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 reviewer comments is found below in *"italic"*. Due to the amount of comments, the response is partly short, which does not mean that we didn't value the opinion of the reviewer, but rather to keep the discussion well-arranged.

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.

Reviewer 1

General comments

The results are correctly presented, and relevant conclusions are raised about Aeolus performance and the processing algorithms. However, the large number of plots makes the manuscript not easy to follow for the reader. Additionally, some figures should be reformatted in order to make them clearer to read and interpret.

We reformatted all the figures, optimized the style and increased the font of the included text. We now hope that they are much better readable.

Specific comments

The abstract is well written and structured. The concepts are well handled and a good overview of what has been done is provided. However, more precise information about the period used for each validation should be provided.

Thanks, for the hint. We clarified more precisely now the analysed periods. To make the analysis more complete, we added all overpasses for baseline 14 in the Leipzig trend analysis shown in Fig. 17.

Additionally, some concepts are not properly introduced (e.g., baselines, Mie cloudy, Rayleigh clear). On the other hand, some minor rephrasing could be performed to improve the understanding of the text.

We tried to describe it better, however, having in mind that this is the abstract and not everything can be explained there.

The introduction section is equally well written. The manuscript is well referenced and this review is significantly valuable. However, again the period considered in the validation activities is not totally clear, as it is specified when they start but not when the end.

We have clarified this now in more detail.

Additionally, the satellite is introduced in this section, but no information is given about the scene classification, quality flags, Aeolus errors, for example. This information is partially lacking also later in the text. More information about the satellite should be included in the introduction or in Section 2, together with the detailed description of the locations and the instrumentation.

As the manuscript is submitted to the Aeolus special issue and many publications covering the validation of Aeolus are already within this special issue, we prefer to refer to other papers instead of describing everything again. But we have extended the description in Sec. 3 to give more details.

On the other hand, this description is very well presented and detailed. In fact, Section 3.1 is overly detailed and it is not clear why this much details are needed for this specific study when so few details were given about Aeolus measuring technique and data processing.

In our opinion, very detailed information on Aeolus is not necessary as shown in many other manuscripts in this special issue and other papers. On the other hand, the methodology for the retrieval of winds from Doppler cloud radar is rather new and the exact implementation (three methods) not published elsewhere. Thus, we have the feeling that some more details are needed. However, we tried to shorten this section a bit more and specify the reasoning in the introduction of this specific section.

The methodology used for the validation is robust and has been widely tested in previous studies. However, more information about the overpasses is lacking (e.g., mean distance, impact of the orbit shift), which can be easily introduced. The presented case studies help to clarify the procedure followed. Nevertheless, the editing of Section 4 should be improved.

Thanks for the hint, we have added the missing information in section 3 and also motivated the presentation of the case studies. Furthermore, we rephrased parts of Sec. 4 for a better understanding.

Regarding Sections 5 and 6, mainly the editing of the figures should be improved. Some plots are hard to interpret, and the information is difficult to read, given the amount of tiny text included. All of the plots are interesting for the manuscript discussion. However, the authors should think of a better way of including that large number of plots. Additionally, given the number of plots, it is difficult to link the text with the figures while following the discussion.

Thanks for the advice. We re-designed the figures the best way we could. We hope now that they are much better readable, even though we know that the amount of information in some figures is really high.

On the other hand, there is not a clear criterion about when to use "Mie", "Mie cloud", "Rayleigh" or "Rayleigh clear". For example, "Mie cloudy" is used in the figures, while "Rayleigh" (and not "Rayleigh clear") is used, and while "Mie and Rayleigh winds" are used in the discussion. Some standardization should be performed or at least it should be made

clear. Nevertheless, the discussion of Section 5 and 6 are valuable and interesting results were raised.

Thanks again for the statement. We standardized the wording throughout the complete text.

Technical corrections

- Line 5. Specify when does the validation activities end. For Leipzig the orbit mode of the overpass considered was specified. But not for Punta Arenas.

The radiosondes are still being launched, so no end date to be stated yet. But we stated the validation period now later in the text and added 4 more overpasses for the Leipzig comparison to complete baseline 14.

- Line 12. Specify what is meant with "the era of the first laser". Indicate this period.

Obsolete due to rephrasing.

- Line 16. The concepts of "clear" and "cloudy" should be previously introduced.

We think it is too extensive to explain these concepts in the abstract. But it is better explained now in the main text.

- Line 22. This sentence should be rephrased.

Done.

