Response to comments of Anonymous Referee #2 on *AMT* preprint "Exploiting Aeolus Level-2B Winds to Better Characterize Atmospheric Motion Vector Bias and Uncertainty" by Katherine E. Lukens et al. Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-277-RC4, 2021

Thank you very much for your careful and helpful comments. We have revised our manuscript following your suggestions.

[Below, quotes from your comments are repeated verbatim in *green italic*. Changes in quoted text are indicated as additions and ~~deletions~~. Comments from author to author are in this color.]

The paper presents statistics of AMVs vs Aeolus observations from a dataset of collocations. The aim is to evaluate AMVs, with the ultimate goal of improving AMV quality. Aeolus provides an unprecedented dataset in this respect, in particular allowing comparisons of AMVs against other observations in regions where it was previously not possible (ocean, remote land regions).

Assessments of the quality of both AMVs and Aeolus data are highly relevant, given they are widely used as input to NWP systems. While there is hence clear merit in producing comparison statistics, I feel a scientific paper requires a clearer interpretation of these statistics than is presently provided. Furthermore, one stated goal of the paper is to guide improvements to AMV quality, but it is not clear to me whether the paper indeed provides new insights into where improvements can be made. It should be possible to address these aspects through a very major revision of the text, particularly in the results and conclusions sections, though some further analysis may also be required to draw firmer conclusions.

We have followed your suggestions as detailed in our responses below.

Main general points:

1. The paper needs to be clearer on which new insights the study provides and which overall conclusions can be drawn. Presently, the text in section 4 largely textualizes values of statistics given in tables and figures, and it is difficult to grasp what the overall interpretation of these values is and how they link to overall conclusions. One stated goal of the paper is to improve AMV quality. What do we learn about this from the study? For your consideration, Cotton et al (2021) provides a detailed list of features noted in monitoring of AMVs versus NWP, with some of them more clearly attributable to AMVs than others. Could any of these features be investigated with the collocation dataset, hence addressing the stated goal of the paper to aid the development of AMVs?

We have thoroughly rewritten Section 4 (as well as Section 5), and the key interpretations and conclusions are given in the following quote from the manuscript:

"The main findings from comparing GOES-16 AMVs with RAY and MIE winds are the following. Aeolus MIE winds show great potential value as a comparison standard to characterize cloudy AMVs. MIE comparisons generally exhibit smaller biases and uncertainties compared to RAY, reflecting the higher accuracy of MIE winds and cloudy AMVs in cloudy scenes as well as larger collocation errors for RAY winds in cloudy scenes. This is attributed to a combination of smaller Aeolus MIE uncertainties and smaller collocation/representativeness errors due to the cloudy/cloudy sampling effect, that is, the fact that both Aeolus and AMV winds are, by definition, sampling similar cloudy scenes at similar altitudes. The contribution of

Aeolus MIE uncertainty to the overall SDCD is small; in fact, removal of Aeolus uncertainties further reduces the small MIE SDCD without much change to its vertical distribution, suggesting that for MIE comparisons, the dominant factors contributing to the total error consist of AMV random errors and representativeness/collocation errors. Additionally, the AMV-Aeolus MIE comparisons depict a relatively new finding that is also noted in Cotton et al. (2020, 2021) and is largely thought to be attributed to AMV height assignment errors: a negative speed bias in the IR and WVcloud AMVs in the tropics. The fact that comparisons with Aeolus exhibit this feature hints at the usefulness of Aeolus MIE winds as a standard for comparison to characterize AMVs. (It should be noted that because the period of study is relatively short, the datasets are not large enough to examine in detail many of the "features" identified and studied in the NWP SAF AMV monitoring. However, it could be possible to verify the identification of such features in NWP comparisons with Aeolus observations by using a larger collocation dataset, which the authors are preparing and making publicly available.)

