Reply to Reviewer #2 :

We thank Reviewer #2 for their careful and candid feedback. The comments were helpful in clarifying aspects of the scientific discussion and in prompting us to address some shortcuts and approximations in the original text. We have revised the manuscript accordingly and provide detailed responses to each point below.

The work described here flows from a unique and novel data source, a wonder of the age, and seems fairly sound. However, the figures are not all they could be, and not fully explained, lacking any statistical analysis or mentions of evident artifacts. Many (incomplete) data handling details and other science issues are sprinkled though a rather smooth vague narrative, with too little scientific circumspection. With a modest amount more care, perhaps aided by senior author input, this paper could become as excellent as the data deserve.

The manuscript suffers from three problems common in the dissertation-to-literature translation, detailed further below but listed thematically here:

1. Key technical details for understanding the results are too buried. In place of crisp exposition and taut circumspection are meandering vague mentions of issues around methods, defensive at times and elsewhere a sales pitch for the arbitrary trade-offs settled on. This may be how a committee explained a complex recipe to a student, but is not ideal for a paper addressing peer researchers.

2. Meandering threads are also present as a means for a student to telegraph to a committee their awareness of a reading list, cited often in vague mentions rather than claim-supporting paraphrases of the content. This may be good for dissertations written to a captive readership, but is not ideal for peer researchers. As just one instance, not all winds on larger scales than the filter belong in a named list of Mode Types: the wind is the wind.

3. Findings and interpretations are a bit vague, unobservant of the actual figure set, and repetitive (for instance the “especially Indian Ocean” trope appearing several times). False color scales are unhelpfully distorting, and questions raised by the figures and results are not pursued with a scientifically committed vigor. This may embody the rationally bounded commitment of a student to the physical problem at hand, but ideally a senior coauthor might exert the leadership to bring more depth of inquiry and perspective.

We acknowledge the reviewer’s thematic concerns regarding the clarity of technical details, the narrative style, and the depth of the scientific interpretation. In response, we have performed a substantial rewrite of several key sections. Specifically, Section 2 (Data and Methods) has been reworked to provide a better description of the data processing, free from promotional language. Section 3 and 4 (Results) now feature improved figures with perceptually uniform colormaps and appropriate statistical testing to support our claims with greater scientific rigor. The narrative has been tightened to focus on direct interpretation of the figures. Section 5 (Discussion) has been expanded to address the scientific questions raised by the reviewer with more depth and circumspection, particularly regarding the potential for misinterpretation of convective outflow, the physical mechanisms behind our observations, and the limitations of the analysis.

We address each of the reviewer's specific points in detail below.

Issue 1. Clarifying key technical details.
Fig. 1 was very helpful but the text not so clear.

SUMMARY:
The variance of u and temperature T fluctuations, sub-weekly in period and Fourier bandpassed to 1-9km wavelengths in the vertical, are averaged over a UTLS layer (about 11-25 km after vertical smoothing). The 500m common grid is mentioned far from the other data details (line 124) requiring a second read and search. Half the variance (with rescaling factor for T') is energy. Was the averaging density-weighted (like a physical energy interpretation in J/kg should be), or is it just 1/2 a height-averaged variance? The text is silent.

We thank the reviewer for highlighting these missing details. The detail about the analysis grid (now revised to 100m to preserve maximum detail before analysis) has been moved to a more logical position within the dataset descriptions in Section 2.1. We now clarify in Section 2.2 that the final energy is a simple height-averaged variance over the defined UTLS layer and is not density-weighted. While density weighting would be more physically precise for total energy content, the density variation over the relatively narrow UTLS layer (tropopause to 22 km) is modest, and using a simple average (in J/kg) provides a robust and standard metric for comparing wave activity that is consistent with prior literature.

(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 [...] The chosen grid has a vertical resolution of 100 meters and spans a range from 0 to 30 km altitude.

(line 227)

The profile is then averaged over the selected range, representing the Ek, as seen in Fig.1c.

Lidar Wind profiles:
A side-looking spaceborne lidar measures u(z) along its line of sight, which is almost the zonal direction (that wasn't clear to this reader without web searches). Profiles were processed to isolate deviations from weekly 20x5-degree Lon-lat averages. Then Fourier filtering passed shorter than 9km vertical wavelengths (on a 500m grid so 1km is the shortest). The square of that filtered deviation profile was vertically smoothed with a 7km boxcar, then averaged over a layer (line 203 is ambiguous, why "49 points?" of 500m depth?) The layer is summed from 1km below the tropopause (about 14km in tropics) to 22km (is this mass weighted?), to make seasonal maps (smoothed how? not mentioned) and about 3-weekly (looks like, from figures? not mentioned) longitude-time sections. Artifical slow trends due to (squared) instrument noise variance increasing were estimated and subtracted.

