Anonymous reviewer #2:

An assessment of the Aeolus L2B Rayleigh-clear and Mie-cloudy wind product quality is presented by this study, using an impressive network of L-band radiosonde stations over China, complemented by ERA5 data. The study focuses on the seasonal variations of the Aeolus wind product quality, including a comprehensive analysis of the impact of wind direction and presence of clouds in different altitude levels. The presented results are valuable for further analysis of satellite products over China, including aerosol and air quality studies. The manuscript is well written, but some major revisions are needed before publication, which are described in the following:

1. In my opinion, the phrase 'Aeolus detection performance' is a bit misleading. I would expect that this is used for minimum detectable aerosol layers or more the limitations of the instrument. But you are analysis the wind product performance. I would advise to avoid this phrase and rather use 'Aeolus wind product performance' or 'Aeolus wind data quality'.

2. You define the relative error as representative for the detection performance (product quality…). Other studies, like Martin et al. 2021, Baars et al., 2020, Witschas et al. 2020, Lux et al. 2020 all use the scaled Median Absolute Deviation (MAD) as a measure of the product quality, which should be closest to the random error of Aeolus winds, compared to the ECMWF model (as done in NWP monitoring). Your results would fit better in the selection of Aeolus validation studies, if you would also show the scaled MAD.

3. You applied a collocation distance of 2.5 deg lat/lon for valid overpasses. However, the cal/val implemtation plan (https://earth.esa.int/eogateway/documents/20142/1564626/Aeolus-Scientific-CAL-VAL-Implementation-Plan.pdf) suggests to use a horizontal collocation distance of 100 km around the ground site, which is much stricter than 2.5 deg. I understand that depending on the wind direction, the radiosonde might drift closer towards the actual Aeolus wind profile, but some further justifications of the 2.5 deg criteria would be desirable. Particularly for Mie-cloudy winds you can have significant deviations when using such a large distance.

Detailed comments:

Line 15: Relative error 174% higher than…

> I would recommend to mention mean random wind errors of Aeolus against L-band radiosonde data. For external readers, a random error in m/s is more easily understandable than a %.

Line 25: which uses a single-view detection method to scan the global three- dimensional wind field from space…

> ALADIN does not scan the 3D wind field, just the Horizontal line-of-sight (HLOS) component of the 3D wind field. Please rephrase.

Line 32: Level 2B products provide HLOS wind data after actual atmospheric correction and deviation correction

Better: fully processed HLOS wind profiles after correction of temperature and pressure effects.

Line 34: relevant researchers

dedicated calibration and validation team have carried out ...

Line 48: After the implementation of the new M1 deviation correction scheme, the effect of system thermal performance changes on Aeolus' seasonal fluctuation is theoretically excluded.

… is significantly reduced which led to systematic errors lower than 1 m/s (Rennie and Isaksen, 2020)

Line 55: Data and methods:

I would recommend to rearrange the chapter just a little bit. Start with Aeolus L2B, L-band, ERA5 and put the description of the collocation criteria together with Fig. 1 under 2.4, when you explain the data matching.

Line 76: What is the wind uncertainty of the L-band data. Please add a reference for the radiosonde wind bias here.

Line 85: You mention that ERA5 data is often used as reference in meteorological data analysis. However, we know that modelled wind predictions have high uncertainties too, particularly in cloudy situations. Please add some information of the expected model wind uncertainties in this chapter.

Line 99: Conversion from ERA5 to Aeolus HLOS:

In Eq. 1, you use the wind direction and horizontal wind speed of the radiosonde data and the Aeolus azimuth angle to reproject the RS data to Aeolus data. I don't understand, why you do the same with the ERA5 data, where you have the full u and v wind vector information.

Wouldn't be the correct conversion: $HLOS = -u \cdot \sin(\phi) - v \cdot \cos(\phi)$?

Line 115: Please spell check the variable V_ture, it should be V_true I guess?