"Regarding GOES-16 RAY comparisons, sampling differences may play a role in the higher correlation between Aeolus RAY winds and WVclear AMVs, since they both represent similar clear-sky scenes. This is especially true in the tropics and NH extratropics where MCD are small and SDCD are comparable to AMV error values compared with high-quality rawinsonde winds. It is likely that collocation errors play a larger role in the RAY SDCD for IR and WVcloud AMVs due to the cloudy/clear sampling effect, where clear-sky Aeolus winds are collocated with cloudy AMVs and thereby observe different scenes, yielding larger errors. In addition, the removal of Aeolus uncertainties from the SDCD considerably reduces the RAY SDCD, particularly for IR and WVcloud comparisons, indicating that Aeolus contributes a substantial fraction of the total SDCD in the presence of clouds.

"Polar AMVs have smaller MCD for MIE compared to RAY, although Antarctic AMVs have larger SDCD than the Arctic. In fact, GEO and LEO comparisons in the SH/Antarctic exhibit the largest SDCD of all regions examined. Large wind shear is evident in the SH/Antarctic throughout much of the atmospheric column, and this can dramatically affect AMV height assignment errors. Indeed, AMV errors are shown to generally increase with increasing AMV wind speed, as do corresponding Aeolus errors for RAY winds, suggesting that both contribute to the larger SDCD observed in layers of high wind speed. Additionally, larger RAY MCD aloft could be attributed to larger collocation/representativeness errors due to IR AMVs and RAY winds viewing different scenes. The possible mischaracterization of very cold surface temperatures as clouds may also be a factor. For GOES-16 MIE comparisons in the SH, AMV errors are larger and increase with AMV speeds > 40 m s^{-1} while Aeolus MIE errors are small and remain relatively constant. This implies that the large systematic differences in MCD at upper levels in the SH extratropics are most probably attributed to larger AMV errors in combination with the wind-shear height assignment error effect."

Additionally, comparisons with Aeolus show another noted feature in monitoring AMVs by Cotton et al. (2020, 2021): a pronounced negative wind speed bias in the tropics for Meteosat-8 is evidenced in this study by large negative MCD and correspond to large SDCD in all regions (**Fig. S1**). This feature is evident in both RAY and MIE comparisons. The fact that comparisons with Aeolus exhibit another known feature hints at the usefulness of Aeolus winds as a standard for comparison to characterize AMVs.

More details are given in the responses below. After carefully considering Cotton et al. (2020, 2021) we decided the direct comparisons possible were limited because all our statistics are for HLOS winds, not vector winds.

2. What is the basis for stating that "AMVs compare well to Aeolus winds"? What does "well" mean in this context? It appears that the authors compare Aeolus/AMV difference statistics directly to values from AMV/sonde or AMV/NWP comparisons, despite very different uncertainties in the respective comparison datasets or the collocation methods. Uncertainties in Aeolus data are alluded to (incl. biases), but it is not clear how they have been taken into account.

Thank you for pointing out that the word "well" is not used clearly in this context. In various places we now state the "Overall, GEO and LEO AMVs are found to compare as well with Aeolus RAY and MIE winds as they do to conventional data sources and NWP products."

Additional AMV/NWP statistical comparisons are included in redrawn Figs. 5,7,9 (**new Figs. 7,10,13**). These figures now show profiles of uncertainty estimates with the Aeolus uncertainty removed. A description of this process has been added to the text. We have also added:

"The removal of the Aeolus error estimate results in a smaller SDCD, which still includes AMV random and representativeness errors and collocation error. The SDCD are larger for RAY comparisons than for MIE comparisons in terms of both the original (or total) and adjusted values. Although the Aeolus L2B uncertainty is highly dependent on the time period and processor used to determine the HLOS winds, it is the correct uncertainty estimate for our study."

3. The statistics presented are affected by collocation/representation error, as well as biased sampling, and my impression is that this may play a considerable role. This aspect should be discussed and, if possible, an attempt at quantifying the magnitude of these aspects should be made.

