

Response to Reviewers

Overview

We thank the reviewers for their helpful comments on our study. Based on their comments, the manuscript has been substantially improved since its initial submission. We briefly summarise the improvements here in an overview, before responding to the reviewers' specific comments.

The key improvements to the paper include, but are not limited to:

- A new section investigating equatorial waves, in response to comments by reviewers #2 and #3. This includes (a) new analyses of Kelvin wave variance during the QBO disruption (b) new analyses of the symmetric and antisymmetric power spectra, and (c) a consideration of the equivalent depth h_e and vertical wavelength L_z of Kelvin waves. These results help to clarify equatorial wave processes during the disruption, are consistent with our existing material and other studies, and have improved and expanded the scope of the our study.
- Statistical tests which have been applied to our results, including to the comparison of Aeolus with ERA5 and radiosonde measurements, and to our calculation of the equivalent depth.
- A more comprehensive description of important Aeolus and ERA5 biases which might influence the results in this study, with additional information on these for the reader's benefit.

We now respond to the specific comments by each reviewer below.

A Reviewer 1

This study provides a good investigation of the ability of the Aeolus data to contribute to the understanding of the QBO dynamics. To this end, Aeolus observational data from various range-bin settings were compared with ERA5 reanalyses and high-quality radiosonde observations. The results and discussions highlight both opportunities and limitations of the Aeolus DWL mission. The manuscript is well written, concise, and includes appropriate figures. I recommend publication after the following minor issues and suggestions are resolved or considered.

We thank the reviewer for this overview and recommendation.

2. Data and Methods *I think the order of the data description could be changed. In the subsection about the radiosonde data, there is already written about Aeolus measurement geometry, overflights and a special Aeolus setting. It would be easier for the reader to understand this if he/she has learned about Aeolus before.*

Agreed, the order of the sections has been updated.

2.2. Aeolus: *Somewhere it should be mentioned that Aeolus is a polar orbiting satellite.*

Agreed, we have mentioned this in the text.

Line 72: Please consider making two sentences out of this (The satellite's orbit is sun-synchronous with 15.6 orbits each day and a repeat cycle of 7 days. For the duration of the observing period in this study there is a close overpass to the site of the Singapore rawinsonde station between 22:55 and 23:00 UTC every Wednesday.)

We agree and we have changed the text to match the reviewer's suggestion.

Consider offering the reader a reference for more detailed information about the Aeolus data products and processing (e.g., Reitebuch et al., 2018 and Tan et al. 2008).

We have added the suggested references to the manuscript, and have also included the L2B Algorithm Theoretical Basis Document (Rennie et al., 2020) as well for the reader's reference.

Are there important changes between Baseline 11 and Baseline 14 that the reader should know about?

45 It is not expected that there will be any major differences in these baselines that would affect our results here. However, we have added details of the Baseline 13 parameterisation which particularly improves Rayleigh cloudy winds, and the Baseline 14 correction which addresses a seasonal ascending/descending node bias. We only mention these details for information for the reader, and because we refer to both improvements later in the manuscript. We think it is
50 important to mention that different processing baselines are available for the Aeolus dataset, and that future users should note which version they are using.

What are random error sources that could affect the data quality during the analyzed period?

The main sources of random error that could affect the Aeolus data quality are (a) signal photon shot noise, which arises from fluctuations in the number of arriving photons from the expected
55 average and (b) unwanted electronic noise, which is caused by other electronic components in the detection chain of the instrument. Both act to decrease the signal-to-noise ratio (SNR) and/or broaden the spectral width of the backscattered signal. Wind errors for single-point Aeolus measurements are found to be typically around 5 to 9 ms⁻¹ in this study and in previous literature (e.g. Rennie et al. [2021], Martin et al. [2021] and Lux et al. [2022]), although note that aggregated
60 wind measurements from multiple points can be much more accurate. This information regarding sources of random error has been added to the manuscript.

Please add a sentence as to why only Rayleigh winds are considered. In the upper troposphere in the tropics, I would expect good quality Mie winds at cloud top level or within thin clouds.

We have added a sentence to explain that this was done to simplify the processing of the analysis
65 at the start of the study, and we have mentioned that including Mie winds in future studies could be useful for the reasons the reviewer has given.

2.3. ERA5

Consider including the formula how to calculate ERA5 HLOS wind.

We have added the formula as suggested by the reviewer.

