Response to Referee Comment #2 on

Correction of wind bias for the lidar on-board Aeolus using telescope temperatures

The authors thank reviewer #2 for carefully reading the paper and providing valuable input. In the following, referee comments are repeated in green and answers by the authors are provided directly below in black.

General comments:

This is an excellent paper that describes empirical correction of Aeolus wind bias based on temperature gradients across the primary mirror. The work is important because correction of bias is an important consideration for assimilation of data into numerical forecast models. Two methods are investigated: one based on comparisons of measurements with the ECMWF model, and one derived from measurements of the velocity from ground hits of the transmitted laser pulse. The paper is well-organized and provides details on the correction methodology as well as performance of the correction methods described. Although the results are unique to Aeolus, and therefore are likely of somewhat limited impact for other instruments, the analysis showing the impacts of temperature gradients across the mirror and the conclusion that empirical corrections can be successfully applied are potentially important for addressing unanticipated problems in that crop up in future missions. Although I think the paper could be published as is, there are a few places in the text where a bit more detail and explanation might be useful to the reader. I leave the decision on whether to request these changes to the discretion of the editor.

Specific comments:

Line 207: It isn't clear to me how the 86 km averaging of the Rayleigh channel is taken into account when comparing the AUX_MET data with the Level 2B results. The text implies that the nearest neighbour from the model is compared to the L2B data, but the discussion seems unclear to me on issues such as 1) Are the O-B statistics comparing an 86 Km average with a single point from a 9 km grid-spaced data set, and 2) Is the level 2B HLOS measurement placed at the centre of the 86 Km swath? Perhaps I'm missing something here, but it seems that some clarification on the details of the comparison would be useful here.

The L2B processor uses a nearest neighbor approach in the horizontal dimension and just uses the closest profile provided in the AUX_MET file in the selected time window. In the vertical dimension a spline interpolation is used to get a value at the proper altitude. We do not do any area averaging and do not use something like an observation operator. More precisely, the centre-of-gravity (CoG) location of the L2B winds is used to derive the values from the AUX_MET for the O-B statistics. The horizontal CoG location of the L2B winds is determined by the CoG of the signals that were included in the accumulation. Due to the classification into clear/cloudy measurements this may deviate significantly from the center of the 86 km group length for the Rayleigh channel (or 14 km group length for the Mie channel). For the vertical location within the range bin the center position is considered.

ECMWF tested the impact of an averaging operator on the Rayleigh clear wind O-B and found it only improved the stdev(O-B) by 0.04 m/s relative to a point-like operator and didn't have any detectable bias improvement. This test was done using AUX_MET data. So, a point-like observation operator is considered to be sufficient for M1 T related bias correction. Note that the AUX_MET IFS profiles are provided along the orbit every 3 seconds (~21 km spaced); despite the underlying ECMWF model being run at higher resolution of Tc01279 (~9 km grid spacing). The grid-spacing of 9 km does not give a true reflection of the model resolution however, the effective resolution has been estimated in the past to be ~4-8 times the grid spacing (https://www.ecmwf.int/en/elibrary/17358-effective-spectral-resolution-ecmwf-atmospheric-forecast-models). This effective resolution of 40-80 km probably explains the negligible improvement in O-B statistics by accounting for Aeolus' footprint (averaging). Further information about the O-B calculation was added to the manuscript:

> files is provided every 3 seconds along the orbit at 137 model levels interpolated (nearest-neighbor) to the Aeolus track. To compute observation minus background (O-B) statistics, t~~T~~he L2B processor uses a nearest-neighbor approach in the horizontal and uses the closest profile in the AUX_MET file in the selected time window. In the vertical dimension a spline
>
> 8
>
> interpolation is used to get a value at the proper altitude. ~~nearest neighbor of the AUX_MET data to the L2B wind results is used to compute observation minus background (O-B) statistics on a global scale.~~ Thes~~e~~ O-B statistics ~~differences~~ have been used to analyze the systematic and random wind errors of the Aeolus observations at a global scale (Martin et al., 2021). The

Line 218: The authors should perhaps provide some evidence for the statement "O-B values are averaged over all range gates which is justified by the lack of altitude dependency". One can make a case that the physical effect that creates the temperature gradients won't change with altitude, but it isn't clear whether the statement is based on that assumption or that a comparison was used to make the case for the lack of altitude dependency.

The M1 temperature gradients only change from observation to observation. For a fixed observation the M1 temperatures are constant for all altitudes. Thus, there can be no altitude effect induced by changing M1 temperature gradients. This was clarified in the text as follows:

> can be performed. Afterwards, the O-B values are averaged over all range gates ~~which is justified by the lack of altitude dependency of the M1 bias effect~~ to yield the $E(O-B)$ value. This is justified by the lack of altitude dependency of the M1 bias effect as the measured M1 temperatures are constant over all altitudes. Figure 3 shows the typical distribution of the centre-of-gravity altitudes of the L2B Mie cloudy and Rayleigh clear wind results. This plot indicates that for the Mie channel large fraction of wind results in the lower altitudes contribute to the $E(O-B)$ statistics. In contrast, the Rayleigh wind results

Line 255: If the bias structure is strongly dependent on the atmospheric scene, that would appear to limit the effectiveness as the scene changes from day to day. I assume that the effects are a function of the time scale of the changes in cloudiness versus the temperature response of the mirror, but perhaps a bit more discussion here could be useful.

It is true that the observed bias pattern strongly depends on the atmospheric scene. However, the correlation of the bias with the M1 temperatures, and thus the underlying physical effect, does not change too much from day to day. It was found that the instrument's sensitivity towards telescope temperature variations only changes slowly with time. This manifests as a slow drift of the model coefficients with time. As a result, it is possible to train a model on day N and use it to predict the bias for day N+1.