We did consider some method to estimate the collocation/representation errors, and we have added a discussion on how this might be done:

"We note that it might be possible to estimate the statistics of the collocation and representativeness errors. The collocation difference may be considered to have three independent components: the error of the AMV winds, the error of the Aeolus winds, and the difference between the truth evaluated for the AMV and the Aeolus winds. We can isolate the first component, the AMV error, if we know the other two components, and we already have estimates of the second component, the Aeolus wind error in the L2B data. The last component is the error due to representativeness and collocation differences. The differences in time and location give rise to the collocation error. The difference in the shapes of the observing volumes gives rise to the representativeness error. If we simulate the AMV and Aeolus observations from a high-quality forecast or analysis or simulation, taken to be the truth, then we can calculate estimates of the combined representativeness and collocation errors. If the truth fields are simply interpolated to the observation locations then the calculated estimates are for the collocation errors alone."

Specific points:

1. Abstract, L26-28: The two sentences appear to contradict each other - on the one hand it is stated that comparisons are consistent with what is known, on the other hand SDCD is over SH is larger than expected.

The abstract now reads: "~~Overall~~In terms of global statistics, QC'd AMVs and QC'd Aeolus HLOSV are highly correlated for both observing modes. …

~~correspond well with QC'd Aeolus HLOS wind velocities (HLOSV) for both Rayleigh-clear and Mie-cloudy observing modes, despite remaining biases in Aeolus winds after reprocessing.~~ Stratified comparisons ~~Comparisons~~ with Aeolus HLOSV are consistent with known AMV bias and uncertainty in the tropics, NH extratropics, and in the Arctic, and at mid- to upper-levels in both clear and cloudy scenes."

2. Abstract L35-39: While I agree with what is stated in this paragraph, this has been recognised for some time (see, for instance, Menzel et al 1996 http://cimss.ssec.wisc.edu/iwwg/iww3/p197-205_Menzel-Improvements.pdf). It seems odd to give such well-established finding such prominence in the present abstract.

The abstract now reads: "As shown in other comparison studies, the ~~The~~ level of agreement between AMVs and Aeolus wind velocities (HLOSV) varies ~~per combination of conditions including the~~with the ~~Aeolus observing mode coupled with~~ AMV ~~derivation method~~ type, geographic region, and height of the collocated winds, as well as with the Aeolus observing mode. ~~It is advised that these stratifications be considered in future comparison studies and impact assessments involving 3D winds.~~"

3. L43/44 ("The survey recommends that radiometry-based …"): The survey considers both radiometry-based AMVs as well as lidar measurements for addressing the requirement of 3d atmospheric winds. To my knowledge, it does not make a recommendation of one versus the other. Please rephrase.

The text now reads: "The survey ~~recommends~~ found that radiometry-based atmospheric motion vector (AMV) tracking should be ~~able~~ an important approach to address the priority requirement of 3D winds."

4. L91/92: Please remove "UTC" in the context of stating overpass times for Aeolus. As stated in the text these are local times, rather than UTC.

Done. Replaced "UTC" with "LT".

5. L118-124: Given the high relevance of the Aeolus quality to the present investigation, it would be preferable to give a deeper overview of Aeolus quality assessments, and to refer to peer-reviewed papers on the subject where possible. I am not fully convinced that 3d AIRS AMVs are a suitable reference dataset in this respect. In addition, the statements regarding biases derived from the Santek et al (2021) study appear to be contradictory, with Aeolus showing larger bias against rawinsondes than AIRS AMVs on the one hand, whereas comparisons against ERA5 show similar biases.