70 *Please provide some information on the quality of the ERA5 data in the upper troposphere/lower stratosphere. NWP models typically have large uncertainties at these altitudes that could potentially affect the results(?).*

We have added a paragraph providing information on some of the relevant biases in ERA5
75 which may affect its representation of the QBO and/or comparisons made in the upper troposphere/lower stratosphere, and have justified the usage of ERA5 for our study. We have also quantified some of the uncertainty in reanalysis winds at these altitudes, citing Kawatani et al. [2016] which discusses uncertainties in tropical zonal winds in different atmospheric reanalyses (excluding ERA5), and citing Healy et al. [2020], which shows root-mean-square zonal wind differences between GNSS-RO and ERA5 of up to 5 ms⁻¹ at 30 hPa.

80 3. Results

3.1. Aeolus observations of the QBO disruption

Line 168: repeating information (see Line 129)

Yes, we have removed this sentence and modified the sentence at Line 129.

3.2. Validation against reanalysis and radiosondes

85 *Line 208: I'm not totally convinced, that ERA5 stratospheric wind errors are that small (due to large model errors at these levels). E.g., the warm bias is mentioned later in the text, what makes this statement slightly doubtful.*

We agree with the reviewer that ERA5 stratospheric wind errors may be comparable to the HLOS error from Aeolus. We have corrected the text to clarify this, and we also include additional refer-
90 ences to Kawatani et al. [2016] and Healy et al. [2020]. We suggest that much of the tropopause bias seen in Fig. 4b is a result of biases in ERA5, rather than solely being a consequence of biases in Aeolus. This is particularly likely since a bias is also seen between ERA5 and the radiosonde, as reviewer #3 points out.

Line 235/236: Is there any conjecture about this height dependence of the Aeolus bias?

95 We have expanded our discussion of the height dependence of the Aeolus bias to give two citations of validation papers which show a similar morphology to the bias structure found in our study. We also suggest that this height dependence is related to the vertical wind shear which maximises at the same altitude, with the following addition: "Given the high vertical wind shear at these altitudes, as noted by Houchi et al. [2010], it is likely that the apparent dipole
100 in this bias is related to local wind shear effects. On the one hand, Aeolus, through its long horizontal accumulation of measurements which form each wind profile, has the capacity to capture localised wind shear that the radiosonde might miss. On the other hand, radiosondes provide a better vertical resolution and can accurately capture regions of high wind shear, but only at the location where the radiosonde is taking measurements."

105 **3.3 Equatorial waves during the QBO disruption**

Please add a reference about the wave filtering method that is used here.

We have added the following references to justify using the Gaussian band-pass filter in this study to highlight the Kelvin waves: Holton [1973], Polo et al. [2008], Roundy and Kiladis [2006], Blaauw and Žagar [2018].

110 *Figure 6: What are the thick horizontal gray lines (early January 2020 and around March 2020)? I can't find a description in the caption or the text.*

The gray lines are data gaps. The figure caption has been updated to add this information.

Introduction/Discussion

115 *Regarding the current state of Aeolus research consider adding the reference Martin et al., 2023 (<https://doi.org/10.5194/wcd-4-249-2023>) about the impact of assimilating Aeolus observations on QBO (with and without QBO RBS) and ENSO in 2020 in the Introduction or Discussion*

Yes, this is a relevant reference, particularly since it discusses the QBO 2020 RBS. We have added it to the discussion.

Technical stuff

120 *I believe the EGU guidelines say that there should be a space in between for the unit ms^{-1} .*

Thanks, this has been fixed in the text.

Sometimes you use 2019/2020, sometimes 2019-2020 or 2019-20. Please make this consistent.

We have changed all instances to 2019/2020 for consistency.

Line 83: please change 2,000 m → 2.000 m

125 This value is meant to be 2000 m rather than 2 m, so we have left this unchanged.

Line 84: please write out the abbreviation UTLS once

We have added this to the text.

Line 105: please write out the abbreviation ECMWF once

We have added this to the text.

130 Line 155: please write out the abbreviation GNSS-RO once

We have added this to the text.

Line 272: please write out the abbreviation OLR once

We have added this to the text.

