We thank reviewers 2 for his/her time and the valuable comments and suggestions. Please find below the response from us indicated in *italic*.

- Line 10: “proof” should be “prove” or “provide proof”.
  Thanks – changed!

- Line 34: It should be noted that Atmospheric Motion Vectors (AMVs) can also be computed by tracking features in the water vapor field. *Thanks for this information, we added it to the text and added a new reference (Bormann, N., S. Saarinen, G. Kelly, and J. Thépaut, 2003: The Spatial Structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data. Mon. Wea. Rev., 131, 706–718,)*

- Line 153: It isn’t obvious to me that the Rayleigh_cloudy wind product is without value. I realize that the presence of aerosols complicates the wind retrieval in the Rayleigh channel, but I’m not aware that the Rayleigh_cloudy product is deemed totally useless. If the Aeolus project team has stated this then the authors should provide a reference.
  You are right as reviewer 1 has also mentioned and I have also used this wind type for investigations. We changed the text accordingly:
  “Two out of this four wind products, namely the Rayleigh_clear and the Mie_cloudy winds, are the main target for the operational use of Aeolus data in NWP”… “The Rayleigh_cloudy products may also deliver usable wind measurements, but contamination of Mie scattering need to be corrected first which is yet at an experimental stage. Thus, we will use only Rayleighclear and the Miecloudy product for our analysis”

- Lines 157-162: Although references are provided, as a reader it would be nice to have a few sentences describing in general terms how the error threshold and validity flags are computed.
  We added some few sentences concerning that. But we think, the full explanation of the validity flag is not needed when the reference is given:
  “…These thresholds are chosen subjectively, based on the compromise between the number of observations that pass the quality control and the overall quality of the dataset (Rennie and Isaksen, 2020).” …”The validity flag (de Kloe et al., 2016) considers the validity of the products. Several different technical, instrumental and retrieving checks account for this flag. .“

- Line 200: Perhaps I missed it, but it would be useful to state in the text that because the Aeolus lidar beam is not nadir-pointing, the horizontal distance from the radiosonde to the Aeolus measurement volumes changes as a function of height as well as radiosonde movement. It’s a simple and obvious point, but it can’t hurt to note it.
  We have added this information accordingly: “As Aeolus is not pointing nadir but is taking measurements 35° off-nadir, the horizontal distance of the Aeolus observations to RV Polarstern is different for the different heights in the Aeolus wind profile. Also the radiosonde drifts along the wind direction, thus the distance to between the Aeolus measurements and the radiosonde changes during the ascent. The effect of both is illustrated…”

- Figure 3 and Figure 7: In looking at the figures on a laptop, I found it somewhat difficult to differentiate the colors in the Mie cloud and Rayleigh clear plots. Perhaps the authors could use a different technique for separating the plots, such as dashed or dotted lines.
  Thanks for this advise. We reshaped all corresponding figures accordingly so that colors are not needed anymore and we hope that they are now more clearly readable.

- Line 247: The inability of Aeolus to characterize the maximum wind under strong shear conditions near the tropopause is useful to point out. However, it should probably be noted that this isn’t an error in the Aeolus measurement, but rather an averaging effect that obscures an important

parameter.
You are right, we've added: "This is in principle no measurement error of Aeolus."

- Line 253: Changing the range bins on Aeolus to 1 km has potentially negative consequences on the measurement in that it reduces the number of photons available, thus increasing the random error. The authors might want to comment on whether the Aeolus team chose to accept this increase in random error or compensate for it by, e.g., reducing the horizontal resolution.
  This is a good point. But as we are "only" a Cal/Val team and not any decision-making body, we would not like to comment too much on these issues. Nevertheless, we've added: "…but accepting the drawback of an increased random error."

- Figure 5: it would be nice to provide a N-S reference on the plots.
  Thanks, we have added this to the plots!

- Line 353: A sentence explaining why the authors prefer to use MAD as the statistic to represent the random error would be useful.
  As also raised by reviewer 1, we meanwhile provide the scaled MAD as an indicator for the random error. This is explained in the text and also that the MAD and thus also the scaled MAD is less sensitive to outliers in contrast to the standard deviation:
  "The median absolute deviation (MAD) of the distribution is used to calculate the random error of the Aeolus wind observations (Lux et al., 2020; Witschas et al., 2020) because it is less sensitive to outliers than the standard deviation. It is 67.4% of the standard deviation or the other way around, the scaled MAD (MAD times 1.4826) is identical to the standard deviation for a perfectly Gaussian distribution. The scaled MAD is thus used an indicator for the random error for Aeolus observations. The MAD is in case of the Rayleigh clear winds 3.26 m/s, the scaled MAD correspondingly 4.84 m/s.

- Table 2 Caption: The caption seems to be defining medium absolute deviation (MAD) as MAD - random error, which doesn't make sense.
  You are right. Reviewer 1 also raised this point. Thus, we have added a column for the scaled MAD which is representative for the random error and rephrased the caption accordingly.