We have replaced lines 118-124 with the following text:

"Recent studies have compared Aeolus winds with various ~~benchmark~~reference wind datasets (e.g., rawinsondes and ~~reanalyses~~NWP forecasts). For example, Martin et al. (2021) validated Aeolus HLOS winds against rawinsonde and NWP forecast equivalents for 2018-2019. They found that the estimates of global mean absolute biases and standard deviations of Aeolus based on comparisons with rawinsonde, the ECMWF Integrated Forecasting System (IFS), and the German Weather Service (DWD) forecast model reference datasets are all comparable, with bias magnitudes ranging from 1.8 to 2.3 m s^{-1} for Rayleigh and 1.3 to 1.9 m s^{-1} for Mie and standard deviations ranging from 4.1 to 4.4 m s^{-1} for Rayleigh and 1.9 to 3.0 m s^{-1} for Mie. In addition, the biases vary with latitude and season in a similar way from reference dataset to reference dataset, with the largest differences observed in the tropics and extratropics, particularly during the summer/autumn season. Similarly, Straume et al. (2020) quality assessments showed good correspondence between Aeolus L2B winds and ECMWF model winds for September 2018. Even though Aeolus exhibited random errors that exceeded the mission requirements (4.3 m s^{-1} for Rayleigh) or just met the requirements (2.1 m s^{-1} for Mie), the Aeolus winds still had a positive impact on preliminary NWP experiments. (It should be noted that the results from Martin et al. (2021) and Straume et al. (2020) characterize Aeolus winds before they were reprocessed with the significant M1 wind bias correction applied. The Aeolus bias and error estimates should improve when using the reprocessed winds.)."

In addition, we have reorganized the information presented in the corresponding paragraph to provide a clearer overview of Aeolus quality assessments.

6. Section 2.2, first 2 paragraphs: Please add which AMV dataset has been used for the various satellites (there are different producers for some of them). I am assuming it is the operational AMV dataset of each satellite operator. I wonder whether the information contained in these paragraphs would be better presented in a table.

We now add, "AMVs examined in this study (Table 1) are operationally used by the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) and are archived in 6-hour satellite wind (SATWND) BUFR files centered on the analysis times 00, 06, 12, and 18 UTC. All AMVs included in the SATWND files are produced by NESDIS, JMA and EUMETSAT. AMVs derived from sequences of GEO satellite images are observed equatorward of ~60° latitude and are stratified by type, including IR, water vapor cloudy channel (WVcloud), and water vapor clear channel (WVclear) AMVs; visible band AMVs are not used in this study. Polar AMVs (observed at latitudes poleward of 60°) are derived from cloud-tracked IR channels in areas covered by three consecutive LEO satellite images."

Related to this comment, we also changed lines 327-329 to: "The other GEO satellites are not further examined as they exhibit larger SDCD (Meteosat-8 and -11), have a much smaller sample size (Himawari-8, ~~INSAT 3D,~~ Meteosat-8 and -11), or are not actively used in NCEP operations (GOES-15 and INSAT 3D)."

Additionally, we have replaced much of the text in the first two paragraphs of Section 2.2 with a **new Table 1** that lists the total collocation counts per satellite as well as the number of observations (and % of total counts) that pass QC for each AMV type and Aeolus observing mode.

7. Table 1: I find the information condensed in this table very heterogeneous and inconsistent, and I am not convinced that it indeed provides a useful and adequate summary of (all) available monitoring statistics for AMVs. I find the table problematic for a number of reasons:

a. While I appreciate the need to condense the information provided, the choice of very broad entries (e.g., all AMVs, all levels, global) appears questionable, given that AMV monitoring statistics vary significantly by season, level, channel, satellite/producer, etc (as apparent from the present paper and many other studies).

b. The ranges indicated for some of these statistics are also rather large, and it is difficult to know what these ranges are referring to (presumably some of the variability noted above). At the same time, the very precise numbers given for some datasets (e.g., GOES-16 IR) also do not seem appropriate given the variability with seasons.

c. It is not clear why certain references have been selected for some AMV datasets, but not for others (e.g., Cotton et al 2020 and the general NWP SAF monitoring provide monthly statistics for each operational AMV dataset, not only GOES-16 IR). Also, I am sure other papers could be used here to contribute statistics.

d. Please note that values given in Cotton et al (2020) are either against the Met Office or the ECMWF system, but not the GFS. The web-address given for the Cotton et al (2020) reference should be updated to https://nwp-saf.eumetsat.int/monitoring/amv/nwpsaf_mo_tr_039.pdf).

