Atmos. Meas. Tech. Discuss., referee comment RC4
https://doi.org/10.5194/amt-2021-277-RC4, 2021

Comment on amt-2021-277

Anonymous Referee #2

Referee comment on "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

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.

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?
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.
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.

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.
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.
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.
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.
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.
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.
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).
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.
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.
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?
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?
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.
12. L201: Which QI has been used for quality control in this study?
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 1.25º or 140 km, but the former would lead to progressively smaller cells at high latitudes and is incompatible with the 140km.
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?
15. L274/275 ("Overall, GEO AMVs correspond very well with RAY and MIE winds…"): See general point 2 above.
16. Figure 4: the text on the plot is very small and hence difficult to read (ie axis labels, and summary statistics).
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.
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?
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?

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?

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.

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.

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

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.

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.