General comments:

The authors compared the Aeolus wind measurements with Ka-radar, wireless soundings, ECCC and ERA5 reanalysis data, and the results show that the Aeolus wind measurements are in good agreement with other wind measurements or reanalysis data. This paper has important implications for the application of Aeolus wind measurements in the Arctic, where wind measurements are currently scarce. However, the paper requires significant revision before publication, and the specific issues are described below.

1. The authors use other soundings and reanalysis data as a benchmark for comparison with the Aeolus wind data, but the data quality of the other data is not presented in the paper. Perhaps the other wind data also have large biases in the Arctic, and then the authors' use of them as a comparison benchmark would make none of the comparisons in this paper credible. Especially for high altitude wind fields, the data sets may have different data quality performance at different heights. Moreover, the reanalysis dataset itself contains assimilation of existing sounding data. The authors need to fully justify the reliability of the data in each of the datasets used in this paper.
2. The authors' presentation of the spatio-temporal matching process between the datasets is not clear enough. It is suggested that the data matching process be introduced as a separate section and the information about the data used in this paper be summarized. A clear and detailed data matching process is desired to be seen.
3. the authors' analysis of Figure 5 is not rigorous enough and the conclusions drawn from Figure 5 are not reliable enough, see Specific comments L296.
4. The authors need to clarify the practical significance of the discussion of the statistical distribution of the wind products themselves in Figures 7 and 8. It might be more valuable to discuss instead the distribution of the difference between the Aeolus data and the other data.

Specific comments:

L82:*1 December to 31 January 2020*
Wrong time markings

L115: ECMWF has recently published the first reprocessed data (2B10)
Data links should be added after the data introduction.

L125: The quality control recommendation following the Guidance for Aeolus NWP Impact Experiments (Rennie and Isaksen, 2019), including the threshold for L2B estimated observation errors.
The recommendation given by NWP is a threshold range. Authors should submit specific thresholds to be used when processing data.

L129: We further reject the outliers by excluding all the data when the difference between the observations and ECCC-B or ERA5 is greater than 30 m/s.
For the screening threshold of 30m/s, the authors need to give an explanation, either from data

analysis or literature.

L162: ECCC-B is then linearly interpolated to Aeolus measurement locations and times.
The process of linear interpolation needs to be clarified. In addition, the main comparison data in the latter section does not provide complete information on the vertical resolution, and due to the large number of datasets used in this paper, it is recommended to use a table after section 2.5 to summarize the important information of each dataset used.

L165: Reanalysis ERA5
For the ERA5 dataset, assimilation relationships between it and the other datasets used in this paper should be added.

L190-L210:
From the latter, there is very little overlap between Ka-band radar or LIDAR with Aeolus data. The main object of comparison is the radiosondes, so its data should have been presented in more detail. In particular, the horizontal drift problem of radiosondes, which seems to be not taken into account by the authors, may also lead to a large bias. In addition, is the introduction and use of Ka-band radar and ground-based LIDAR data in this paper meaningful? It seems that the absence of these two does not affect the logic and conclusions of this paper.

L246：On 24 September, Aeolus measures westerly winds in reasonable overall agreement with the other data.
The agreement between Aeolus and the other data in Figure 2b does not seem to be obvious, and the deviation in some data points is already close to 50%.

L249:
Figures 3 and 4 are identical and need to be modified. The comparison of Ka-band radar in Figure 3d needs to be addressed for its significance, since there are only about 10 data points. Also the information on the fitted straight lines, standard deviation and sample size in the figure can be added, while the information in the supplemental Figure S1 will be placed in the original text in the form of a table.

L270:
The data consistency performance of the two sites is different, and the data consistency of ERA5 and ECCC-B with Aeolus in the Whitehorse site didn’t change, and the conclusion given by the authors as well as the explanation is not reasonable enough.

L275-L292:
The authors' discussion of Ka-band radar consistency with Aeolus and its causes seems unnecessary for this paper, as there are too few overlapping data points and no valid conclusions are drawn.