- Line 25. What is meant with "concept" in this line? Please, rephrase.

Done.

- Line 37. Indicate in some way that "the technology" refers to Aeolus. Although Flament et al. (2021) and Siomos et al. (2021) provide an interesting address of Aeolus optical products at an earlier stage (presented in conferences), there are other relevant studies which entail a more in-depth study and validation of Aeolus optical products (published in the same Special Issue as the present work). Please, consider including the works of Abril-Gago et al. (2022) (https://doi.org/10.5194/acp-22-1425-2022) and Gkikas et al. (2022) (https://doi.org/10.5194/amt-2022-205), although the latter is still in review.

Done!

- Line 68. Indicate when the radiosondes stopped being launched. Also indicate that for Punta Arenas, ascending and descending node orbits were considered.

The radiosondes are still being launched, so no end date to be stated yet. But we stated the validation period. Orbit notes are now also mentioned.

- Line 71. Indicate when did the campaign end and whether the occasional radiosondes were considered in the validation or not. These radiosondes were not mentioned in the abstract.

Done.

- Line 84. Figure 1 is not referred anywhere in the manuscript. Please, consider mention it and include some discussion or remove it.

Changed.

- Line 93. Include also the radiosonde models here.

Done!

- Line 101. What is meant with "primarily". Were the radiosondes used or not?

Rephrased. Thanks.

- Line 114. Then, the Doppler cloud radar provides one horizontal wind profile each hour (of the PPI considered)? Provide more details about this. How was ensured the temporal collocation of this measurements with Aeolus products? At which time were they performed (i.e., every hour on the hour o'clock)? Discuss about the temporal representativeness of the measurements.

The scans have been performed at minute 35 each hour and lasted for 60 seconds. This is now stated. The other issues are already discussed in Sec. 3.2. and Sec 4.1

- Line 115. Discuss the reliability of the measurements for winds larger than 10.56 m/s, specially given one of the case studies provided.

10.56 m/s are the radial velocity. The elevation angle was 85°. Thus, we can cover horizontal wind speeds up to 10.56 m/s / cos 85° = 121 m/s. It is expected that winds higher than this threshold did not appear in the troposphere.

- Line 120. Give the information somewhere in Section 2.2.2 about when the radiosondes where launched (i.e., how much time before or after the overpass). A deeper discussion about the spatiotemporal collocation of the radiosondes is lacking in the text. Was the horizontal drift considered? Discuss about the representativeness of these measurements.

We have stated now the launch time and added some discussion on the spatiotemporal colocation.

- Line 131. Rephrase the information about the Lockheed Martin LMS6 radiosondes.

Done

- Line 133. The radiosondes where launched, but where they used? After reading the manuscript it is not clear at all. In several places it is stated that Aeolus products were validated with radiosondes in Punta Arenas, but no statistical results are given. Line 133 is a good place to make this clear.

See comment above. We have not used them for statistics, but for case study validation also with respect to our cloud-radar-derived wind retrievals. We made this clearer in the text now.

- Line 135. Please, consider rephrasing Section 3.1 to be more concise. A detailed description is given about how the Doppler cloud radar obtains the horizontal winds, but few or no information is given about how Aeolus or the radiosonde obtain these winds.

See reply to the general comment above. We tried to more concise, but reviewer 2 asked for even more information. Therefore, the section was revised and we think we achieved a good compromise now. Furthermore, in section 3.2, a detailed description on the validation strategy is given.

- Line 136. Please, rephrase. Mention TROPOS here gives the wrong idea that the instrument was placed in Leipzig for the activities.

Thanks. We added "Punta Arenas" to make it clearer.

- Line 180, subscript. A 100 km criteria was mainly used, except for Punta Arenas during the period after the orbit shift. Please, discuss this fact and how this could affect the representativeness of the winds captured by each instrument. Also mention and discuss the 100 km criteria recommended by ESA for Aeolus wind product validation activities.

We added some further discussion with respect to the orbit shift and the therefore slightly increased radius. Furthermore, we added references for the choice of the 100 km radius threshold:

- *Geiß, A., et al.: Methodology and Case Studies for the Validation of Aeolus Observations by means of Radar Wind Profilers, 2019.*
- *Cossu, F., et al.: Characterization of Aeolus wind measurement errors, 2022.*

- Line 182, Figure 3. Specify that the zoomed in products are the ones within the 100 km criteria. Also specify which classification they belong: cloudy or clear.

