reading and their very constructive criticism. We are confident that due to their comments the manuscript could significantly improve.

We addressed all issues brought up by both reviewers in the manuscript. A detailed response to the comments of reviewer 2 is found below in *"italic"*.

Furthermore, we want to mention that during the revision of all statistical plots as requested by both reviewers, we realized that for the early baselines with the error unit in the Aeolus files being still m/s (before Baseline 8), Aeolus data with an error estimate higher than the stated thresholds were still included. We have removed those values now. The message remains the same, but some numbers have changed (improved). This has mainly affected Sec. 6.1.

Furthermore, we removed reference measurements (mainly radar) with an absolute error higher than 10 m/s these reference observations, so that a more meaningful comparison is possible. It has not changed much on the general statement, but some outliers (caused by the ground-based systems) are removed leading to a "better" systematic and random error statistic.

Response to reviewer 2

This manuscript presents the meaningful validation results for Aeolus via the long-term simultaneous measurements of ground-based instruments located at Leipzig (NH) and Punta Arenas (SH). It is impressive that the validation was performed for all product baselines (mainly from Baseline 05) after the switch-on of FM-B and the data products from different baselines are compared. Depending on mission time, baseline versions and the orbit types, the comparison results of this work are sufficiently presented and recognized appropriately.

The manuscript is well written and its contents are of high quality and scientific interest. The benefits of this study would be great for the accurate observations/high performances of Aeolus. Hence, I recommend the acceptance of this manuscript after the necessary revisions.

The specific comments are listed below:

1. Figure 1: This figure is not cited in this manuscript. This figure is somehow not necessary and the Aeolus Cal/Val stations are changed/updated worldwide. Thus, I recommend to remove this figure.

Thanks for the hint, indeed the reference got lost during the working process. However, we still want to leave it as it might be not evident for every reader where Leipzig and Punta Arenas is located.

2. The authors state that the Streamline Doppler lidars are applied in the supersites. Have the authors considered to include the simultaneous wind measurements from lidars into the comparison?

Good point. We operated a Streamline Doppler lidar in Punta Arenas the whole campaign period while in Leipzig one instrument was available only a short time. However, due to the very low amount of particles in Punta Arenas, the performance of the Streamline was not optimum for the Aeolus validation. Mostly, wind retrievals were restricted to the local boundary layer. But due to the relatively long distance to the Aeolus ground track (partly

more than 50 km) and the complex orography, it was not useful to use this data for the Aeolus comparison. We state this now more clearly in the text.

3. A 35 GHz Doppler cloud radar of type Metek MIRA35 at Punta Arenas, Chile was deployed and its data was used for the validation of Aeolus. In Section 2.2.1 of the manuscript (line 113 to line 115), you stated that "Once per hour, the stare mode (vertical profiling) was interrupted for a Range-Height-Indicator (RHI) and Plan Position Indicator (PPI, also called VAD - Variable Azimuth Display) scan. Only the PPI scans are considered for the horizontal wind retrieval.". Can I refer that the temporal resolution of the horizontal wind measured by the Doppler cloud radar is 1 hour? And how long does it take for one PPI scan, i.e., what is the accumulation time for each horizontal wind retrieval?

We have revised the methodology section, to answer this question, which was also raised by reviewer 1. PPI scans lasted 60 seconds and were performed always 35 minutes after the full hour. So, it is very close to the usual Aeolus overpasses and has a resolution of 1 hour even though the measurement itself is made within 1 minute.

Besides, the maximum unambiguous radial velocity and the resolution of the Doppler cloud radar were described in Section 2.2.1. Can you also provide the detection accuracy of this instrument? I think you can summarize the specifications and parameters of the instrument in table.

Indeed, we could do this, but reviewer 1 even asked for less details on the ground-based instrumentation. Thus, we have to find a compromise with the level of detail for the used cloud radar. Therefore, we think it is better to cite Görsdorf et al for the general Mira 35 concept and state only the most relevant parameter for the TROPOS radar.

1. In Section 3.1 of the manuscript, you described the horizontal wind retrieval method of the Doppler cloud radar and the three different fit methodologies are used to derive the horizontal wind vector in detail. But you just stated that "In the final data set, a best estimate is then computed which selects the method with the lowest error. This best estimate it then used for the comparison with the Aeolus winds." in line 176 to line 177. Can you provide the specific analysis result (e.g., errors of each method) and state which method was used? Is there one specific fit method for all the data set, or should the fit method be compared and chosen for every single wind profile retrieval?