L296:

Are the sample sizes in Figure 5 the same for the three time periods? The use of the expression summer, fall, and winter is not rigorous and should be specific to dates. In addition, the discussion of Figure 5 is inadequate. If solar radiation is used to explain Figure 5, then why did the overall performance of fall 2018 be better than that of winter 2020? The Mie channel also performed better in summer 2019 than in winter 2020, and the Mie channel is also influenced by solar background radiation. The authors seem to have overlooked some phenomena in their haste to reach conclusions.

L355：Since ECCC-B and ERA5 are mutually consistent

This may need to be more fully demonstrated.

L359-L364:
The conclusion of this paragraph does not need to be obtained by data analysis. It is just a mathematical law. The discussion of the longitudinal and latitudinal components also seems to be unnecessary.

Figure 8:
What is the significance of discussing the mean and standard deviation of the wind speed samples themselves? Clarification by the authors is needed. It may be more meaningful to discuss the distribution of the differences between the Aeolus wind measurement data and the ECCC-B and ERA5 data.

L377: Figure 8 shows an overall agreement between Aeolus, ECCC-B, and ERA5
The proof of consistency between data by comparing data distribution characteristics only is not enough.

L391: Figure 9 shows that Aeolus data consistently has more structure than ECCC-B during all three periods and for both Rayleigh and Mie winds.
What does "more structure" mean here? Please explain.

L395: During the boreal summer period, the data in the stratosphere seem to agree less with the ECCC-B data, reflecting reduced sampling, solar background noise that is most effective during summer as mentioned earlier, and other possible errors (Reitebuch et al., 2020).
The derivation of this conclusion is not rigorous, there are many possible reasons for the decrease in the correlation between the Aeolus data and ECCC-B data in the stratosphere in summer, and it is also possible that it is caused by changes in the atmospheric environment in summer, thermal changes in the telescope, etc. I don't think we can make a speculation on the cause from Figure 9. But the authors seem to attribute it to solar radiation in the abstract.

L402: For this reason, in the next paragraph, where we investigate the spatial distribution of the consistency in the lower and upper atmospheric regions, we exclude the Rayleigh winds in the PBL and the Mie winds in the stratosphere.
It would have been more convincing if the authors had made this data trade-off from the analysis

of data used in this paper. Although Aeolus was designed to complement the dual channels at altitude, there are many cases where the Mie channel has higher data volume and data quality at altitude than at lower altitudes, especially in summer. Simply removing the data would be detrimental to the subsequent analysis.

Figure10, Figure11:
The calculation process of RMSD in this paper should be clearer, and it is better to give the formula and the range of RMSD to be considered "consistent". In addition, the radial mutations in these two figures seem to be inconsistent with common sense, especially in Figure 11e, which I hope the authors can explain. Also, the same color scale should be used for all subplots. The number of data samples used for different subplots should be provided, because the valid sample size may vary greatly for different seasons at different altitudes. In addition, the data density of Aeolus is also different for different latitudes, how did the authors deal with this point, and does this lead to lower quality data for lower latitudes?

L446: No significant improvement is seen here because we have implemented a weekly updated dynamic bias correction to the near real time data.
How do the authors explain the slightly higher RMSD of 2B10 data compared to 2B06 in Figure 10 and Figure 11? What is the meaning of the dynamic bias correction mentioned in the paper, please elaborate, and what is the difference between this correction and the reprocessed data correction in the L2B product?

L476: We have found some initial evidence that the estimated error product is also a good predictor of RMSD between Aeolus and the reanalysis, which could be useful for constraining future forecasts.
The "estimated error product" itself is used to estimate the difference between the Aeolus wind product and the true wind field. If it does not predict the RMSD between Aeolus and reanalysis data, then the reanalysis data deviates from the real wind field. I don't understand the purpose of the author's statement, perhaps more information about "constraining future forecasts" is needed.