Reply to Reviewer #3 :

We thank Reviewer #3 for their detailed and constructive feedback. We acknowledge the concerns regarding methodological justification, clarity, and the structure of the manuscript. The comments have been significant guidelines in providing a substantial revision of the paper. We have streamlined the methods section, improved the figures, and rewritten large parts of the results and discussion to be more concise and rigorous.

The premise of this study is very promising and the results, if robust, are of high significance in comparing Aeolus, ERA5 and GNSS-RO gravity wave energy parameters in the tropical UTLS.

Still, the more I go through the manuscript I find too many details in the methodology unjustified, or their interpretation too stretched. Almost half of the text belongs to data and methods section which should be streamlined a lot. The text overall lacks an organized and concise structure, and some method details or datasets (e.g. NCEP reanalysis or the OLR datasets) seem to appear out of the blue.

I have several major comments about methodology that need to be clarified, because some of the results do not look very robust to me from the beginning, and this cascades then to the rest.

Figures could be improved a lot, and the authors should make a big effort in the text to avoid repetitive sentences, unnecessarily long explanations / verbose in methods or results (a lot of examples in minor/technical comments).

Also I feel that in many instances things are presented in a rather bombastic way, e.g. without really specifying where and how these valuable results have applications.

In section 5-6, some of the conclusions might change if some small tweaks in methodology were applied -- the authors make many choices and assumptions in the method -- and many grand statements with what comes out of it. Unless one shows very convincing and robust results (which would require a fair amount of supplement material), in the plots provided in this manuscript I see some inconsistencies that make me remain a bit skeptical.

To be clear, I'd very much like to see this study on such relevant topic published, and I hope the large amount of comments I assembled below are helpful for this. I recommend a major revision, and at least an additional round of reviews will be needed after that since the required changes are very substantial.

#

#

Major comments

#

#

#

1: vertical grid and filtering choice

#

-l.124-126: this grid penalizes ERA5 and RO a lot more than Aeolus, and 'acceptable middle ground' does not really justify your choice in my opinion. Is there any other literature doing this kind of middle-ground approach with other datasets?

-There are undesirable sources of uncertainty if you sub-sample or interpolate onto your 0.5km vertical grid: this might affect the resulting profile if a wave is not well aligned with your 0.5km vertical grid. Also ERA5, Aeolus and RO have each a very different vertical (original) grid alignment with your 0.5km grid.

In my opinion one should err on the side of caution and interpolate to a finer grid that retains all dataset's vertical structures as much as possible, and then filter out the scales that the coarser dataset cannot see, I explain below:

-In section 2.2 you specify that you apply vertical high-pass filter to the data. Why not use a finer vertical grid of e.g. 0.1km, and apply bandpass between e.g. 1km and 9km? This way the uncertainty with sub-sampling is gone, and you remove the shortest vertical scales that Aeolus cannot see to even the field among all datasets. To me, this would be the fairest way to make the comparison by taking the vertical scales resolved by all datasets.

We thank the reviewer for this excellent suggestion regarding our data processing. We have completely revised our data processing pipeline as suggested. All datasets (Aeolus, GNSS-RO, and ERA5) are now interpolated onto a finer 0.1 km vertical grid. Following this, we apply a band-pass filter to retain vertical wavelengths between 1.5 km and 9 km. This new methodology ensures that we preserve the native vertical structures of each dataset as much as possible, while the filtering guarantees that our comparison is limited to the wave scales reliably captured by all instruments. All figures and results in the manuscript have been regenerated using this new, more rigorous method. The relevant part of the Methods section has been rewritten to detail and justify this new approach.

(lines 126-129)

This study specifically utilizes Aeolus Level 2B Rayleigh clear HLOS winds, ERA5 wind components, and GNSS-RO temperature profiles, all brought to a standard interpolated grid to facilitate the accurate comparison and integration of data from the different sources. The chosen grid has a vertical resolution of 100 meters and spans a range from 0 to 30 km altitude. This approach preserves the maximum vertical detail from each dataset before analysis.

