Reply to Reviewer #1 :

We thank Reviewer #1 for the review and thorough work in broadening the discussion and providing feedback on the readability of our manuscript. The efforts that will help us address any weaknesses in our manuscript are greatly appreciated, and we hope our revised manuscript meets the reviewer's expectations.

The manuscript presents an analysis of gravity wave (GW) kinetic energy distributions, derived from new Aeolus satellite wind profiles, that shows great promise in pushing the needle forward in the construction of observational constraints of gravity waves and their impacts on upper troposphere/lower stratosphere circulation. A methodology is presented for deriving the kinetic energy associated with small-scale GWs in regions of deep convection in the tropics over a period spanning June 2019 to August 2022. Comparisons with ERA5 suggest that the reanalysis product underestimates GW-associated kinetic energy; conversely, GW-associated potential energy comparisons between ERA5 and temperature-profiles from an independent instrument (GNSS-RO) show much more consistency, suggesting that the use of kinetic energy highlights a distinct feature of the GW energy spectrum that is not typically assessed (and, incidentally, is not well represented in ERA5). The authors further speculate that this underestimate may reflect lack of assimilated direct wind observations, in contrast to temperatures, which are assimilated. All in all, the manuscript does a good job of presenting a new dataset with all necessary caveats, while also making a generally convincing case that this new data will be valuable. To this end, I recommend acceptance, pending that minor revisions be made to address the following concerns:

#1. Page 4: There is no description of the GW drag parameterization employed in ERA5. In particular, does the model have an explicit parameterization for non-orographic GW drag due to parameterized convection? If so, what is it and how has it been evaluated/performed in past assessments? This will be important in terms of interpreting the dearth of kinetic energy in the model, relative to the Aeolus-derived energy.

We thank the reviewer for this crucial point. We have now added a detailed description of the non-orographic gravity wave drag parameterization used in ERA5 to Section 2.1 (Data and Methods).

(lines 28-34)

For the study period, ERA5 utilizes the non-orographic gravity wave drag (GWD) scheme described by Orr et al., (2010), which is based on a spectral approach (Scinocca, 2003 ; Referred to as S03 in Orr et al., 2010). This scheme does not explicitly resolve convectively generated waves based on model-diagnosed convection; instead, it launches a globally uniform and constant spectrum of waves from the troposphere. The momentum deposition occurs as these waves propagate vertically and interact with the resolved flow via critical-level filtering and nonlinear dissipation. While this parameterization improves the middle atmosphere climate compared to simpler schemes, evaluations have shown it has limitations in fully capturing the required wave forcing, particularly for the Quasi-Biennial Oscillation (QBO) in the tropics (Pahlavan et al., 2021).

#2. Page 7: Presumably the definition of "background" based on "the arguments presented in Alexander et al. (2008b)" apply to past analysis of temperature, not wind, profiles, no? More generally, it would be good for the reader to have a better sense of the sensitivity of the profiles depicted in Figure 1a to choice of grid box averaging domain, the temporal period over which profiles are averaged (currently set to 7 days, etc.), etc. I imagine the authors have already done this sensitivity analysis, so they could consider showing in an appendix figure.

We have now added text to Section 2.2 to explicitly justify the application of the horizontal detrending method to wind profiles, based on the coupled nature of wind and temperature perturbations in linear gravity wave theory. We have also added a statement confirming that we performed sensitivity analyses on the choice of the averaging domain and found the selected 20°x5°x7-day grid to be a robust compromise between noise reduction and signal preservation, consistent with the original rationale of the method. An appendix figure has also been included.

(lines 186-195)

While this horizontal detrending method was originally demonstrated using temperature profiles in Alexander et al., (2008b), its application to wind profiles is theoretically sound. Linear gravity wave theory dictates that wind and temperature perturbations are coupled manifestations of the same wave phenomena, and thus the principle of separating smaller-scale waves from the large-scale background flow via spatiotemporal averaging is equally valid for both fields. Following the arguments presented in Alexander et al., (2008b), this choice is justified by the need to ensure a sufficient number of profiles per grid cell, which minimizes random noise while preserving meaningful variability in the data. Shorter temporal windows would lead to insufficient sampling, while longer windows would smooth out critical small-scale wave features. The grid size is also designed to preserve the spatiotemporal variability of mesoscale gravity waves and equatorially trapped structures, making it possible to separate the background and perturbation components without introducing significant biases.