We apologize for the confusing and incomplete description. The ambiguous language ("14-point moving average over the 49-point profile") was a remnant of an earlier analysis step and has been completely removed. We have replaced it with a clear description: the perturbation profile is band-pass filtered, and then the resulting energy profile is vertically averaged over the entire defined UTLS layer (from 1 km below the tropopause to 22 km). We have added a full description of the gridding and smoothing process for the geographical maps in the Methods section. It involves binning the data onto a 5°x2° grid, followed by a 3-point median and 3-point moving average filter applied sequentially in both dimensions. A note has also been added to the relevant figure captions.

(lines 223-228)

The resulting profile, which is essentially the perturbation squared, is cut to keep the data between one kilometer below the tropopause and 22 km [...] The profile is then averaged over the selected range, representing the Ek, as seen in Fig.1c.

(lines 300-303)

To reduce noise and highlight large-scale patterns, a 3-point median filter followed by a 3-point moving average filter was applied sequentially in both the zonal and meridional directions.

Identical processing was applied to RO T(z) and ERA5 winds and T (interpolated on the same grid, again please mention in the data section not introduction). This makes an excellent baseline of comparisons and opportunity for interpretation!

How many pages did the above take to describe in the manuscript? Too many, a tedious read to fish out key details in order to bring a skepticism the somewhat slick text seemed to lack. Lines 172-174 are a good example of sales tone taking over: “making it possible”…”without introducing significant biases”…”configuration mitigates errors”…”ensuring reliable… and robust…” Declarations of success are not very scientific ways to express understandings of trade-offs. The real strength is *identical* processing of comparison datasets. Tossaway adverbs (e.g. “strongly”, “specifically") also set a slick tone in places, undercutting reader trust in the self-skepticism of those best positioned to see problems.

We agree with the reviewer that the previous version leaned too much into a sales-pitch tone. The revised sections adopt a more traditional scientific style, aiming to be clearer, more restrained, and less verbose. While some adverbs can be subjectively judged as helpful or excessive, we have tried to limit them and hope that the new text better aligns with conventional scientific writing.

Issue 2: Clarifying results

2.1 Maps of KE and PE in the tropics and subtropics (Fig. 4) are clearly smoothed, but no mention is made of how and how much and why. Was raw data really too rough for scientific readers’ eyes, or is the paper trying to be too smooth? It is surprising to this reader how strongly confined to the equatorial belt the energy is even in solstice seasons. Is there some kind of conditioning or weighting behind this feature, or is it truly an aspect of convection as a source of signal (’the ITCZ’?)? Or is there an effect of the Coriolis force somehow suppressing sub-weekly sub-7km layer fluctuations? Silence about the smoothing undercuts reader confidence.

We thank the reviewer for pointing out the lack of detail regarding the map smoothing. We have now explicitly described the map smoothing process in the Methods section. Regarding the strong equatorial confinement, we believe this is primarily a physical feature linked to the ITCZ. However, we acknowledge the reviewer's point that methodological choices can influence this. We have now added a sentence to the Results section explicitly stating that our noise-correction scheme, which is weighted by the latitudinal structure of the raw signal, likely enhances this gradient by design. This provides the reader with the necessary circumspection.

(lines 355-358)

Finally, regarding the strong latitudinal confinement of the signal, while this is primarily a physical feature, our noise-correction methodology may also contribute to it. As detailed in the Appendix D (Part 2), the correction is weighted by the latitudinal structure of the raw signal. This approach, designed to avoid over-correction in low-signal subtropical regions, naturally sharpens the latitudinal gradient at the edges of the tropical belt.

2.2 Why no mention of the obvious artifacts in DJF2019-JJA2020, with zero in weak areas, or JJA2019 with 5+ in those areas? Are we all looking at the same figure here and describing its characteristics from the most obvious to the most subtle? Reader confidence is again at stake.

The reviewer correctly identified artifacts in our original plots. These were a result of a flaw in our initial noise correction algorithm that could lead to zero-flooring or baseline shifts. We have developed a more robust, adaptive noise correction algorithm for this revision, detailed in the Appendix D. The new results, presented in the revised Figure 2, no longer exhibit these non-physical artifacts. The energy fields are now physically consistent across all seasons, making discussion of the previous artifacts unnecessary.

2.3. The color scale for positive-definite variance, especially when discussed in a linear meaning like energy, should not have perceptual jumps like this one. Gray shading would be the honest choice, or a single color. This perceptual nonlinearity may be the source of the several-times-repeated “especially Indian Ocean” trope which otherwise seemed inscrutable to this viewer. Seychelles, Diego Garcia, Indonesia; are the other equatorial oceans really so much better covered with “conventional” wind soundings? And anyway, does ERA really get its UTLS wind variability from assimilating rawinsonde data into its imbalanced flow manifold? Did a desire to say something and move on override a thoughtful scientific assessment of the differences between reanalysis and observations, differences fluffed by the redness of a (not accessibility recommended) color map, and perhaps a misinterpretation of u fluctuations in the upper troposphere in the Maritime Continent wet season (detailed below)?