(lines 206-211)

After said windowing, a band-pass filter designed to retain vertical wavelengths between 1.5 km and 9 km. is applied to the perturbation profile, as seen in Fig.1b and 1c. The upper limit of 9 km isolates GWs from larger-scale planetary waves, consistent with our background removal strategy. The lower limit of 1.5 km is chosen to reflect the effective vertical

resolution of the Aeolus instrument (Ratynski et al., 2023) and ensures that our comparison is restricted to wave scales reliably resolved by all datasets (Banyard et al., 2021). This procedure provides a methodologically consistent basis for comparing GW energy across the different instruments.

#

2: NCEP reanalysis and smoothing (l.198-204)

#

A lot of things appearing out of the blue here without proper justification.

--> Is this the NCEP-DOE Reanalysis 2? It is not referenced either. How come this dataset is not mentioned in section 2.1?

--> Just because it's easier to integrate does not justify using it. I just don't understand why ERA5 is not used with its own tropopause.

--> Also, you don't show anywhere how similar are the results compared to ERA5. It certainly has poorer vertical resolution than ERA5, and this choice just adds an unnecessary layer of uncertainty. Not even some comparison material in a supplement?

"The profile is then smoothed using a 14-point moving average over the 49-point profile"

--> No justification given anywhere for this. Any other studies doing similar things that you could reference here?

--> You should explain what the purpose of this smoothing is. My impression is that it's not even necessary (see last part of my Major Comment #1 for a better option to compare what's resolved by all datasets).

We thank the reviewer for pointing out these methodological weaknesses. We now use the tropopause height derived directly from the ERA5 dataset for all analyses to ensure consistency. The use of NCEP reanalysis has been removed from the manuscript. The 14-point moving average smoothing step has been removed. We agree it was not well-justified. The new band-pass filtering approach (as suggested in Major Comment #1) is a more appropriate and sufficient method for conditioning the perturbation profiles. The Methods section has been updated to reflect these important changes.

(lines 225-227)

The lower bound is set one kilometer below the tropopause to focus on events extending beyond it, balancing Aeolus' resolution with our interest in upper-end dynamics. For consistency, the tropopause height is derived directly from the ERA5 dataset for all analyses. The profile is then averaged over the selected range, representing the Ek, as seen in Fig.1c.

#

3: treatment of GNSS-RO data, details not properly justified

#

l.206-211: details are very vague, e.g. which windowing is used for both in the end? Please state clearly what settings are applied to RO data and Aeolus.

l.213-214:

"Where the Brunt-Vaisala frequency squared (N2) is smoothed using binomial (Gaussian) smoothing of 10th order."

--> This is not justified, where is this coming from? Any reference for this?

"Consequently, the data treatment across various instruments, whether wind or temperature remains consistent"

--> I strongly disagree!

We apologize for the lack of clarity and justification in this section. We have clarified in the manuscript that the same Welch windowing function was applied to all perturbation profiles from all datasets. We have removed the 10th-order binomial smoothing of the N^2 profile. We thank the reviewer for this critical comment and agree that this step was not standard practice. As argued in studies like Alexander et al. (2008b), the spatio-temporal averaging used to derive the background state provides a sufficiently smooth and stable background temperature profile for the N^2 calculation. With these changes, our statement about the consistency of data treatment is now properly supported. The section has been revised for clarity and accuracy.

(lines 237-240)

The main difference lies in substituting temperature T(z) for wind U(z) throughout the background-perturbation decomposition [...] The Welch window was applied to all perturbation profiles (wind and temperature) before filtering to mitigate spectral leakage. The same band-pass filtering strategy and vertical averaging then provide the Ep profile from the temperature perturbations.

#

4: Most figures are low quality

#

Fig.2:

--> please use the degree sign ° and not "DegN/E" (also present in Figs.4,5,6,7... all figures with lon or lat dimension...)

--> color scale is not the best for visibility, a color every 0.1 would improve guessing the exact value by eye.

--> It appears a bit pixelated if one zooms in just a bit.

Fig.3

--> unreasonably large to show only four lines

--> way too many labels on the x-axis

Fig.5

--> label sizes too small

--> too many labels on y-axis

--> odd alignment of a)b)c) with subpanel titles

--> b) panel title size mismatch with the others, looks like put there by hand unlike a) and c)

The figures have been reworked to be as high quality as possible, with the according fixes applied to fig 2 and 3 (now in the appendix) as well as the geographical maps and hovmoller plots. We also ensure that degree signs, label sizes, and colorbar intervals follow the specific corrections the reviewer has suggested.