Line 358: abbreviation DWL is already defined in Line 174

135 Yes, we have removed this from the text.

B Reviewer 2

140 *In their manuscript “Aeolus wind lidar observations of the 2019/2020 Quasi-Biennial Oscillation disruption with comparison to radiosondes and reanalysis”, the authors observe the evolution of the second ever Quasi-Biennial Oscillation (QBO) disruption in novel Aeolus satellite measurements. The observations are validated with radiosonde and reanalysis data. In addition, the reanalysis and satellite data are analyzed with respect to Kelvin wave occurrence before the disruption.*

145 *This study is an excellent show case for the use of novel Aeolus wind measurements for the investigation of dynamic phenomena in the upper troposphere / lower stratosphere (UTLS) region. It clearly demonstrates the capabilities and restrictions of this unprecedented dataset for atmospheric research. With this the article is in general of high scientific significance. Nevertheless, I see a major need for improvement with respect to the scope of the journal and the scientific questions answered by the manuscript.*

150 *In its current state the paper is more a validation of the new Aeolus dataset and with this more in the scope of Atmospheric Measurement Techniques. However, by applying some changes and extending the discussion the paper could help to address important scientific aspects of QBO research. This would make it very suitable to Atmospheric Chemistry and Physics.*

155 *In detail, the paper would strongly benefit from a more detailed discussion on what can be learnt from the differences between ERA5 and Aeolus for the generation of the disruption. Which processes are different (e.g. Kelvin wave activity) and what does this tell us about our current understanding of the physical processes behind?*

We thank the reviewer for this overview and for their recommendations.

Specific comments

L27: Any reference to support this sentence? Maybe Smith et al. (2022)?

160 Yes, we have added Smith et al. (2022) which supports this claim.

L29: Also, here a reference would provide additional information to the reader, e.g. ESA (2020a)

We have added the suggested reference to the paper.

L36ff: This sentence is really long, contains a lot of information and is hard to read. Maybe better split into two or more sentences.

165 Agreed. We have split this sentence into two sentences.

L39: What is your study adding to the current knowledge / understanding of the QBO? What are the research questions you are trying to answer? For the reader it usually helps to briefly address the “why exactly this research” in the introduction to easier follow the manuscript.

Our study contains the first direct wind measurements of the QBO using a space-borne DWL

170 instrument. The additional scientific analyses suggested by this reviewer have helped to shed light on why forecast models may have struggled to predict the disruption, which we have implemented by finding key differences in the representation of Kelvin waves between Aeolus and ERA5. We have added some text to the introduction to address the reasons for this research, as suggested.

175 *L46f: Global and local are antonyms, but how does reanalysis fit into the picture? Is a reanalysis perspective really a measurement perspective? I would suggest to reformulate this sentence.*

We have reformulated the sentence as suggested by the reviewer. The reanalysis of course gives a global perspective, so we've removed the inference that it is neither global nor local.

180 *Section 2.1: Maybe it is worth mentioning that the Singapore Radiosonde Station is commonly used as the gold standard when it comes to QBO analysis as it provides the longest available data record in the tropical stratosphere. Okay, this is somewhere later in the paper. Maybe you could move it forward to the dataset description.*

Agreed, we have moved this sentence into the dataset description as suggested by the reviewer.

185 *L66ff: Is Banyard et al. (2021) really an appropriate reference here? Maybe remove this part of the sentence. The references before already support this sentence sufficiently.*

This sentence is a close-quotation from Banyard et al. (2021) since this describes what we intend to say well and we do not wish to rephrase this, so a citation is necessary here to avoid plagiarism.

L71: I would suggest to better reference ESA (2020b) here.

190 We have added the reference ESA (2020b), although we have left the original reference of Legras (2022) in place since we did use this resource to help give us information about the Hunga-Tonga RBS.

195 *L77f: This is in general correct, but the sentence is confusing here. 35° off-nadir (as mentioned in the sentence before) per se gives mainly the vertical wind component. Only because the vertical wind component is much smaller than the horizontal component, the zonal wind can be derived, which you actually mention, but only one sentence afterwards. So, the order of the sentences here confused me. Maybe just remove this sentence on the advantage for QBO observation?*

Yes. We have rephrased this paragraph to avoid confusion.

200 *L108: ERA5 is available at a temporal resolution of 1 hour:
<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>*

We used 3 hour resolution data in order to reduce unnecessary computational load. In practice such a change makes only an extremely small difference to our dataset and does not affect our results significantly.

205 *Figure 1: How did you fill the data gaps in this plot? E.g. between the weekly QBO settings or during instrument down times (e.g. spring 2021).*

210 A 7-day boxcar filter is used to fill the gaps above 20 km following the QBO 2020 RBS change, and where there are data gaps elsewhere, these are filled using a broader 20-day boxcar filter. We have added this text to the figure caption. As this zonal-mean timeseries requires less stringent constraints for accuracy, we have been able to apply the intended restrictions for NWP centers with some flexibility.

