Response to Referee Comment (RC2) on

Quality control and error assessment of the Aeolus L2B wind results
from the Joint Aeolus Tropical Atlantic Campaign
(https://doi.org/10.5194/amt-2022-223)

We are grateful for the referee's valuable and helpful comments on our manuscript. The responses to the individual comments and questions are presented in the following together with the corresponding changes that will be made to the manuscript.

General comment #1:

The abstract appears to be a bit too verbose and I would suggest to make it more focused on the key results and their implications.

Response to General comment #1:

We agree that the abstract can be formulated more concise and have shortened it a bit accordingly:

> *Since the start of the European Space Agency's Aeolus mission in 2018, various studies were dedicated to the evaluation of its wind data quality, and particularly to the determination of the systematic and random errors of the Rayleigh-clear and Mie-cloudy wind results provided in the Aeolus Level-2B (L2B) product. The quality control (QC) schemes applied in the analyses mostly rely on the estimated error (EE), reported in the L2B data, using different and often subjectively chosen thresholds for rejecting data outliers, thus hampering the comparability of different validation studies. This work gives insight into the calculation of the EE for the two receiver channels and reveals its limitations as a measure of the actual wind error due to its spatial and temporal variability. It is demonstrated that a precise error assessment of the Aeolus winds necessitates a careful statistical analysis, including a rigorous screening for gross errors to be compliant with the error definitions formulated in the Aeolus mission requirements. To this end, the modified Z-score and normal quantile plots are shown to be useful statistical tools for effectively eliminating gross errors and for evaluating the normality of the wind error distribution in dependence on the applied QC scheme, respectively. The influence of different QC approaches and thresholds on key statistical parameters is discussed in the context of the Joint Aeolus Tropical Atlantic Campaign (JATAC), which was conducted in Cabo Verde in September 2021. Aeolus winds are compared against model background data from the European Centre for Medium-range Weather Forecasts (ECMWF) before assimilation of Aeolus winds and against wind data measured with the 2-µm heterodyne-detection Doppler wind lidar (DWL) onboard the Falcon aircraft. The two studies make evident that the error distribution of the Mie-cloudy winds is strongly skewed with a preponderance of positively biased wind results distorting the statistics if not filtered out properly. Effective outlier removal is accomplished by applying a two-step QC based on the EE and the modified*

Z-score, thereby ensuring an error distribution with a high degree of normality while retaining a large portion of wind results from the original dataset. After utilization of the described QC approach, the systematic errors of the L2B Rayleigh-clear and Mie-cloudy winds are determined to be below 0.3 $m \cdot s^{-1}$ with respect to both the ECMWF model background and the 2-µm DWL. Differences in the random errors relative to the two reference datasets (Mie vs. model: 5.3 $m \cdot s^{-1}$, Mie vs. DWL: 4.1 $m \cdot s^{-1}$; Rayleigh vs. model: 7.8 $m \cdot s^{-1}$; Rayleigh vs. DWL: 8.2 $m \cdot s^{-1}$) are elaborated in the text.

General comment #2:

The discussion of the potential implications of the results presented is mostly missing. It would be useful to develop this aspect in the abstract and in the concluding section.

Response to General comment #2:

Thank you for this comment which concurs with the remark from referee #3. To address this point, we have extended the introduction to better describe the objective of the paper and the methodical approach used to achieve this (new text is underlined):

[…] Therefore, a more detailed treatment of different QC schemes and how they affect the resulting statistics is necessary for comparable validation results and for a more objective assessment of the Aeolus wind data quality. Moreover, it is an important aspect with regards to the operational data assimilation in NWP centres and allows for a more rigorous error characterization of the Aeolus winds.

This paper aims to raise the awareness to the influence of the chosen QC schemes on the validation results, particularly when using the L2B EE. It also demonstrates the usefulness of specific statistical tools for the purpose of outlier removal and the assessment of normality, which are necessary to retrieve the Aeolus wind errors in accordance with the MRD. The presented methods are applied in the context of the AVATAR-T validation campaign in 2021 by comparisons against the ECMWF model background winds and the 2-µm DWL wind data. The specifics of the campaign and available datasets are outlined in Sect. 2 together with a description of the L2B Rayleigh-clear and Mie-cloudy EE and their temporal evolution over the past three years. The model comparison in Sect. 3 serves as an example to introduce the reader to the detailed treatment of the Aeolus wind data in terms of QC and error assessment. In particular, the modified Z-score (Sect. 3.2) and normal quantile plots (Sect. 3.4) are discussed as powerful tools for removing gross errors and assessing the normality of the wind error distribution, respectively. In addition, the impact of the QC settings on the results from the model comparison is elaborated (Sect. 3.5). In Sect. 4, the statistical methods are then applied to the comparison of Aeolus wind observations against 2-µm DWL data. The paper concludes with a summary and outlook to future studies of the L2B wind error characteristics in Sect. 5.

Moreover, we have split the last section into two sections named “Discussion and summary” (Sect. 5) and “Conclusion and outlook” (Sect. 6). The latter includes a new paragraph to highlight the relevance of the presented results not only in terms of the validation of Aeolus wind data, but also with regards to its assimilation in NWP centres, along with a short description of the QC scheme that is used in the Aeolus data assimilation at the ECMWF.

> *This work is intended to provide a guideline on how to perform a rigorous QC when working with Aeolus wind data. The presented results have demonstrated that a careful QC scheme is crucial for rejecting gross errors and, in turn, for providing an accurate estimation of the wind data quality. The shown statistical methods form the basis for a standardization and objectification of the Aeolus wind validation and will be applied in forthcoming studies involving DLR's wind lidar instruments. Furthermore, apart from the better comparability among different validation studies, the investigation fosters the analysis of the individual channel error characteristics and stimulates the refinement of the QC schemes that are currently used in the assimilation of Aeolus wind data into operational models. Both aspects are important to further improve the impact of the Aeolus products for NWP centres around the world.*
>
> *In this context, it should be noted that the operational assimilation of Aeolus wind data at the ECMWF involves a multi-step QC scheme which also largely relies on the imperfect L2B EE. It comprises a first-guess check, which rejects observations with very large (O-B) departures (5σ), followed by the so-called variational QC (VarQC) method (Andersson and Järvinen, 1998). The VarQC assumes that the distribution of the normalized wind error, i.e., the (O-B) wind error divided by the assigned observation error, takes the form of a Gaussian function including an offset. The assigned observation error is proportional to the EE and additionally considers a representativeness error of 2 $m \cdot s^{-1}$ for the Mie winds (Rennie et al., 2021). Finally, there is a blacklist in the ECMWF assimilation which removes Rayleigh winds below 850 hPa pressure altitude as well as Rayleigh-clear and Mie-cloudy winds with EE larger than 12 and ~5 $m \cdot s^{-1}$, respectively. The multi-step approach ensures effective removal of the largest gross errors, but the VarQC assumption does not well represent the Aeolus normalized wind error distribution, especially for the Mie winds. In this regard, the use of the modified Z-score may help to improve the performance of the QC in the Aeolus data assimilation.*

<u>Specific comment #1:</u>

The title of the last section should probably be Discussion and summary instead of Summary and conclusions.

<u>Response to Specific comment #1:</u>

As described in the Response to General comment #2, we have split the last section into two sections named “Discussion and summary” (sect. 5) and “Conclusion and outlook” (sect. 6). The former section wraps up the results from the manuscript in accordance with its title.

Specific comment #2:

L.573: should it be "Black bars"?

Response to Specific comment #2:

Yes, we have changed the text accordingly.