We modified the plot so it should be clear now. Furthermore, we homogenised the wording so adding clear or cloudy is not necessary anymore.

- Lines 182 and 183. Indicate the mean distance between the overpass and the station and their orbit mode. This information will provide an interesting discussion about the representativeness of the comparison results. Also discuss how does this mean distance and the overpass times change (or not) with time, and specially with the orbit shift. Two overpasses are mentioned for Leipzig, but only one is used. Mention this in the manuscript.

See reply to general comment. We revised the respective section. Should be clearly stated now.

- Line 184. Provide more details about the scene classification.

See reply to general comment. After revising, it should be clear now. We prefer to refer to the already published description to not be too repetitive.

- Line 185. Discuss if the 87 and 15 km horizontal resolution has been constant during the satellite mission. This could be linked to the new baseline releases.

We refer now to Table 1 to make it clearer.

- Line 192. Give more details about the "before the resolution was 87 km". This comment is related to the previous one.

The date is stated there and together with the added information with respect to Table 1, we think it is sufficient information.

- Line 194. Provide more details about the temporal collocation of the profiles. Additionally, this sentence should be rephrased.

We have rephrased the text accordingly.

- Line 200. This sentence needs rephrasing. Specify that the HLOS wind was averaged in height. Additionally, more information should be provided about Aeolus bins and bins thickness.

Thanks for the advice. We have revised the sentences and added also references with respect to the Aeolus range bin settings:

- *Stoffelen et al, 2005*
- *Straume et al, 2019*
  -
  - Line 201, Table 1. Please, check if the start date for the operational processing of B12 is in May. Additionally, why were the subscripts not included in the "Additional info" column? If they are kept as subscripts, they should be numbered in some way.

Checked! First data of B12 appear on the online dissemination centre for 26 May 2021. I guess you mean with subscript the footnotes and indeed it is a good idea, we change that and added it to the additional information of Tab.1.

- Line 214. Type "Laser" with lowercase letters.

Done

- Line 219. Discuss here about ESA's recommended 100 km spatial criteria.

This issue was now already earlier in the manuscript. We refer to that now.

- Line 242. Please, rephrase.

Done

- Line 243. The profile was taken at 09:35, but it is not specified how much time it covers. Are these profiles always taken around half past? This temporal collocation of the Doppler cloud radar should be discussed before.

After revising the methodology section, your question should be answered now. The PPI scans last 60 seconds and are performed always 35 minutes after the full hour. So, it is very close to the usual Aeolus overpasses (and this was the reason why the scan was performed at this time).

- Line 244, Figure 4. The presentation of this figure should be improved. The arrow indicating the east direction of the middle plot is overlapped with a line and is barely visible. Additionally, the wind direction included in that plot is hard to interpret. In the left plot, the Doppler cloud radar can not be distinguished from the others. It should be specified which "Distance to lidar" the authors are referring to: distance to ground-track, distance to observations… In addition, it "lidar" should be replaced with "radar". It is not clear which time is given as "Ground-based instruments". The caption states "How are winds retrieves from the Doppler cloud radar", which is not really what Figure 4 presents. Furthermore, "Resulting winds retrieved with the Doppler cloud radar scans in regions cloudy occurrence" should be rephrased.

We have revised the whole plot, so it should now be of much better quality addressing all the mentioned issues.

- Line 248. Which frequency is that?

You are right, the statement is misleading. We mean high spatial (vertical) resolution. We rephrased that.

- Line 250. Please, rephrase.

Done

- Line 251. Some values should be provided in stead of only stating "excellent agreement". If GDAS data is going to be used, then some previous introduction should be given.

We rephrased the sentence to make it more clear. GDAS data is only used on case-study-basis for consistency checks, which should be now made evident from the text.

- Line 252. I recommend to change the tense of "have been" to "were".

Done

- Line 255. The discussion is very valuable. However, I would not make reference to the 87 km averaged Aeolus observation rather to the 100 km radius used, as that distance average is different for the Mie channel. Additionally, I would say that an optically thin cloud scenario could be also the case, so that Aeolus light could still penetrate.

The whole paragraph was changed and the additional scenario mentioned. Thus, we hope it is now easier to follow our argumentation.

- Line 262. It is not true for all the cases that the Rayleigh clear winds agree within the uncertainty range above 15 km.

That's true, we rephrased an also discussed the potential occurrence of smoke at this altitude.

- Line 264. Some observations are significantly lower than the GDAS, so the information is not totally true.

True, information added.