L116f: This would probably be better visible if the Aeolus dataset would be extended until launch in 2018.

215 We have only used data from after the switch from laser FM-A to FM-B occurred in June 2019, since at the time this analysis was conducted, data from the first FM-A data period was not trustworthy. Since then however, data from this period has been reprocessed to remove the existing biases, so if this analysis was done in future we would indeed extend back until launch

in 2018 as the reviewer suggests.

L161: Why only for reanalysis?

We have added 'and NWP models' to the text.

220 *L185: ..., as *mentioned / described* on Kawatani et al. (2016), ...*

We have changed the text as suggested.

L185ff: I had to read these sentences a couple of times, before understanding their meaning. Maybe rephrase?

225 Agreed, we have rephrased this paragraph to make it clearer, and moved the sentence about the importance of using the Singapore radiosonde into the Data and Methods section.

L193: Why 3D?

Removed to avoid confusion.

230 *L204f: This is a courageous statement. Weiler et al. (2021) and Abdalla et al. (2020) only look into this on global average. According to ESA (2021), there could be higher regional biases. In addition, there are small bias differences between ascending and descending orbits in certain months (also ESA, 2021).*

We agree with the reviewer and have added a distinction between global and regional biases in the text. We have also modified the text to change "kept to a minimum" to "kept low".

235 *L206: I fully agree to use negative HLOS in figure 3, but please also state this in the figure or caption.*

We have clarified this in the figure caption.

240 *L211ff: You nicely describe the evolution of easterly winds in the profiles. The weakening of westerly winds is also nicely visible in profiles b-e. Maybe you could also guide the reader through this point first, before coming to the evolution of the easterlies. Otherwise, why do you show the profiles b-e?*

Yes, we agree that this is important. We have added a sentence to guide the reader through the weakening of westerly winds which occurs before the disruption easterlies develop, so that the full context of the disruption itself is provided.

245 *L219: Aeolus measurements are performed separately for each range bin and, unlike the derivation of temperature and aerosol backscatter and extinction, the derivation of wind from lidar measurements does not rely on an iterative profile reconstruction. Thus, cloud contamination in higher atmospheric bins cannot originate from clouds in lower atmospheric bins. To me it seems more that your limit of 250km is quite a large range (in validation studies often 100km is used as collocation criterion) and the atmosphere varies within this distance. There are many points in your profile comparisons where two or more Aeolus winds in the same range gate at the same time are quite far from each other (not only in figure 3g & 3h). This is probably due to the variability of the wind in latitudinal direction. Maybe the altitude offset in figure 3f and 3i are also due to a not perfect match in location.*

255 Firstly, we concur with the reviewer's assessment of the Aeolus measurement profile retrieval. Regarding the comment on cloud contamination issues, we speculate that the higher random error in the higher atmospheric bins might result from cloud contamination at those altitudes, rather than from the clouds at lower altitudes, which were removed through quality control. Aeolus' along-track accumulation spans approximately 87 km, and data may still be flagged as clear-sky even if some clouds are present. Considering the presence of convection at lower levels, isolated deep convection might also occur along this 87 km track, albeit insufficient to flag high-altitude data as cloud-contaminated.

260 Secondly, we agree that 250 km is quite a large range so we have redone the analysis for 150 km collocation. We found that this is a good trade off between sufficient data for our analysis

265 and close collocation (The average minimum proximity of the Aeolus overpass to the radiosonde station is ~ 80 km). A similar radiosonde comparison by Ern et al. (2023), albeit implemented for a slightly different purpose, imposes collocation criteria of $\pm 2^\circ$ latitude and $\pm 10^\circ$ longitude which, although a larger area, should still provide a good comparison of the larger-scale dynamics. Interestingly we find that the altitude offset in figure 3f and 3i is still present in the 150 km collocation results, which suggests it may not be due to latitudinal variability. In addition, the
270 150 km collocation shows, if anything, a slightly weaker similarity to the radiosonde, especially in figure 3e, likely due to the reduced number of profile points used. Nevertheless, Fig. 3 and 4 have now been updated to constrain data to a maximum of 150 km from the Singapore radiosonde launch site.