The reviewer's point is well-taken. We have revised all false-color maps to use the cmocean(haline) colormap, which is perceptually uniform and scientifically appropriate.

The reviewer asks two questions about the mechanism behind ERA5's underestimation of kinetic energy. We thank them for pushing us to be more precise in our scientific assessment.

The reviewer is correct that *all* equatorial oceans are poorly covered by conventional *in-situ* wind soundings (rawinsondes), which are primarily launched from land. Our emphasis on the Indian Ocean stems from it being one of the largest and most dynamically significant oceanic regions with an almost complete lack of such soundings, in contrast to the Pacific and Atlantic which have more island stations and aircraft routes. However, the core issue is the general sparsity over all oceans.

The reviewer also correctly intuits that ERA5 does not simply "assimilate rawinsonde data into its imbalanced flow manifold." In data-sparse regions, the analysis is dominated by the model forecast (the "first guess") and adjustments derived from assimilated *mass* field observations (like temperature from GNSS-RO and satellite radiances). The data assimilation system then attempts to infer the corresponding wind adjustments through its background error covariance matrices. These statistical relationships are primarily designed to represent large-scale, quasi-balanced (rotational) flow. They are known to be much less effective at specifying the smaller-scale, divergent component of the wind field, to which convectively generated gravity waves belong, especially in the tropics.

Therefore, our central argument, which we have now clarified and greatly expanded in the Discussion (Section 5), is this: even when ERA5 correctly assimilates the *temperature* signature of a gravity wave from GNSS-RO (constraining its Ep), the system lacks both the direct wind observations and the appropriate dynamic constraints to generate the corresponding divergent wind perturbations. This leads directly to the observed Ek deficit. The problem is not that the model is "missing" rawinsondes *per se*, but that it lacks any source of direct wind information to correct its background state for these specific, dynamically important wave modes.

(lines 587-605)

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 direct wind profile measurements not available from other observing systems (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.

2.4. Fig. 2 is about the anisotropy of waves in ERA5. Is this the second thing a reader wants in a paper about new observations? A value of 2 means isotropic waves, <2 suggests E-W elongation of the variance ellipse, >2 a N-S elongation. Here is a case where a color scale with a perceptual steep part could make sense, but 2 belongs there. Here a less meaningful choice was made. The focus is estimating something like the KE "missing" from the zonal (or LOS) component only, rather than the issue of isotropy which implies something interesting about sources. But all only in the reanalysis, before the first comparisons to even make that dataset as a relevant one. Might a better choice for a second figure be continuing raw obervations (Figs. 3-4) with discussions of the obvious artifacts? Let the ERA5 comparisons wait.

We agree with the reviewer that placing the ERA5 anisotropy analysis as Figure 2 was premature and distracted from the main observational results. We have moved this figure to the Appendix (as Figure C1). Furthermore, we have improved the figure as suggested by centering the diverging colormap on the isotropic value of 2.

2.5. Fig. 5 is a nice comparison, although again distorted by the use of false color for a linear positive quantity. Variance is a jumpy quantity from squaring the data (Fig. 1 shows this nicely) so statistical significance is tricky and surprising. It is customary to use a RATIO rather than a DIFFERENCE of variance (which has no meaning), subject to the F-test for significance. Many students are surprised by how hard it is to pass the F-test with quite a few degrees of freedom. The student should consult a table and appreciate the issue. No statistical testing is evident in the work, surprisingly. That can be overdone or a distraction, but none at all seems weak, again undermining reader confidence.

The reviewer makes an excellent point regarding the statistical comparison of variances. We agree that a ratio subject to an F-test is the standard and most rigorous method. We have now performed this analysis and included the ratio plots with F-test significance stippling in the Appendix E (Figure E1). For the main manuscript, we have chosen to retain the difference plots (Figure 2c and 6c) but have now added stippling to indicate statistical significance based on a two-sample t-test. We believe that for the narrative of this paper, the *absolute difference* in energy (in J/kg) is a more direct and physically intuitive way to communicate our central finding: that ERA5 underestimates the kinetic energy by a certain amount in convective regions. The ratio plot, while statistically pure, can sometimes obscure this absolute magnitude (e.g., a ratio of 2 could mean a difference of 1

J/kg or 10 J/kg). By providing the rigorous ratio analysis in the supplement and adding statistical testing to the difference plots in the main text, we believe we have addressed the reviewer's concern for statistical rigor while maintaining the clarity of our primary message.

