***Response to Referee Comment (RC3) 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 appreciate the referee's thorough reading and elaboration of our manuscript. The comments are very valuable and helpful for improving the quality of the paper. Our responses to the individual comments are presented below together with the corresponding changes that will be made to the manuscript.

General comment:

*I agree with the comment of RC-2 regarding the mostly missing discussion on the implications of the improved QC. While, in my mind, the paper presents results with high scientific value (understanding the instrument characteristics, improving the outcomes of validation and the outcomes of DA efforts), the importance of the improved QC are not strongly highlighted either in terms of motivation, or in terms of impact on the scientific or applications efforts (e.g. operational DA).*

Response to General comment:

Thank you for this advice which concurs with the comment from referee #2. We have revised the last part of the introduction to elaborate on the motivation of our study and the used methodology (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*

*<u>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).</u> 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.*

Regarding the implications for the Aeolus wind data assimilation, it should be pointed out that the DA at the ECMWF involves a complicated QC which relies on a multi-step procedure to ensure effective outlier removal. To start with, there is a first-guess check of the (O-B) wind error with respect to the model background, rejecting Aeolus winds with departures greater than $5\sqrt{\sigma_O^2 + \sigma_B^2}$, where $\sigma_O$ is the assigned observation error (1.4 times the EE for the Rayleigh; 1.25 times the EE plus 2 $m \cdot s^{-1}$ representativeness error for the Mie). $\sigma_B$ denotes the background error which is derived from Ensemble of Data Assimilation (EDA) statistics and can vary from about $\sim\sigma_O$ for Mie-cloudy to $\ll\sigma_O$ for Rayleigh-clear winds, so that the first-guess check typically discards wind data with deviations larger than ~5 $\sigma_O$ to $5\sqrt{2}$ $\sigma_O$. The first QC step is followed by the so-called variational QC (VarQC) method (Andersson and Järvinen, 1998) which assumes that the distribution of the normalised wind error, i.e., the (O-B) error divided by $\sigma_O$, takes the form of a Gaussian plus flat distribution (Gaussian function including an offset) to account for gross errors. The VarQC applies a weighting to the Aeolus observations depending on the normalized wind error. However, since for the current settings, observations are only down-weighted significantly for departures larger than 4.71 $\sigma_O$, in most cases the VarQC has only a small filtering effect in addition to the first-guess check. Finally, there is a blacklist in the ECMWF assimilation which removes Rayleigh-clear 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. Hence, this multi-step QC scheme used in the ECMWF DA has some similarities to the two-step QC approach described in the paper, as it combines a rather relaxed EE threshold with a filter that rejects winds with large departures from the model wind. However, as this multi-step QC scheme also largely relies on the imperfect EE and does not directly aim at a Gaussian wind error distribution, there is probably some room for improvement with regard to the DA in NWP.

Due to the complexity of the QC scheme used at the ECMWF, we refrained from elaborating on it in detail in the text. However, we have split the last section into two new sections "Discussion and summary" and "Conclusions and outlook", and have added a short paragraph to the latter including a brief description of the ECMWF QC 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:

*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>

*When using the ECMWF winds as the truth against which the Aeolus winds are compared, up to a 12-hour ECMWF forecast is used so that the model data are independent from the Aeolus winds (they have not been assimilated yet). However, Aeolus data have been assimilated in the previous model runs (cycles). What is the possible impact of the fact that the previous cycles have already assimilated the Aeolus winds?*

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

That's a very good question which also concerns the assessment of the impact of Aeolus wind retrievals, e.g. on ECMWF global weather forecasts, as discussed by Rennie et al. (2021). In this work, it is stated that since the start of the operational assimilation of Aeolus winds at ECMWF

> *"[…] the background departures are no longer independent of past Aeolus winds; it is unclear if this affected the Aeolus error estimates but there is no obvious discontinuity in the time series, so it probably did not."* (Rennie et al, 2021).

Given this statement, we assume that the influence of the assimilation of Aeolus winds in previous model runs on the (O-B) statistics is only minor. A more detailed investigation would be required to verify this assumption, but goes beyond the scope of this paper. Aside from this, the fact that even model background is not entirely independent of the Aeolus wind observations emphasizes the relevance of the performed Cal/Val activities using ground-based and airborne instruments such as the 2-µm DWL.

<u>Specific comment #2:</u>

*The text on P. 23 that describes Fig. 10 has several inconsistencies with the figure - e.g. Figure 10 does not have orange bars; it is said that the 1-step QC statistics are given in a black inset while it is gray.*

<u>Response to Specific comment #2:</u>

We have corrected the text as follows:

> *Finally, the PDFs of the Mie and Rayleigh wind errors are presented in Fig. 10, indicating those wind results that are filtered out by the EE threshold (red bars) as well as those that are additionally filtered out by the modified Z-score (<u>black</u> bars). The statistical results that are provided in the boxes refer to the different subsets without QC (red), one-step QC using solely the EE threshold (<u>grey</u>) and two-step QC additionally applying the modified Z-score filter (blue/green).*