275 *Figure 4: Which collocation method is used for the comparison of these data points? Aeolus vs radiosonde, probably the same 200km, but what did you do for the comparison to ERA5? Interpolate the model onto the Aeolus and radiosonde location? This might explain the better agreement between ERA5 and Aeolus vs radiosonde and Aeolus.*

280 We have used the same collocation method as for figure 3, now updated to contain all collocations less than 150 km. The profiles in figure 3 are a selected subset of all the profiles used for figure 4. The ERA5 data has been projected on the HLOS measurements as described in the Data and Methods section.

L251ff: Maybe a good idea for a next study. ECMWF forecast / analysis data with and without Aeolus assimilation is available at ECMWF.

285 We thank the reviewer for their good suggestion, this is indeed an idea for further study which the authors would be interested in carrying out.

L255: What exactly do you mean with like-for-like comparison? Are you interpolating ERA5 data to Aeolus measurement locations?

290 Yes, that's precisely what we do. Although other studies may compare data from Aeolus against ERA5 on different grids, since we're focusing on the perspective of Aeolus we have interpolated ERA5 onto each Aeolus measurement location. We have altered the text in the Data and Methods section to clarify this, and added the formula used to convert ERA5 u and v into a synthetic ERA5 HLOS wind. In the text here, we have removed the phrase 'like-for-like' to avoid confusion, and have referred the reader to the ERA5 subsection for clarity.

Figure 5: Maybe you could show the ERA5 contour lines in the whole plot.

295 We can only show ERA5 contours over the same area as the Aeolus measurements since we are effectively making a comparison between (a) Aeolus measuring the real atmosphere and (b) Aeolus measuring the 'ERA5 atmosphere'. This means that the 'ERA5' dataset has the same spatial and temporal extent as the Aeolus dataset.

300 *L257: You not only see this difference in onset time, you also clearly see stronger winds and wind gradients in the Aeolus data in the troposphere before the disruption (e.g. -5 m/s line is at a higher altitude in Aeolus data). These stronger wind gradients might have an impact on the Kelvin wave propagation (you discuss afterwards), so I think they should be mentioned here.*

305 We thank the reviewer for this useful comment. It has partially motivated some of the further analysis added later in the paper. We agree, and the updated version of figure 5 shows this more clearly. We have added more information about the relationship between these stronger wind gradients and Kelvin wave propagation to the text.

L270: ... highlight only symmetric wind structures ... (antisymmetric waves with respect to the equator are removed due to your averaging from -5° to $+5^\circ$ latitude; for a detailed analysis of equatorial waves in the Aeolus dataset, you could have a look at Ern et al. 2023)

310 That is correct. We have added this clarification.

L291 – 296: This paragraph mixes different things and draws conclusion which I either do not

understand correctly or a very far-fetched. I would suggest to remove the whole paragraph.

315 We have removed most of this paragraph, and only retained the sentence relating to the comparison between the different QBO disruptions, since we believe that this is particularly relevant to this study.

animation S2: Title of plot and caption are not in line with each other. It should be $\pm 5^\circ$ latitude as everywhere in the manuscript.

320 We thank the reviewer for pointing this out. The Hovmoller and cross-section plots are plotted from the equatorial bin of a 5° latitude resolution gridded dataset, so these should both be ± 2.5 degrees. A gridded dataset is used to make the analysis easier to run. The zonal-mean time series of the QBO for figure 1 and 5 is for $\pm 5^\circ$ since this Aeolus data has not been binned in the same way; in order to maximise the SNR and accuracy as much as possible. Re-running code to ensure these match would take a long time and is unlikely to change the timeseries figures significantly.

325 *Animation S2 and Figure 7: Why don't you apply the same temporal filtering as in the Hovmöller plots? You want to show the Kelvin waves, but these are barely visible due to the dominating feature of the Walker circulation. By applying a similar filter as in Figure 6, this strong dipole should vanish and the Kelvin waves should become clearly visible.*

330 We agree with the reviewer here and have updated animation S2 and figure 7 to apply the same temporal filtering as in the Hovmöller plots as suggested. The result of this is that the Walker circulation no longer dominates and the Kelvin waves have become more clearly visible.

L328ff: You could perhaps reformulate the sentence to stress the importance of a future wind lidar measuring up to at least 30km or higher for QBO research.

We have added this to the text and given two references in support of this.

335 *L336: ... this change in random error ... (the high random error itself is a problem especially for short analysis periods and perturbation analysis, for these a bias would be less of a problem)*

We have edited the text to add this, although this paragraph has now been moved to the Data and Methods section.

L348: Why only reanalysis model?