(lines 198-201)

We performed sensitivity tests with varying grid sizes and temporal windows to confirm that this configuration provides the best possible background state when prioritizing Aeolus retrieval (see Fig. A1 in Appendix A).

#3. Equation (1): This notation becomes slightly confusing/counterintuitive as the text moves on, since the meridional component often goes to zero due to the pointing vector retaining its approximate angle at ~100 degrees. In other words, V_HLOS would be more intuitively referred to as U_HLOS (or something similar) since, indeed, it primarily reflects the zonal component of the flow. Is there any particular reason why "v" is used instead of something more generic? I suggest changing.

We agree with the reviewer that the notation was confusing. To improve clarity, we have changed `v_HLOS` to `u_HLOS` throughout the manuscript to better reflect its quasi-zonal nature. All corresponding equations have been updated accordingly.

#4. Figure 14, lines 354-355: The first sentence of this paragraph does not make sense to me. In particular, the bit referring to "ERA5 shows a considerable reduction" is vague. Reduction relative to what? Please clarify.

We thank the reviewer for pointing out the vagueness in our original description. We agree the sentence was unclear. We have completely rewritten the discussion of this figure (now Figure 2) to be more direct, quantitative, and clear. Instead of "considerable reduction," we now explicitly compare the peak energy values and geographical structures observed by Aeolus with the more diffuse and lower-energy patterns in ERA5, providing specific energy values (in J/kg) to make the contrast unambiguous.

(lines 340-344)

In stark contrast, Aeolus reveals a picture of much more localized and intense Ek hotspots. For example, during JJA 2020 and SON 2020, Aeolus observes a well-defined hotspot over the Indian Ocean with Ek values exceeding 10-12 J/kg, whereas ERA5 shows only a diffuse enhancement in the same region with values rarely exceeding 5-7 J/kg. Similarly, the DJF 2020/21 hotspot over the Maritime Continent is markedly stronger and more geographically confined in the Aeolus data.

#5. Figure 5: The temporal resolution labeled on the y-axes of these hovmoller plots is too high/unnecessary as it crowds the figures. Please show only every other two or three months. Same comment applies to Figure 7.

We agree with the reviewer. The y-axes on evert Hovmöller diagrams have been updated to display fewer monthly labels, improving the readability as suggested.

#6. Figure 16, Discussion concluding Section 3.2: The discussion here seems weak and understates the disagreement between the Aeolus and ERA5 Ek temporal patterns. The second-to-last paragraph highlights the common features between Aeolus and ERA5, but I think the plots look very different. In particular, the hotspots coincident with low OLR are totally missing in ERA5 (Fig. 5b). The phrasing in the text, however, seems to suggest that the differences are only minor. Please rephrase.

On re-reading, we agree with the reviewer that our original text was misleading and significantly understated the differences between the Aeolus observations and ERA5. We have rewritten this section to emphasize the disagreement. The new text explicitly states that ERA5 "completely fails to capture the intense, high-energy hotspots" and that the high peak energy values are "entirely absent in the reanalysis." To further strengthen this point, we have added a statistical significance test (a two-sample t-test), with results shown as stippling in Figure 3c, to formally demonstrate that the differences are not random but represent a fundamental and systematic underestimation by ERA5.