The predictive capability, i.e. how far in the future the bias correction can still be used, is limited by the drift of the internal reference of the instrument and the slow changes of instrument's sensitivity (as mentioned above and discussed in Sec. 3.2 of the manuscript).

Line 268: This is the same issue as noted above. The time scale of the OLR changes would seem to be important if the results from the prior day are being used to correct the bias for a given measurement period.

See the response above.

Line 359: The meaning of the sentence "Note that for the reprocessing data from the same time period is used to derive the fit coefficient." isn't clear to me. Perhaps I missed an earlier reference to reprocessing.

For the reprocessing there is no need to predict the M1 induced bias ahead. The advantage for reprocessing is the availability of the complete data set for 24 h, while for NRT processing only the last 24 h are available. As a result, it is possible to apply the regression model to the same bunch of data that was used to train the model. This further improves the performance as no out-of-sample predictions with unseen data have to be performed. This was clarified in the text:

> 2019. Note that for the reprocessing data from the same time period is used to derive the fit coefficients. The big advantage for reprocessing is the availability of the complete data set which makes it possible to apply the MLR model to the same time period that was used to train the model. This even further improves the performance of the bias correction scheme as no predictions with unseen data have to be performed. The left scatter plot indicates the high correspondence between the model prediction and the measured bias values. This is demonstrated by the high R^2 value of 0.78. The right plot of Figure 8 is

Line 438: In the sentence "Note that the constant offset of about 3 m/s between O-B and ZWC values is due to the different calibration procedure between L1B and L2B winds and is not considered to be a problem for the bias correction since this offset could be corrected in the data processing", reference is made to the different calibration procedures. A reference of a bit of explanation would be useful here.

The calibration of L2B Rayleigh winds includes Rayleigh-Brillouin scattering correction, based on so-called AUX_RBC files (Dabas et al. 2008, Rennie et al. 2021 L2B ATBD)). The AUX_RBC file contains a look-up table for instrument Rayleigh responses as a function of atmospheric pressure and temperature. The AUX_RBC file is derived from an internal reference calibration measurement, representative for the internal path of the instrument. In contrast to that, the calibration of ground return winds is based on calibration measurements that are representative for the atmospheric path of the instrument in nadir mode (Reitebuch et al. 2018, L1B ATBD). So, differences in the frequency offset between the atmospheric and internal path are responsible for the observed offset between L1B and L2B winds. The references to the ATDBs and the paper by Dabas were added to the manuscript.

Line 443: Regression theory is not my specialty but some metric or reference for why 659 samples is not sufficient to use the original model might be useful here.

In case the sample size is small compared to number of covariates, overfitting can occur such that the regression model tends to describe the noise rather than the physical relationship in the data. In such a case, the capability of the model performing predictions with unseen data is drastically reduced. For the ZWC approach, differently sized linear regression models were generated and for each model the predictive skills, i.e. the ability to perform out-of-sample predictions, were evaluated. Based on that, the presented regression model was found. This information was added to text:

> not considered to be a problem for the bias correction since this offset could be corrected in the data processing.
> In contrast to the MLR model defined in Eq.1 a slightly different approach is used to describe the ZWC winds as a function of the M1 temperatures. Due to the lower sample size a simplified model with fewer independent variables has to be used. In case the sample size is small compared to the number of model coefficients, overfitting can occur, meaning that the model tends to describe the noise rather than the physical relationship in the data. In such a case, the capability of the model performing predictions with unseen data is drastically reduced. To avoid this issue, different MLR model combinations were tested and for each combination the skill in predicting the bias was evaluated. It was found that a grouping of the thermistors into two groups which describe the temperature at the outer and inner parts of the M1 mirror provides the best results: $G1 = mean(AHT27, TC20, TC21)$ and $G2 = mean(AHT24, AHT25, AHT26, TC18, TC19)$. The bias correction model is then described as follows:

Line 462: It seems the "and" before "without M1 correction" in the caption for Figure 11 could be eliminated.

Thank you. The caption was corrected for the revised version of the manuscript.

Line 477: Use of the dash to indicate the temperature range causes confusion when followed by a negative number in "0.3°C – -0.1°C". Perhaps another way to articulate the range could be employed.

This was changed in the manuscript. Now, words are used to describe the ranges: "0.3°C to -0.1°C".

Line 523: It isn't clear to me why it would not have been possible to observe the increase in random error without the bias correction. Perhaps a sentence of explanation here would be useful.

The daily averages of the M1-uncorrected STD(E(O-B)) values are dominated by the M1 effect. Figure 13 of the manuscript shows daily averages of the STD(E(O-B)) for Rayleigh clear HLOS winds before and after the M1 correction. The decrease of blue curve, showing the statistic before the M1 correction, could be misinterpreted as a decrease of the random measurement error. In fact, the decrease is related to the change of the M1 temperature conditions, having less impact on the bias at the end of the period compared to the beginning. Only after correcting for M1 effect, the "true" random error reveals. Following information was added to the text:

> of the accurate M1 bias correction, it would not have been possible to observe the increase of the random error based on wind error statistics as the M1 effect dominates the $STD(E(O-B))$ values. The daily averages of the uncorrected $STD(E(O-B))$

> values (blue curve in the top plot of Figure 13) show a decrease from the beginning of the period until October 2019, related to the changing seasonal influence of the M1 temperature induced bias. This decrease could be misinterpreted as a decrease of the random wind error. Only after correcting for the M1 effect, the true random error evolution reveals.