- Line 265. The following discussion gives the wrong idea that only cloud region could be analyzed. However, this is not true, as several radiosondes were launched and could be compared to both Aeolus Mie and Rayleigh winds. Then, it is worth mentioning here that the radiosondes could address the comparison in the other regions.

We tried to state it clearer, in fact no radiosondes were used by us for the long-term validation at Punta Arenas.

- Line 274. Information about the time when the overpass, Doppler cloud radar profile and radiosonde launch took place is lacking in Section 4.2.

Information is given now

- Line 277. Please, rephrase "considerable normal" or the whole sentence. Why using km/h now and not m/s as in the whole manuscript?

Rephrased and units unified.

- Line 280. Is there any evidence about the polar vortex shift over the region?

Unfortunately, the weather maps we used at this time to characterize the scenario are not anymore available online. Thus, we skip this explanation as it is not the focus of this paper.

- Line 285. I would recommend not to use "perfectly", especially because some points do not fit perfectly.

Changed

- Line 287. More information is lacking in the manuscript about how the radiosonde horizontal drift was solved, especially for the analysis in Leipzig. Please include this discussion in Section 3.2, for example. It could be interesting to include here a discussion with ESA's specifications of Aeolus performance under high wind conditions.

In Leipzig, the radiosonde has been launched directly for Aeolus overpass as now described in Sec. 2.2.2, so that we considered that the 1--2-hour ascent of the radiosonde is representative for the atmospheric conditions during the Aeolus overpass and no systematic biases is introduced in the long-term validation due to the radiosonde drift. This is now stated in the text. The same accounts for the dedicated radiosondes in Punta Arenas, which were however, more irregularly launched.
With respect to high wind speeds, the mission requirement document of Aeolus states:
MR-90: The wind observation profiles shall have a dynamic range of +/-150 m/s along the HLOS direction
and
MR-95: The wind observation profile performances shall be applicable over a dynamic range of +/-100 m/s along the HLOS direction.

So, considering this fact, Aeolus should have been designed in a way to be able to observe such high wind speeds. We have no further information other than that. But if the reviewer has more information and is willing to share that, we are happy to include them in the manuscript.

- Line 288. Give an approximate tropopause height here. Again, I would recommend not to use "perfectly".

Rephrased – thanks! Tropopause was at 9.5 km.

- Line 291. How can this be possible? Each Aeolus profile takes around 87 km, so how could two profiles fit in 100 km? It seems so simple that I might be wrong. I am especially concerned about the word "full" here.

The validation radius is 100 km radius. Rayleigh winds are retrieved with 87 km horizontal resolution and coordinates are given for the start, end and the center of this 87-km observation. We used the center coordinates to check if the Aeolus profile is within the 100 km radius. Following that approach, usually 2 Rayleigh wind observations per profile are used for validation. Sometime it can be even 3 or only one, depending on the distance of the ground track to the ground location. We clarified that in the text in this line but also in Sec 3.2.

- Line 298, Figure 5. The Doppler cloud radar profile is barely visible. Again, why using km/h and not m/s? Lidar should be replaced with radar as well.

We redesigned this Figure and hope it is better visible now.

- Line 298. I understand the limitation raised in this line. However, how can you be confident about it?

Correct. We changed "confident" to "consider".

- Line 314, Figure 6. A homogenization of the criteria used to name the channels is needed: use "Mie" or "Mie cloudy", but not both without any explanation before. Also rephrase (or not) "Rayleigh" accordingly.

    Done

     Additionally, including the error bars makes the plot harder to read. However, they could be kept, but further editing should be done to make the plots easier to interpret. Moreover, the units of the different coefficients and errors should be included. The same can be said to Figures 7, 9, 10, 11, 12, 13, 14 and 15. Especially Figures 9 to 15 need to be edited to make them easier for the reader to interpret.

We have redesigned all Figures – see general comment above! We hope the readability of the plots have now improved a lot.

- Line 313. Specify "Mie wind" and do not use "that wind". Additionally, discuss the comment about the slope and the mean bias.

*We added Mie wind now. Furthermore, we added: "as expected for a Gaussian distribution"
to make the discussion with respect to slope and bias clearer.*

- Line 356, Figure 7. Replace "Ground Doppler" with Radiosonde" in the figure
  caption.

Obsolete after Figure redesign.

- Line 356, Table 2. Why was omitted the uncertainty of the Punta Arenas' slope?

We simply forgot, thanks for spotting that!