We acknowledge that the apparent pixelation or blurriness described by the reviewer. Such issues arise primarily from the PDF conversion process currently used to assemble the manuscript draft. Despite multiple attempts, it has been difficult to completely avoid compression artifacts when exporting to PDF from Word, especially for complex maps and plots with fine detail. We are deeply sorry for the inconvenience this has caused in the review process. We would like to reassure the reviewer that these issues will not occur in the final submission: for the production-ready version we will provide all figures separately as individual high-resolution vector or high-quality raster files (following Copernicus/EGU guidelines on figure preparation).

#

5: pages 11-14

#

-whole pages 11-12: this can be briefly summarized in the main manuscript and all the details moved to a supplement, including Fig.3

-The noise correction makes a quite long list of assumptions, could the authors provide some results/comparison of Aeolus results without noise correction for reference?

-Fig.4: the stark contrast of MAM 2019 and MAM 2020 does not give a reader a lot of confidence in your method. I am skeptical of how realistic the evolution of the left column is (the noise-corrected AEOLUS HLOS*).

-l.351-352: "The geographical distribution and evolution of energy hotspots are largely similar between the two datasets"

--> I disagree, the evolution of their strength, even in relative terms, seems quite different: e.g. compare the last 3-4 rows.

We acknowledge the concerns about the original results. A primary reason for the "stark contrast" between years was a flaw in our original noise correction. We have developed a much more robust, spatio-temporally adaptive algorithm (detailed in Appendix D). All results and figures have been regenerated with this new method, along with the new vertical grid and filtering. The new results no longer exhibit the unrealistic jumps between seasons. We have moved the full mathematical derivation and validation plots (including the original Figure 3) to a new, comprehensive Appendix D. The main text now contains a concise summary of the method's principles. Appendix D also now includes a plot of the uncorrected Aeolus data for reference, as requested. With the new results, we have completely re-written the interpretation in Section 3.1.

#

#

Minor / technical comments

#

#

Abstract

"revealing opportunities to refine reanalysis products and model parameterizations, as well as improving the energy ratio."

--> too vague, is there any specific recommendation here?

The following paragraph has been modified

The combination of Aeolus and GNSS-RO data allows for an observationally-based examination of the partitioning between kinetic and potential energy, highlighting discrepancies with reanalysis products that could inform future model parameterization development.

-l.2: cite ERA5 reference here

Done

-l.28-30: but the Podglajen study is from before ERA5 was around, please rephrase sentence for consistency.

Done

-l.45-46: "short-wavelength waves are primarily lower frequency gravity waves, as dictated by the dispersion relation"

--> To avoid confusion please specify that it's short vertical wavelength, and give a ballpark number of the range of vert. wavelengths you are referring to.

--> Also in the next sentence, specify what vertical wavelenghts can be captured by Aeolus.

The following paragraph has been updated

These waves with short vertical wavelengths (typically 2-10 km) are primarily lower-frequency gravity waves, as dictated by the dispersion relation, and exhibit relatively large amplitude wind variability. The Aeolus satellite, equipped with its Atmospheric LAser Doppler INstrument (ALADIN), is able to measure global wind profiles up to an altitude of 30 km, providing insights into the behavior of gravity waves with vertical wavelengths down to ~1.5-2 km in these critical atmospheric layers (Banyard et al., 2021; Rennie et al., 2021; Ratynski et al., 2023).

-l.59-61: calling it "climatology" from 3 years sounds a bit stretched...

--> perhaps simply state this as an observational estimate for Jun.2019-Aug.2022

--> also this is an example of a repetitive sentence. E.g. 'and its link with deep convection' could be removed without any loss of information

The sentence has been changed to :

By comparing direct measurements with ERA5 data, we reveal certain limitations in the reanalysis's ability to represent tropical gravity wave dynamics. We will look at the most recent reprocessed Aeolus baseline 2B16, providing data from June 2019 to August 2022

And the other sentence was removed

-l.70: I would support sub-subsections for each separate dataset and methods.