Line (394-407)

The difference between the two datasets, shown in Fig.3c, quantifies this discrepancy. The plot is overwhelmingly positive, indicating a systematic and significant underestimation of GW kinetic energy by ERA5 throughout the tropics. The regions of greatest underestimation, where the difference exceeds 10 J/kg, align almost perfectly with areas of deep convection, as identified by the low Outgoing Longwave Radiation (OLR) contours. The OLR represents the amount of terrestrial radiation released into space and, by extension, the amount of cloud cover and water vapor that intercepts that radiation in the atmosphere. It is a widely used and reliable proxy for deep

convection due to its strong correlation with diabatic heating (Zhang et al., 2017), reinforcing the conclusion that ERA5's primary weakness lies in representing convection-driven wave activity.

To confirm the robustness of this finding, a two-sample t-test was performed for each grid cell. The stippling in Fig.3c indicates where the mean Ek from Aeolus is statistically significantly higher than that of ERA5 ($p < 0.05$). The pervasive stippling across nearly all convective hotspots underscores that the observed differences are not random fluctuations but represent a fundamental deficiency in the reanalysis. This finding strongly suggests that without the assimilation of direct, high-resolution wind profile data like that from Aeolus, reanalysis models struggle to resolve the full spectrum and intensity of gravity waves generated by localized, powerful convective events. An alternative display of Fig.3c as a ratio, along with an F-test, can be found in Appendix E.

#7. Section 4: Doesn't the ratio of Ek/Ep (shown for ERA-5 in Fig. 8a) suggest that these two quantities are extremely different and not meaningful to compare with each other? I appreciate that the authors want to move beyond traditional (conservative) analysis and attempt to do a bit more, but Figure 8a suggests that the two quantities are in much more disagreement than the discrepancy predicted by llinear wave theory (i.e., factor of 4, not factor of 2). My suggestion here is to introduce Figure 8a earlier as a way to more directly address the concerns with comparing potential and kinetic energy (within a self-consistent product like ERA-5).

We agree with this suggestion that significantly improves the structure of our argument. We have restructured Section 4 as suggested. We now introduce a figure showing the Ek/Ep ratio from ERA5 alone *first*. This serves to demonstrate that even within a self-consistent model, the ratio is highly variable and deviates from simple linear theory, thus motivating why a direct one-to-one comparison of energy magnitudes is insufficient. We then proceed with the observational comparison between Aeolus Ek and GNSS-RO Ep.

#8. Last paragraph on page 19 (lines 467-470): How do you know it's the failure to assimilate the winds directly that's causing the poor representation of GW-associated EK? In principle, one might be able to capture these features using a convective non-orgraphic gravity wave drag parameterization within the ERA-5 model, no? In other words, the assimilation is one way to correct the problem, but an alternative approach is to tackle the model bias directly. However, without having more knowledge about the underlying GW drag parameterization in the model it's hard for the reader to know how many degrees of freedom are afforded to the modeler. Can the authors please comment on the role played here by model bias? And how this is/is not handled by the GW drag parameterization?

We thank the reviewer for this critical question. Our primary argument is based on the inconsistent performance of ERA5 on potential versus kinetic energy.

Our primary argument stems from the inconsistent performance of ERA5 across different assimilated and unassimilated variables. The key piece of evidence is that ERA5 successfully reproduces the potential energy (Ep) field, which is strongly constrained by assimilated GNSS-RO temperature data (as shown in our Fig. 6c). However, it fails to generate the corresponding kinetic energy (Ek) in the very same convective regions, a quantity for which it lacks direct observational constraints.

If the problem were primarily a model physics bias (e.g., the GWD parameterization failing to generate sufficient wave energy), we would expect both Ep and Ek to be systematically underestimated. The fact that only the unassimilated, wind-derived component is deficient strongly points to a failure in the data assimilation system's ability to generate the correct divergent wind field from the available mass (temperature) field in data-sparse regions. We have significantly expanded the Discussion section to elaborate on this reasoning, referencing known limitations of data assimilation systems in the tropics concerning background error covariances and the rotational/divergent wind balance.

Line (584-607)

An additional tool at our disposal to solve the case is the global distribution of Ep, through the use of independent GNSS-RO instruments. Our analysis confirms that the assimilation of GNSS-RO data in ERA5 is highly effective, with minimal discrepancies observed between the reanalysis Ep and direct GNSS-RO observations (Fig.6c). This key finding allows us to arbitrate between two potential causes for the Ek discrepancy: a lack of direct wind data assimilation versus inherent biases in the model's physics (e.g., its GWD parameterization).