Our reference now points to https://nwp-saf.eumetsat.int/site/monitoring/winds-quality-evaluation/amv/amv-analysis-reports/, the landing page for all the ARs.

I suggest that the authors critically review the material presented in this table. My impression is that the numbers are primarily used to put the results of the Aeolus/AMV comparisons in broad context, but that these comparisons mostly stay at a rather qualitative level. To stay in line with this qualitative use, the table could also be removed and replaced with a simple statement of typical values found in collocation statistics.

We have removed Table 1 and replaced it with the following text: "AMV performance metrics vary significantly by season, level, channel, satellite/producer, etc. (e.g., Santek et al., 2019; Daniels et al., 2018; Cotton et al., 2020; Key et al., 2016; Le Marshall et al., 2008). For example, typical values of AMV wind speed bias acquired from seven different data producers and verified against rawinsonde winds can range from -1.8 m s^{-1} to 0.3 m s^{-1}, and wind speed uncertainty represented by standard deviation can range from 4 to 6.5 m s^{-1}, with higher vector wind root mean square errors of 6-9 m s^{-1}. Even for a single satellite, e.g., GOES-16 or Aqua, speed bias and uncertainty can vary geographically as well as vertically."

8. Table 1: Which QI has been used to quality-control the AMVs (forecast-dependent or independent) for the studies shown? The choice of QI can have a significant impact on monitoring statistics.

The text now states that "For all AMVs, a forecast-independent quality indicator (QI) of at least 80% is used to filter and retain the high-quality data."

9. L187-192 ("Since it takes approximately 92 min... closest in the vertical to the AMV."): I struggle to understand these sentences. Are the authors saying that if multiple Aeolus winds fulfil *the collocation criterion then the Aeolus profile closest is space is used, and within that profile the observation closest in pressure?*

Yes, as stated at L177: "Then, if multiple Aeolus observations still meet all collocation criteria, the observation closest in pressure to the AMV observation is kept for analysis." After this sentence we add "There is no need to consider closeness in time given the collocation criteria and the Aeolus orbit." And eliminate the text in L187-192.

10. L187-192: Are Mie/cloudy and Rayleigh/clear winds collocated separately here or are they treated together? Ie, could the same AMV be collocated once with a Mie/cloudy Aeolus wind and once with a Rayleigh/clear wind?

At the beginning of this section we add, "AMV collocation datasets are prepared separately for RAY and MIE winds. (A single AMV might appear in both data sets.)"

11. L194-203: I note that the text does not mention an outlier removal (ie removal of collocations that show particularly large deviations). Please confirm that no outlier removal has indeed been applied. I note the absence of egregious outliers in Figures 4 and 6, hence the question.

At the end of this paragraph we add, "No explicit outlier QC is applied and since there are no extreme outliers (seen below in Figs. 6 and 9), the QC that is applied is sufficient to eliminate them."

12. L201: Which QI has been used for quality control in this study?

See our response to point 8.

13. Fig. 2: Please clarify what grid cells have been used in this plot and whether they are of equal area size. The caption states that each grid cell is 1.25° or 140 km, but the former would lead to progressively smaller cells at high latitudes and is incompatible with the 140 km.

The figure captions now read, "within a grid cell at 1.25° (approximately 140 km in the N-S direction) horizontal resolution".

14. Tables 2-5: Given the considerable variations shown by channel, wind type, level in Fig. 5 (and other Figures), how useful are the statistics given in these tables? Also, it is very difficult to grasp the information conveyed in this way - would replacing the table with a graphical display help?