We acknowledge the suggestion to introduce sub-sections for each dataset to improve readability. However, since all datasets are now mapped to ERA5's resolution, and the GNSSRO treatment has been streamlined to match Aeolus, there are effectively only two types of processing. We believe that organizing the methods by Ek/Ep categorisation is preferable, as it avoids unnecessary repetition between ERA5 and its counterparts and keeps the text more concise.

-l.72: range bin settings and other specifications should have an earlier reference.

--> Also please update the Rennie and Isaksen 2020 reference to the 2024 ESA contract report (which includes all information from the 2020 TM). Check throughout the manuscript.

--> https://www.ecmwf.int/en/elibrary/81546-nwp-impact-aeolus-level-2b-winds-ecmwf

The reference for these details is now prominently placed at the end of this introductory block, ensuring that all preceding technical information is immediately and clearly sourced. We have also updated the reference to the latest 2024 ECMWF report by Rennie and Isaksen throughout the manuscript, as requested.

-l.87-88: please confirm whether you got the data on that native resolution?

The reviewer is correct to point out the distinction between the model's native grid and the data product we use. The data products we downloaded from the ECMWF archive were pre-interpolated onto a regular 0.25° x 0.25° latitude-longitude grid. We have revised the manuscript to state this explicitly and avoid any ambiguity.

-l.94: best candidate (by far in my opinion), especially when compared to other reanalysis products.

The sentence has been reworded as “the best candidate”.

-l.95: "standard" --> you mean the 137 hybrid levels? Standard is usually associated with the 37 standard pressure levels, I recommend not using this term here to avoid confusion.

We confirm that ERA5 data was retrieved on native 137 model levels, not "standard" levels. Text corrected

-l.99: GNSS-RO datasets --> please list which missions are included + their references, and I presume COSMIC-2 dominates the overall data amount? If so, mentioning a bit

COSMIC-2/FORMOSAT-7 and other third-party RO missions were evaluated and monitored by ROM SAF (several technical reports exist), but no routinely generated ROM SAF products based on those data were disseminated in the 2019-2022 operations reports. The satellites used in this analysis are the Metop constellation (Metop-B & Metop-C continuously and Metop-A up to 15 Nov 2021). Refences have been added.

(lines 108 -114)

For the study period of June 2019 to August 2022 these datasets are dominated by the Metop constellation: Metop-B and Metop-C throughout, with Metop-A contributing until its retirement in November 2021 (von Engeln et al., 2011). These datasets are derived from the bending angles of GNSS signals as they pass through the Earth's atmosphere and are observed by low Earth-orbiting satellites. It provides global coverage with a high vertical resolution, sub-Kelvin accuracy, full diurnal coverage, and all-weather capability. The vertical resolution of GNSS-RO temperature profiles is fundamentally limited by diffraction and varies with altitude, typically ranging from ~0.5 km in the lower troposphere to ~1.4 km in the middle atmosphere (Kursinski et al., 1997)

-l.111-113: perhaps merge with l.97-98 at the beginning of the paragraph, otherwise to me feels a bit repetitive.

Done

-l.103-104: defined by bending angle gradient, which increases near inversion layers / humidity gradients

--> I recommend to refer to Kursinski et al. 1997 here --> https://doi.org/10.1029/97JD01569

The reference has been added.

-l.115-118: feels very repetitive and could be streamlined

Removed as it added not information

-l.130-131: overselling and too vague, remove or specify recommendations to enhance reanalyses and models from the results of your study.

The sentence has been removed

-l.131-133: just say they are independent datasets, this sentence can be streamlined and toned down.

The sentence has been simplified.

################ 2.2 Methods and limitations

-l.139-150: regarding the trickiness of background state removal, I miss a discussion about research that used GNSS-RO and Aeolus to study Kelvin waves, their vertical scales and (in the case of Randel et al., 2021) the behavior of the small-scale residual.

These references are very relevant to your study's methodology, the more so since they use the same datasets as you.

--> Randel and Wu (2005) --> https://doi.org/10.1029/2004JD005006 (using GPS-RO)

"Vertical wavelengths of ~6–8 km are observed near and above the tropopause in December 2001 to January 2002 (Figures 6a and 6b), while shorter vertical wavelengths (~4–5 km) are observed in May and August–September 2002 (Figures 6c and 6d). "

--> Randel et al. (2021) --> https://doi.org/10.1029/2020JD033969 (using COSMIC-2)

"strong residual variance occurs in the longitudinal shear zones of Kelvin waves" and this small-scale residual T variance is associated with GWs.