340 Yes, this is the case for NWP models as well. We have added this to the text.

L349: Why is your analysis not spanning the whole Aeolus measurement period, so why is the data before summer 2019 missing?

345 As mentioned earlier, we have only used data from after the switch from laser FM-A to FM-B occurred in June 2019, since at the time this analysis was conducted, data from the first FM-A data period was not trustworthy. Additionally, our analysis does not span the whole Aeolus measurement period for a very important reason, which is that we are focused primarily on the QBO disruption of 2019/2020. It is important that the focus of the paper does not shift away from the disruption (and therefore relevant for publication in ACP) towards a more general techniques paper on the characteristics of Aeolus measurements in the tropical UTLS (which would be more relevant in AMT). Please see comments by reviewer #3 for more details.

350 *Discussion: This is not a general discussion of the results of the study. What you describe here are drawbacks of the Aeolus mission. Thus, I would suggest to either rename the section or revise its content.*

355 We agree with the reviewer that there is an overemphasis on the drawbacks of the Aeolus mission here, and so we have revised the contents of the discussion section accordingly. We think it is important to still mention key limitations of the study which may affect the results, but some of the content has been moved to the Data and Methods section, partially in response to a comment by reviewer #1 as well.

L379f: Why is this in agreement with the data validation?

360 We agree that this was confusing, and since we discuss the data validation in the previous paragraph, we have removed this clause for clarity.

L385: Maybe better: ... have been discussed.

We have changed this as suggested.

L389: Maybe good to stress here the importance of measurements up to at least 30km.

365 We have added this to the Discussion in response to an earlier comment by this reviewer.

370 *Conclusions: You nicely describe the data, but what have we learned from a scientific point of view (except that Aeolus is a well-suited dataset for observing wind related phenomena in the UTLs)? Why did it come to this QBO disruption? Where is the difference between ERA5 and Aeolus and what can we learn from this difference for our understanding of the underlying physical processes? These are questions I would expect to be answered in an ACP manuscript.*

375 We thank the reviewer for this comment, which has motivated much of the additional scientific analysis we have conducted. We have learned that there are key differences in the vertical wind shear and Kelvin wave variances between ERA5 and Aeolus which help to explain the observed lag in disruption onset in the reanalysis. We have therefore made the manuscript more suitable for publication in ACP due to the greater focus on the dynamics of the QBO disruption.

C Reviewer 3

380 *The manuscript by Banyard et al. presents wind measurements of the 2019-2020 QBO disruption from the first spaceborne Doppler wind lidar ADM-Aeolus. The general topic is relevant for publication in ACP and the paper is interesting, demonstrating that the QBO is generally well captured in Aeolus data. However, I find the manuscript very focused. It would benefit from including a broader analysis and discussion of the new information brought by Aeolus on tropical lower stratospheric dynamics compared to a state-of-the-art reanalysis. Some differences between the ERA5 and Aeolus are already mentioned but not much commented.*

385 *For these reasons, major revisions (additions) are required before the paper can be considered for publication. My recommendation to the authors would be to make more quantitative statements and expand the comparison between Aeolus and ERA5 to other modes of variability than the QBO (e.g., equatorial waves) which are only alluded to in the present manuscript. The authors should also consider the recent preprint by Ern et al. (ACPD) on a related topic.*

390 We thank the reviewer for their comments, and in response we have made significant changes to make the paper more quantitative and have expanded the comparisons, as described in the Overview at the top. We have considered and referenced the recent preprint by Ern et al. (ACPD) where appropriate.

Main comments

395 *1) Lack of significance test : in a few instances the author claim that a bias/difference is significant, but do not include a statistical significance test. For instance the near-tropopause wind bias at Singapore (I238-240) is not clear to me, since it is also present in the ERA5-Singapore radiosonde comparison.*

400 The authors agree that a statistical test is useful to quantify, for example, the near-tropopause wind bias at Singapore (L238-240), and we have included a Student's t-test of $p < 0.001$ to highlight altitudes where the bias is statistically significant. A t-test is suitable here since the distributions are near-normal (as shown in the violin plots). As mentioned in the response to reviewer #1's comment at L208, we suggest that much of the tropopause bias seen in Fig. 4b is a result of biases in ERA5, rather than solely being a consequence of biases in Aeolus. This is particularly

likely since a bias is also seen between ERA5 and the radiosonde, as reviewer #3 mentions here.