This is an important point. To be clear the first paragraph of Section 4 now includes: "In agreement with previous studies, our results confirm that the level of agreement between AMVs and Aeolus winds varies per combination of conditions including the observing scene type (clear vs. cloudy) coupled with AMV type, geographic region, and height of the observable. Moreover, the findings highlight the value of using Aeolus MIE winds as a comparison standard to characterize AMVs. For context, we begin with summary statistics for samples that include all conditions."

In addition, Tables 2-5 have been replaced with **new Figs. 2 and 3** that display the same statistical information presented in the tables. The statistics are easily compared for each satellite in each geographic region and clearly show that RAY comparisons tend to have larger SDCD than MIE, and MCD and SDCD in the SH are generally larger than the other regions. The results are summarized in the text.

15. L274/275 ("Overall, GEO AMVs correspond very well with RAY and MIE winds..."): See general point 2 above.

The text now reads: "~~Overall, GEO AMVs correspond very well with RAY and MIE winds, with~~ The main points from the summary collocation statistics of RAY and MIE winds with AMVs are the following: …"

16. Figure 4: the text on the plot is very small and hence difficult to read (ie axis labels, and summary statistics).

Figures 4, 6, and 8 (**new Figs. 6, 9, and 12**) have been redrawn.

17. 4.1.1: A common thread throughout this sub-section seems to be the finding that WV clear AMVs compare better with the Rayleigh/clear winds than the cloudy IR or WV winds (stated multiple times). I think some critical discussion of this finding would be useful. The fact that cloudy AMVs were found in a region where Aeolus indicates a clear scene suggests that either the AMV height assignment is erroneous or that collocation/representation errors are likely to be larger (as the Aeolus wind must originate from a different area than the AMV). So by design these statistics for cloudy AMVs are expected to be less favourable than the ones for clear AMVs. Without further analysis the statistics will give little insight in the relative quality of clear-sky AMVs vs cloudy AMVs in general.

We have expanded our discussion of this finding. The manuscript now reads:

"Of the three AMV types, the best match is for WVclear AMVs, with the comparisons exhibiting the smallest SDCD values in each geographic region, that in turn are comparable to known wind speed SD and RMS of all GEO AMVs relative to rawinsonde winds (Santek et al., 2019). This is expected since WVclear AMVs and Aeolus RAY winds are most probably sampling similar clear-sky scenes, and clear scenes are more homogeneous over time and space scales, which in turn implies smaller collocation differences. Ideally, one would expect samples large enough to provide statistically significant collocation differences between RAY winds and WVclear AMVs only; as it turns out, collocation differences are also statistically significant for IR AMVs (see Fig. 7). In these cases cloudy AMVs are collocated with Aeolus RAY winds that represent clear scenes, and since they do not observe the same type of scene, Aeolus and/or AMV representativeness errors are most probably larger (hereafter we refer to this as the cloudy/clear sampling effect).

"To better isolate the AMV error, the Aeolus error estimate is removed from the SDCD at each level, resulting in mean profiles of adjusted SDCD (long dashed lines in Figs. 7b, 7e, and 7h) that include AMV errors and collocation/representativeness errors. Overall, the adjusted SDCD for all AMV types exhibit similar magnitudes and distributions in each geographic region throughout the vertical. WVclear comparisons have slightly smaller adjusted SDCD at upper levels, suggesting that sampling differences may play a role in the higher accuracy observed for WVclear AMVs, given that WVclear representativeness errors are likely small due to Aeolus RAY and WVclear AMVs observing similar scenes. Aeolus RAY uncertainty is larger in the presence of clouds and appears to have a considerable impact on the corresponding SDCD, as the reductions in IR and WVcloud SDCD (~1 m s^{-1}) are larger than for WVclear (0.5 m s^{-1}).