Several lines of evidence from our study point towards the lack of wind assimilation as the dominant cause. Firstly, the fact that ERA5 accurately reproduces Ep fields demonstrates that the underlying model can represent the thermodynamic signatures of wave activity when properly constrained. Conversely, the largest discrepancies are found in kinetic energy, a purely wind-based quantity, and are concentrated over data-sparse regions like the Indian Ocean, precisely where Aeolus provides unique wind information (Banyard et al., 2021).

Secondly, while ERA5's non-orographic GWD scheme has known limitations and is not directly forced by diagnosed convection (Orr et al., 2010), it is unlikely to be the sole reason for the missing Ek. Such a parameterization bias would be expected to manifest as a systematic error across different variables or regions, or as a persistent model drift requiring large, ongoing corrections by the assimilation system (Dee, 2005). However, our findings show a targeted deficiency: the model performs well on assimilated temperature (Ep) but poorly on unassimilated wind (Ek) in the very same locations. This sharp contrast strongly suggests the problem is not a wholesale failure of the model's physics to generate wave energy, but rather its inability to correctly partition that energy into kinetic and potential components without direct wind constraints.

In data-sparse areas, ERA5 must rely on its internal background error covariances to infer wind adjustments from the assimilated mass field (Hersbach et al., 2020). These statistical relationships are primarily designed to represent large-scale, quasi-balanced (rotational) flow and are known to be less effective at specifying the smaller-scale, divergent component of the wind field to which convectively generated gravity waves belong, especially in the tropics (Žagar et al., 2004). Consequently, while the assimilation of GNSS-RO constrains the thermodynamic (Ep) aspect of the wave, the system lacks the necessary information and dynamic constraints to generate the corresponding divergent wind perturbations, leading to the observed Ek deficit. This process evidently fails to capture the full spectrum of high-Ek wave modes generated by convection.

#9. Discussion: No mention is made of how these observations might be used to develop constraints on the momentum fluxes (which is what modelers seek most). Is that something that the author has considered? This is a challenging question, so I am not seeking any complete answers here; I am just wondering if the author can speculate in a sentence or two how to potentially bridge V_HLOS with the momentum fluxes.

We thank the reviewer for this forward-looking question. Constraining momentum fluxes is indeed a key goal for the community. We have added a new subsection to the Discussion to speculate on this pathway.

(line 687 – 701)

Looking forward, a critical application for such observations is the constraint of gravity wave momentum fluxes, which are essential for global circulation models. However, deriving momentum flux estimates directly from single-component wind measurements like those from Aeolus presents significant theoretical and observational challenges. The vertical flux of horizontal momentum (e.g., ⟨u'w'⟩) fundamentally requires simultaneous knowledge of both horizontal (u') and vertical (w') wind perturbations. Aeolus provides only a projection of the horizontal wind and, crucially, contains no direct information on the vertical wind; in fact, w' is assumed to be negligible in the standard data processing (Krisch et al., 2022). This represents the primary missing piece of information for a direct flux calculation.

A potential pathway to overcome this limitation involves creating synergistic datasets, for instance by combining Aeolus wind data with simultaneous, collocated temperature measurements from instruments like GNSS-RO. In principle, gravity wave polarization relations could then be used to infer the missing wind components. However, this approach is not a simple remedy and relies on strong, often unverifiable, assumptions about unmeasured wave parameters, including the horizontal wavelength, intrinsic frequency, and the stationarity of the wave field between measurements (Alexander et al., 2008a; Chen et al., 2022).

Therefore, while Aeolus does not directly measure momentum flux, its unprecedented global measurements of kinetic energy provide an additional observational constraint. Such observations are a critical prerequisite for developing and testing the more complex, multi-instrument techniques that will be required to eventually constrain the global gravity wave momentum budget