- Line 356, Figure 8. It should be described what the different rectangles mean.
  Additionally, a reference here to each of the following sections would be appreciated.

Very good advice. We have done accordingly!

- Line 358. In line 352 it was stated that the comparison with radiosondes (for Punta
  Arenas) was not worth it, thus it was not included. Why mentioning radiosondes here?
  In the following sections no results of the radiosonde comparison are given.

Well spotted. Thanks! We removed it.

- Lines 360 to 364. Although the information is more or less clear, this sentence should
  be rephrased to ease the understanding.

Done!

- Line 373, Figure 9. Using the same axis limits for all of the plots would help to
  compare the plots and interpret the results. At least, the axis should be symmetric, in
  order to make the 1:1 slope easier to observe. On the other hand, the font size is really
  small and most of the text in the plots is difficult to read. The same can be said for
  Figures 10, 11, 12, 13, 14 and 15.

Redesigned – see general comment above. It really helped a lot. Thanks.

- Line 375. Please, make reference to the square drawn in Figure 8 which corresponds
  to this period.

Done

- Line 379. Why is it obvious? Please, discuss. Also discuss why it indicates an
  improvement in the quality flags and error calculations.

*Ok, we state the numbers now and with the new plots it should be more obvious. We also
discussed the quality flag issue now.*

- Line 393. Please, discuss why the Mie winds were already much more reliable for B05
  and B06.

According to Weiler, F et al, 2021, the impact of the telescope temperature variations were ten times less for the Mie winds than for the Rayleigh winds due to the instrumental nature of the two techniques. We discuss this now in the Manuscript.

- Line 399. Specify when does this decrease took place?

Done

- Line 408. Please, make reference to the square drawn in Figure 8 which corresponds to this period.

Done.

- Line 412. It is not clear why the Leipzig data was not used here.

Because we focus here on the ascending vs. descending orbit. As in Leipzig only the ascending orbit was available, we did not include it. We rephrased it and it should be clearer now.

- Line 425. Please, make reference to the square drawn in Figure 8 which corresponds to this period. Also, rephrase the sentence, please.

Done

- Line 436. Why was Leipzig data discarded?

Because we focus here on the ascending vs. descending orbit. As in Leipzig only the ascending orbit was available, we did not include it. We rephrased it and it should be clearer now.

- Line 447. Please, rephrase.

Done, thanks!

- Line 458. Given the intercept value and the biases obtained, no significant differences were really obtained? Please, discuss.

According to our understanding, the most significant change with Baseline 11 were

1. *Sat-LOS velocity correction implemented*
2. *New wind type classification (copied from ESA confluence): "different SNR thresholds for classification of Mie and Rayleigh" and "and an option to transfer Mie SNR results to the Rayleigh channel was added",*

Nevertheless, as discussed, we do not see a significant improvement but in fact a slight worsening. We stated this now more intensively and name some reasons, but we cannot make an in-depth discussion on the reasons due to the missing details.

- Line 477. Is there any specific reason why a 28-days moving average was used?

We intended to go for a 28-day smoothing window to ensure 4 weeks of overpasses (i.e. 8 overpasses in total and 4 for each orbit node). However, during the review phase we realized that the plot shown in the discussion paper was created with a 50-day smoothing averaged. This number was unfortunately hard-coded in the analysis code. We fixed this now, however the 28-day average which was then created was much noisier than the 50-day average (especially for Punta Arenas Rayleigh). Thus, we left this resolution window at 50 days which covers 7 weeks (14 overpasses) as a compromise. We state this now correctly in the text.

- Line 478. "The most recent available baseline …" can be understood, but rephrasing would be appreciated.

Done.

- Line 504. Please, rephrase.

Done.

- Line 526. It would be really interesting if you discuss further about the fact that the Doppler could radar only captures the wind within clouds. How does this affect the representativeness of the results? Are they representative to all vertical regions of the atmosphere? You will not find much cloudy regions in the stratosphere, for example.

Thanks. We discussed this now.

- Line 530. Specify that the occasional radiosondes were launched in Punta Arenas. However, if these radiosondes were not used in the comparison, I would omit the information from the Conclusions.

You are right, done!

- Line 554. Please, make reference to the availability of the Doppler cloud radar and radiosonde data from Punta Arenas.

We are still working on the upload of the radar data to EVDC which is more challenging than originally expected. If this does not work, we will publish the data at zenodo. If the reviewer likes, we can make the data available for him via an anonymous ftp account already now.

- Line 695. The DOI link should be corrected.

Done, thanks!