"Regarding GOES-16 RAY comparisons, sampling differences may play a role in the higher correlation between Aeolus RAY winds and WVclear AMVs, since they both represent similar clear-sky scenes. This is especially true in the tropics and NH extratropics where MCD are small and SDCD are comparable to AMV error values compared with high-quality rawinsonde winds. It is likely that collocation errors play a larger role in the RAY SDCD for IR and WVcloud AMVs due to the cloudy/clear sampling effect, where clear-sky Aeolus winds are collocated with cloudy AMVs and thereby observe different scenes, yielding larger errors."

18. 4.1.2: Related to the above, I note that in this section clear-sky AMVs are excluded from comparisons with Mie/cloudy Aeolus winds, with the argument that clear-sky AMVs measure wind in clear scenes only. The choice is inconsistent with the choice made in 4.1.1, where comparisons of cloudy AMVs vs Rayleigh/clear Aeolus winds were included. Could the authors elaborate on the reasons for these two different choices?

We now add, "To increase the size of our collocation data set, we compared all types of GOES-16 AMVs to both Rayleigh-clear and Mie-cloudy winds. In addition, we do not show results from WVclear AMV collocations with Mie-cloudy winds as correlations for this category of collocations are poor and the sample size is very small (see Table 1), and this result may be unreliable. With a larger data set it might be possible to compare Rayleigh-clear and Mie-cloudy winds to clear and cloudy AMVs only, respectively. Additionally, winds retrieved from tracking clear-sky and cloud motions represent different dynamical features and tend to behave differently. For example, the recommended time interval for tracking cloud motions is 10-15 minutes to capture short cloud lifetimes and rapid intensification/deformation, while the recommended time interval for clear-air motions of 30 minutes is suitable to capture variations in jet streams and other clear-air features (Schmetz et al., 2000)."

19. 4.1.2: Similar to the point above, the effect of the sampling imposed by looking at Mie/cloudy vs cloudy AMV collocations should be discussed here. By design this is a sample where Aeolus and AMVs agree in terms of a cloud being present at a particular altitude. So this sample of AMVs would be expected to have smaller height assignment errors (as the height assignment has effectively been quality-controlled by Aeolus), and representation errors are likely to be smaller (as AMVs and Aeolus are more likely sampling similar areas). This will contribute to favourable comparison statistics. Of course, the smaller random error in the Mie/cloudy wind is another reason for smaller SDSCs compared to values shown in 4.1.1. Based on Aeolus uncertainty estimates, is it possible to quantify which aspect is the dominant factor?

We now add, "MIE SDCD are considerably smaller than those for RAY comparisons, and this is attributed to the general higher accuracy of Aeolus MIE wind retrievals. Another possible reason is that MIE comparisons might generally have smaller collocation errors: because collocated Aeolus MIE winds and IR and WVcloud AMVs are by definition more likely sampling similar cloudy scenes at similar altitudes, we expect the Aeolus and AMV random and representativeness errors to be small (hereafter the cloudy/cloudy sampling effect).

"…Aeolus MIE winds show great potential value as a comparison standard to characterize cloudy AMVs. MIE comparisons generally exhibit smaller biases and uncertainties compared to RAY, reflecting the higher accuracy of MIE winds and cloudy AMVs in cloudy scenes as well as larger collocation errors for RAY winds in cloudy scenes. This is attributed to a combination of smaller Aeolus MIE uncertainties and smaller collocation/representativeness errors due to the cloudy/cloudy sampling effect, that is, the fact that both Aeolus and AMV winds are, by definition, sampling similar cloudy scenes at similar altitudes. The contribution of Aeolus MIE uncertainty to the overall SDCD is small; in fact, removal of Aeolus uncertainties further reduces the small MIE SDCD without much change to its vertical distribution, suggesting that for MIE comparisons, the dominant factors contributing to the total error consist of AMV random errors and representativeness/collocation errors."