Line 124: You define the relative error as representative for the detection performance (product quality…). Other studies, like Martin et al. 2021, Baars et al., 2020, Witschas et al. 2020, Lux et al. 2020 all use the scaled Median Absolute Deviation (MAD) as a measure of the product quality, which should be closest to the random error of Aeolus winds, compared to the ECMWF model (as done in NWP monitoring). Your results would fit better in the selection of Aeolus validation studies, if you would also show a scaled MAD.

Fig. 3: It is recommended, also during several Aeolus validation workshops, to have HLOS Aeolus on the y-axis and all reference observations, models on the x-axis. Please swop Aeolus and L-band on the fig.3 (a,d). I would further advise to have Aeolus-L-band first,

Aeolus-ERA5 in the middle, and ERA5-L band as last plot. You are describing (a,b…) but the letters are not shown in the plot.

Lines 146-153: You are mentioning the space time matching problem. In the Cal/Val implementation plan, it is recommended to use a collocation criteria of +-100 km (which is much less than 2.5 deg) and +-60 min in time. One possibility could be to show the results for all measurements and one table for all within the 100 km distance. This may improve your results.

Fig. 5: In order to compare it easier to the random error evolution, presented by Rennie and Isaksen, 2020, I would recommend to show the scaled MAD, additionally to the relative error. As mentioned before, other studies one the scaled MAD as the main estimate of the wind product random error.

Lines 194-196: The fact that the relative errors are larger in summer 2020 compared to summer 2019 is mainly caused by the decrease in the output laser energy of ALADIN. This effect is less pronounced for Mie-cloudy winds, because they are not as strongly depended on the laser energy.

Lines 205-210: On the other hand, the different situation of the Mie-cloudy group and the Rayleigh-clear group does not seem to support the explanation that the Aeolus system itself causes seasonal changes in detection performance…

> What do you mean with that? Please add some further explanations on that. NWP monitoring results and other independent validation studies show that the Mie-cloudy bias is much less effected by thermal variations compared to the Rayleigh-clear bias. This can be explained by technical differences of the Mie spectrometer compared to the Rayleigh spectrometer.

Line 216: The wind direction of the atmospheric wind field has an obvious seasonal trend.

> Can you provide a reference publication for this? I believe there are some manuscripts showing the seasonal trends of the wind direction in China.

Line 224-225: Based on the previous data matching work, we calculate the angle α between the real horizontal wind direction (provided by 225 ERA5 data) and the Aeolus HLOS direction (provided by L2B data) of each Aeolus valid data point…

> How do you calculate the wind direction from the HLOS data? Or do you mean that you use the azimuth angle, given in the L2B product?

Line 240: …*resulting in decrease in the energy of the laser beam and signal-to-noise ratio (SNR)*:

> Instead of decrease in energy, better say: “resulting in a decrease of the laser return signal and…”

Line 242: we takes the effective data → we take the effective data…

Line 256: Mie-cloudy data does not distinguish between aerosol and clouds. You mention the impact of aerosols here, but some further comments about the potential impact of low-level aerosol layers on the wind product quality would be desirable. Could seasonal variations of aerosol layers also influence the results?

Fig. 12: Please ad (a) and (b) to the image.

Line 319: Avoid to use double reasonable.

311-320: Have you thought about the effect of horizontal cloud variability on the Mie-cloudy wind errors? High clouds tend to be more homogeneous than low level clouds, which can vary strongly within one ~10 km Mie-cloudy wind result. This could potentially cause the higher relative errors in lower altitudes. Furthermore, the lidar signal is usually attenuated, if there are high clouds, thus I don't think that there will be Mie-cloudy valid winds from low altitude levels below high-altitude Mie-cloud valid winds.

Line 323: Besides, as the first spaceborne wind lidar, the analysis of factors affecting the detection performance of Aeolus will help provide a reference for the follow-up development of spaceborne wind lidar.

> Could you be a bit more specific what of your findings will contribute to the follow-on development of future spaceborne wind lidars? Just add some more details here.