405 2) The authors touch the topic of equatorial wave representation in ERA5, but they could make more quantitative statements. They might also wish to change the colormap used for Figure 6. Given that one of the advantages of Aeolus is global sampling, the authors could without much effort quantify the differences as a function of longitude, beyond the location of Singapore. This is of interest, in particular since earlier studies (e.g., Baker et al., 2013 ; Podglajen et al., 410 2014) found that reanalysis uncertainties and wind errors in the tropical UTLS were larger in regions without assimilated radiosonde observations, resulting in a pronounced zonal structure. Note that the need for wind observations to better constrain the flow in tropical regions is what motivated Aeolus in the first place.

415 The authors agree that the topic of equatorial wave representation should be explored further, especially in the context of the disruption and with more quantitative statements. To this end, a new section has been added examining the differences in equatorial wave spectra between Aeolus and ERA5 and differences in Kelvin wave equivalent depths with height. We have considered the reviewer's suggestion of further analysis of the differences between Aeolus and ERA5 as a function of longitude; however the recent preprint by Ern et al. (2023, ACPD) covers this longitudinal variation extensively and we have thus taken a different path, whereby we focus on the 420 QBO disruption and Aeolus' capability to capture it and the QBO. The colormap for the middle panel of figure 6 has been changed to make the structure clearer.

3) Regarding the last sentence of the abstract : "This analysis highlights how Aeolus and future Doppler wind lidar satellites can deepen our understanding of the QBO, its disruptions, and the tropical upper-troposphere lower-stratosphere region more generally." , it is not shown in the 425 paper that Aeolus can help clarify the mechanisms of the disruption, at least directly. I imagine spaceborne lidars would help deepen the understanding of the QBO by putting observational constraints on equatorial waves and tropical gravity waves, but this is not really the focus of the paper.

430 The authors have made changes to the paper to make it more suitable for publication in ACP by exploring the role of equatorial waves in the development of the disruption to a greater depth. This better justifies this sentence in the abstract compared with previously, so we have kept it as in the original manuscript.

Line 1 : The abstract could mention some key number (bias, etc.)

435 We have added this information to the abstract, along with the key results from the new material.

Line 102 : please describe the method briefly here

We have moved these details a little earlier in the section and modified the text so that it describes the correct method.

440 Line 238-240 : Is this a significant bias ? It does seem smaller than the difference between Aeolus and the radiosonde (panel a), which by the way, also maximizes near the tropopause.

445 As we have previously mentioned, we have performed a Student's t-test on these results to demonstrate that the bias is statistically significant. This is the case both for the difference between Aeolus and the radiosonde (panel a) and the differences between the observing instruments and reanalysis (panel b and c). Though significant, this does not give any direct indication as to the cause of the biases in relation to each dataset, i.e. it is not possible to say that one bias is the addition of the other two since the distributions are not additive.

Line 264-265 it is suggested that the sampling by aeolus does not induce much bias. You could easily prove this with a figure comparing ERA5 with aeolus sampling and the actual zonal mean

450 The biases in ERA5 that are caused by regional differences in observation sample sizes are discussed by Kawatani et al. [2016] and are mentioned earlier in our study. Given that Aeolus' data sampling is mostly homogeneous (aside from data gaps due to instrument downtime or

quality controls), we believe it is fair to make this assumption. We have not added the suggested figure since we do not think it adds much strength to the paper. Mapping the locations and density of input observations for ERA5 would take quite a bit of work to complete, and is outside the scope of this study.

Fig 4 : the pressure scale is wrong in this figure. What do the dots correspond to? Some of the numbers in the legend would find their place in the main text.

The pressure scale has been corrected and the dots have now been referenced in the figure caption. We are not sure which legend the reviewer is referring to, but we have placed all relevant numbers off the figure in the caption. Median differences and standard deviations have been updated to correspond to the new 150 km constraint.

Fig 5 : Could you rather plot the mean wind in contour and the difference in colors ? This would be more quantitative. There seems to be a delay, does it hold for the later period or is this just a feature of the disruption?

The authors agree with the suggestion by the reviewer and have changed the plot so that it shows the mean wind for Aeolus and ERA5 in contour and the difference between them in colour. In general, Aeolus has stronger winds than ERA5, but the zero wind line does not always follow the line of zero difference.

Fig. 6 : Is there a zonal-mean structure in the error (see main comment 2) ? Also, I do not understand the need for time filtering. Eastward propagation appears clearly in your unfiltered plot.