20. L450-453: The relatively large systematic differences over the SH extra-tropics appear to be attributed to AMVs, as the authors suspect height assignment errors. Are there any reasons to believe that Aeolus winds could be in error in this particular region?

To address this comment, we now include in **new Figs. 7, 10, and 13** profiles of the mean AMV wind speed (not projected onto HLOS) from which one can infer vertical wind shear by the gradient of the AMV wind speed with respect to pressure. In addition, we include figures of the decomposed AMV error and corresponding Aeolus error with respect to AMV wind speed (**new Figs. 8, 11, and 14**). The manuscript now states:

"MCD are largest in the SH extratropics and are statistically significant throughout the vertical, ranging from -1.0 m s^{-1} at low levels to < -3.0 m s^{-1} above 300 hPa. Strong wind shear corresponding to an intensified jet is inferred at upper levels (Fig. 10g). The larger MCD aloft are associated with increases in adjusted SDCD with height, which are on the order of 4-5 m s^{-1}. Moreover, the large MCD represent over 8.5% of the corresponding HLOSV at upper levels and could be attributed to larger AMV height assignment errors corresponding to stronger storm tracks in winter. This is exemplified in Fig. 11e where Aeolus MIE errors are shown to be small (2 m s^{-1}) and remain relatively constant with increasing AMV wind speed, while the adjusted SDCD are larger and increase with AMV wind speeds > 40 m s^{-1}. The results imply that the large systematic differences in MCD at upper levels in the SH extratropics are most probably attributed to a combination of larger AMV errors in combination with strong wind shear."

21. L454-455: The biases exceeding -3 m/s over the SH extra-tropics are not small, and they are not in line with the ranges given in Table 1. This seems to be acknowledged later in the same paragraph (L459-460), but the sentences in question expresses the opposite.

Thanks for catching that. The text now reads, "Statistics of AMV minus MIE collocation differences are generally consistent, albeit with some notable exceptions, with those for AMV comparisons with high quality rawinsonde winds."

22. L536-538 ("Overall, GEO and LEO AMVs are found to correspond very well with Aeolus RAY and MIE winds... range of known biases and uncertainties of AMVs"): See general point 2 above.

The text now reads, "Overall, GEO and LEO AMVs are found to compare as well ~~correspond very well~~ with Aeolus RAY and MIE winds as they do to conventional data sources and NWP products."

23. L550 ("GOES-16 AMVs are found to compare well with RAY and MIE winds"): As above, see general point 2.

The text now reads: "The main findings from comparing GOES-16 AMVs ~~are found to correspond well~~ with RAY and MIE winds are the following."

24. L552-553 ("WVclear AMVs perform best..."): See earlier point 17. This is likely at least partially due to biased sampling, and without further analysis it would be inappropriate to conclude that WVclear AMVs are more accurate than WVcloudy AMVs. This should be clearly addressed when interpreting the results. A similar comment applies to abstract L29/31.

See our response to point 17.

25. L570 ("... Aeolus could be used as a standard for the comparative assessment of AMVs pending additional bias corrections to the Aeolus L2B winds"): I am not sure what the authors are saying here. Are the results presented not reliable, as additional bias correction for Aeolus winds is required? Or do the authors think that their results suggest that additional bias correction is needed for Aeolus? I don't think there is sufficient evidence for either statement, so I am puzzled what is meant here. A similar comment applies to abstract L25/26.

Our results are reliable. We now add: "The Aeolus project has done much to eliminate errors of all types, but some improvements are expected. For example, some of the bias corrections currently applied depend on ECMWF forecasts and the analysis of Liu et al. (2021, TLSBC) demonstrates that additional bias correction for Aeolus are possible, and that such corrections can improve NWP analysis and forecast results (Garrett et al., 2021, Impacts)."

Supplemental Figure

Figure S1: As in manuscript Fig. 1 but showing Meteosat-8 vertical profiles of AMVs collocated with Aeolus RAY (top) and MIE winds (bottom) in the tropics.