Thank you very much for the hint. The method which was finally used for the Aeolus validation depended on the atmospheric conditions. In most cases, the standard approach in line with Päschke et al. (2015) has been used. Rarely, one of the other fitting methods was set as best estimate, e.g., in case when Doppler folding occurred or the error with method 1 was higher as for the other methods. We extended the description in this section, so that it should be now better understandable:

"All three methods are performed for each range R to calculate the horizontal wind vector. In a final step, a best estimate is computed, which selects the method with the lowest error out of the three methods. In the data set, the retrieval results from all three methods plus the best estimate is stored.

This best estimate is then used for the comparison with the Aeolus winds, however not considering cloud-radar-derived HLOS winds with an error higher than 10 m/s."

2. For the comparison criterion, why the temporal threshold of 1 hour is chosen? Theoretically, the ground-based radar performed continuous measurements of wind fields (except stare mode), hence there is no time difference between radar and Aeolus observations. For radiosonde-based comparison, assuming that we have a mean wind speed of 10 m/s for the horizontal wind and 100 km radius criterion, then it would take 100 km/10m/s=10000 s= ca. 2.8 h that the atmosphere moves above your ground site. Probably, the temporal threshold could be larger than 1 hour for radiosonde-referenced comparisons. In other words, the temporal threshold and spatial threshold criteria should be somehow consistent.

Thanks for bringing this up. Indeed, it was probably confusing formulated. For the radiosondes, which have been launched dedicated to the Aeolus launch, no temporal restrictions are made. For the Cloud radar observations, however, we feel that the temporal threshold of 1 hour is needed as in Punta Arenas atmospheric conditions (because winds are available in clouds only) do change much faster than, e.g., the general advection pattern over Leipzig which is assumed to be completely recorded with the radiosonde. We clarified this now in the text

3. Line 191 to line 194: the descriptions are "1–3 wind profiles fulfil this criterion of being within 100 km radius of the observational site (see green box in Fig. 3)" and "one can have up to 13–20 "Mie winds" for one altitude range around the 100 km of the ground-based location (see red box in Fig. 3).". But in the green box in Fig. 3, 7 wind profiles are presented, and I think there are more than 20 wind profiles in the red box. The wind profiles in the boxes are not all within 100 km radius of the observational site. You should mark the profiles fulfil the criterion further.

Very great suggestion! We modified Fig. 3 accordingly and show only Aeolus observations within the 100 km radius.

4. The legends in Fig. 9-Fig. 15 are unreadable. Alternatively, authors can enlarge them or list them in tables (like table 2 did) to help the readers to better understand the comparison results.

We modified the plots accordingly. And hope they are now much better readable.

The technical corrections:

1. The full stop at the end of the title should be omitted.

Done.

2. Line 1: …have been performed in Leipzig (51.12 N, 12.43 E), Germany, and at Punta Arenas (53.35 S, 70.88 W) should be have been performed in Leipzig (51.12 °N, 12.43 °E), Germany, and at Punta Arenas (53.35 °S, 70.88 °W). The degree signs are missing.

Added! Thanks!

3. Line 6: Does the “DACAPO-PESO” campaign have its full name? Please provide.

Provided now in Sec 2.1.1.

4. Line 181: According to your following description in line 182 to line 183, should it be two overpasses per week for each station?

Changed.

5. Line 192: “…vertical resolution” should be “… horizontal resolution…”

Well spotted, thanks!

6. CAL/VAL team (line 45) and CAL VAL teams (line 224) have to be consistent throughout the manuscript.

Done.

7. Figure 4: the Y-axis labels of the center and right panels should be consistent. Both should be “Height” or both be “Altitude”.

Thanks, the figure was completely revised and should now meet the requirements.

8. Figure 6: the legend is hardly readable. Please polish it.

Done, as for all other Figures which have been revised.

9. The units must be written exponentially. Hence the “m/s” should be changed to “m s^{-1}”. Please check this issue throughout the manuscript.

Changed!

10. Figure 16: the colors “pink” and “purple” sometimes are hardly to be distinguished. Please use some other colors.

Changed! We tried our best to find distinguishable colours and hope that it is now better readable.