We disagree with the reviewer that time filtering is not necessary to more clearly show the eastward propagation of waves, since in the unfiltered plot the standing wave of the Walker Circulation dominates and masks the Kelvin waves. (There is also a contradictory comment from reviewer #2)

lines 271-277 : The phrasing is a bit odd, suggesting that this long period was not an optimal choice for Kelvin waves. You cite a few papers which show that the typical stratospheric Kelvin waves commonly seen in Hovmoller diagrams indeed have planetary wavenumber 1-2, periods of 10-20 days or more, as first described by Wallace and Kousky and reported in many papers and textbooks. Convectively-coupled Kelvin waves in the OLR are higher frequency but this is not what dominates at tropopause altitude where free-travelling waves are prominentl. I would shorten this discussion.

The reviewer makes an important point, and we have rephrased this discussion accordingly. The distinction between convectively-coupled Kelvin waves in the OLR (and therefore covering the large altitude range of 10-18 km as mentioned by Alexander et al. [2008]), travelling with periods of 7-10 days, and free-travelling waves in the TTL and stratosphere, with a larger range of periods, is an important one to make. We have tried to keep the discussion concise whilst still mentioning the details we think are relevant. Our choice of filter is optimal for the dominant wave periods seen in the Aeolus data, and is broad enough to capture the vast majority of the Kelvin wave spectrum at these altitudes.

References

S. P. Alexander, T. Tsuda, Y. Kawatani, and M. Takahashi. Global distribution of atmospheric waves in the equatorial upper troposphere and lower stratosphere: Cosmic observations of wave mean flow interactions. *Journal of Geophysical Research: Atmospheres*, 113(D24), 2008. doi: 10.1029/2008JD010039.

M. Blaauw and N. Žagar. Multivariate analysis of kelvin wave seasonal variability in ecmwf I91 analyses. *Atmospheric Chemistry and Physics*, 18(11):8313–8330, 2018. doi: 10.5194/acp-18-8313-2018.

- 500 S. B. Healy, I. Polichtchouk, and A. Horányi. Monthly and zonally averaged zonal wind information in the equatorial stratosphere provided by gnss radio occultation. *Quarterly Journal of the Royal Meteorological Society*, 146(732):3612–3621, 2020. doi: 10.1002/qj.3870.
- J. R. Holton. On the frequency distribution of atmospheric kelvin waves. *Journal of Atmospheric Sciences*, 30(3):499–501, 1973. doi: 10.1175/1520-0469(1973)030<0499:OTFDOA>2.0.CO;2.
- 505 K. Houchi, A. Stoffelen, G. J. Marseille, and J. De Kloe. Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and the ecmwf model. *Journal of Geophysical Research: Atmospheres*, 115(D22), 2010. doi: 10.1029/2009JD013196.
- Y. Kawatani, K. Hamilton, K. Miyazaki, M. Fujiwara, and J.A. Anstey. Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses. *Atmospheric Chemistry and Physics*, 16(11):6681–6699, 2016. doi: 10.5194/acp-16-6681-2016.
- 510 O. Lux, B. Witschas, A. Geiß, C. Lemmerz, F. Weiler, U. Marksteiner, S. Rahm, A. Schäfler, and O. Reitebuch. Quality control and error assessment of the aeolus l2b wind results from the joint aeolus tropical atlantic campaign. *Atmospheric Measurement Techniques*, 15(21):6467–6488, 2022. doi: 10.5194/amt-15-6467-2022.
- 515 A. Martin, M. Weissmann, O. Reitebuch, M. Rennie, A. Geiß, and A. Cress. Validation of aeolus winds using radiosonde observations and numerical weather prediction model equivalents. *Atmospheric Measurement Techniques*, 14(3):2167–2183, 2021. doi: 10.5194/amt-14-2167-2021.
- I. Polo, A. Lazar, B. Rodriguez-Fonseca, and S. Arnault. Oceanic kelvin waves and tropical atlantic intraseasonal variability: 1. kelvin wave characterization. *Journal of Geophysical Research: Oceans*, 113(C7), 2008. doi: 10.1029/2007JC004495.
- 520 M. P. Rennie, L. Isaksen, F. Weiler, J. de Kloe, T. Kanitz, and O. Reitebuch. The impact of aeolus wind retrievals on ecmwf global weather forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147(740):3555–3586, 2021. doi: 10.1002/qj.4142.
- 525 P. E. Roundy and G. N. Kiladis. Observed relationships between oceanic kelvin waves and atmospheric forcing. *Journal of Climate*, 19(20):5253–5272, 2006. doi: 10.1175/JCLI3893.1.