--> Zagar et al. (2021) --> https://doi.org/10.1029/2021GL094716 --> "Aeolus assimilation modifies the representation of vertically propagating Kelvin waves in the tropical UTLS" (Aeolus)

A discussion encompassing previous research, the limits of current methods and the excepted caveats has been implemented.

(lines 136 – 166)

The following section discusses the retrieval of GW kinetic energy, Ek. A primary challenge in this retrieval, particularly in the tropical UTLS, is the robust separation of GWs from other dominant, synoptic-to-planetary scale equatorial waves, such as Kelvin waves. Observational studies using GNSS-RO data have consistently shown that Kelvin waves, with typical vertical wavelengths in the range of ~4-8 km (Randel et al., 2021; Randel and Wu, 2005), are a prominent feature of the tropical temperature and wind fields. This presents a potential for spectral overlap with the longer vertical wavelength portion of the GW spectrum that this study aims to capture.

[…]

The separation of the wind or temperature profile into a background state and perturbations using HD is intended to isolate fluctuations characteristic of gravity waves by filtering out larger-scale and slower-evolving processes like the mean components of Rossby and Kelvin waves. This selection relies on the distinct scale and structural characteristics of GW perturbations. However, the work by Randel et al., (2021) using dense COSMIC-2 RO data reveals further complexities. They found that "residual" small-scale temperature variances (analogous to our perturbation fields) exhibit coherent maxima in the longitudinal and vertical shear zones of large-scale Kelvin waves. This suggests that the local atmospheric environment shaped by Kelvin waves, particularly variations in static stability (N^2), can modulate the amplitude of smaller-scale variability, potentially including GWs. Furthermore, the assimilation of Aeolus wind data itself has been shown to directly impact the representation of vertically propagating Kelvin waves in numerical weather prediction models, especially in regions of strong vertical wind shear (Žagar et al., 2021). This implies that Kelvin waves are indeed present in the Aeolus observations and that their characteristics might differ from those in reanalyses not assimilating Aeolus data.

#################

-l.160-161: "Aeolus now provides the necessary tools to apply the same approach for GW Ek."

--> Sorry to be picky here, but what tools does Aeolus bring now that it didn't before. You use the same approach (your tool) to calculate GW Ek from Aeolus (data, not a tool). Such phrasing is just unnecessary verbose.

We reworded the sentence to remove unnecessary verbose speech.

- Fig.1: please include the U(z) notation in the labels, and specify which datasets you take U from.

The figure has been updated.

-l.216-219: you should state all this when introducing Ek{hlos}.

--> Fig.2 belongs in a supplement

The text has been moved up next to the introduction of Ek_hlos and the ex-Fig.2 has been moved to the Appendix.

--> And wouldn't it be fairer to compare EK from ERA5 U to EK_HLOS??

The proposed comparison would involve examining ERA5 zonal kinetic energy (Ek_u_ERA5) against ERA5 HLOS-projected kinetic energy (Ek_HLOS_ERA5). That ratio would assess how well the HLOS projection specifically captures the zonal wind component within the model.

However, our analysis in Appendix B (e.g., Fig. 2) had a different objective. We aimed to quantify how representative a quasi-zonal HLOS measurement (like Aeolus's) is for the total gravity wave kinetic energy, which includes both zonal and meridional components. To do this, we calculated the ratio of ERA5's HLOS-projected kinetic energy (Ek_HLOS_ERA5) to its total kinetic energy (Ek_TOTAL_ERA5, including both u′ and v′ components). This approach was intended to estimate the fraction of total kinetic energy captured by a quasi-zonal line-of-sight measurement and to characterize the spatial and temporal variability of gravity wave anisotropy in the ERA5 model. The ratio explicitly highlights regions where meridional wind perturbations are more significant, therefore pointing where an HLOS-only measurement would miss a larger share of the total energy. We believe this provides a necessary context for interpreting the absolute values measured by Aeolus.

-l.228-231: belongs also in a supplement in my opinion

Moved to appendix.

-l.245-249: a lot of verbose here, show the figure in a supplement and move the text there

Moved to appendix.

-l.343-344: another example of verbose.

The sentence has been removed.