2.6. Here is an actual scientific error I suspect: In the wet season over the Maritime Continent, convection is strong and localized and organized on island-strait and diurnal mesoscales that models struggle to represent. The intense divergent outflow of convection in the upper few km of the troposphere (squared) is not UTLS gravity wave energy! The 14-point smoother smears squared wind from a 4.5km layer of the upper troposphere into the averaging layer. Might the pat, recipe-like data analysis choices (Issue 1 above) and an insufficiently critical success-declaring mindset be combining here in a genuine misinterpretation? Sensitivity to this vertical smearing of non-wave wind variance should be assessed.

Based on this interesting doubt, we have performed a sensitivity analysis by recalculating our results using more conservative vertical averaging layers that exclude the upper troposphere (Tropopause + 1 km and Tropopause + 2 km) for Aeolus. The results, presented in the Supplementary Information and referenced in Sections 2.2 and 5, show that the spatial patterns of the energy hotspots are preserved (spatial correlation r > 0.83) and that the vast majority of the energy (~88%) persists well into the stratosphere. This supports our conclusion that we are observing vertically propagating gravity waves, not just tropospheric outflow.

(lines 563-572)

Another consideration in this study is whether the large Ek values observed by Aeolus, particularly over convective hotspots, could be an artifact of misinterpreting non-wave tropospheric outflow rather than stratospheric gravity waves. Our sensitivity analysis (see Fig. B1 and B2 in Appendix B) directly refutes this concern. By shifting the analysis layer upward to begin 1 km and 2 km above the tropopause, we confirm that the geographical patterns of the energy hotspots are remarkably stable (spatial correlation r > 0.83), and that the vast majority of the peak energy (~88-91%) persists well into the stratosphere. If the signal were dominated by shallow tropospheric outflow, the energy peaks would have collapsed when the analysis layer was moved above the tropopause. The fact that a strong, structured signal remains provides compelling evidence that we are observing vertically propagating gravity waves that have penetrated the lower stratosphere. This validates our central conclusion: Aeolus is capturing a significant field of convectively-generated stratospheric gravity wave kinetic energy that is largely absent in the ERA5 reanalysis.

2.6b Figure 7, again a ratio (F tested) would be more meaningful than a difference. The lack of RO signal in the Maritime Continent wet season further strengthens my belief that the Ek is a misinterpretation of convective outflows.

We believe the lack of a strong potential energy (Ep) signal from GNSS-RO in these regions, combined with the strong kinetic energy (Ek) signal from Aeolus, is not evidence of an artifact, but is instead a key physical finding of our paper: that deep convection preferentially generates waves with a high Ek/Ep ratio. Our sensitivity analysis (point 2.6) confirms the Ek signal is stratospheric. Therefore, the combined observations suggest these are high-Ek, low-Ep gravity waves, a phenomenon that challenges simple linear theory and highlights the unique value of direct wind measurements. We have strengthened this point in our Discussion.

(lines 519 – 543)

Fig. 7 presents the first observationally-derived long-term study of the Ek/Ep ratio, comparing Aeolus's HLOS Ek and GNSS-RO-derived Ep […]

The regions with the highest ratio values are systematically co-located with areas of deep convection, as indicated by the low OLR contours. This is particularly evident over the Indian Ocean […] and over the Western Pacific. This observation suggests that, in areas with similar seasonal characteristics, gravity waves tend to transport more kinetic energy during convective events, which amplifies their influence on the overall energy dynamics […]

This observational result stands in contrast to the picture presented by the ERA5 reanalysis in Fig.4. While ERA5 also shows variability in its Ek/Ep ratio, its regions of highest ratio are often located outside the main convective centers. This suggests that ERA5 misrepresents the physical link between deep convection and the partitioning of wave energy.

Given that ERA5 successfully assimilates GNSS-RO temperature data (and thus has a reasonable representation of Ep), this discrepancy points to a fundamental difficulty in the reanalysis's ability to generate the corresponding kinetic energy component in the right locations. Without direct wind profile assimilation in these data-sparse convective regions, the model's parameterizations and background error covariances fail to create the intense, localized kinetic energy associated with convective gravity waves.

2.7. The discussion of wave sources seems shallow. There are various mechanisms including temporally varying convective heat sources (which might set vertical wavenumber and frequency), mountains and transient mountains of lofted air in shear (which might set horizontal wavenumber and phase speed), and more. Would they be anisotropic? The phrase "trade winds" appears in the context of Fig. 2 (anisotropy), as if the surface wind direction has something to do with anisotropy at the tropopause level. Does it? Might one of the senior authors add a little depth?

We agree that this discussion lacked physical depth. We have substantially revised Sections 2, 3, and 5 to address this by providing a clear physical mechanism for how the large-scale background wind imposes an anisotropy on the upward-propagating wave field, addressing the "trade winds" question. We added phenomenological descriptions of distinct convective generation mechanisms (thermal forcing and the "moving mountain" effect) in the results section where these sources are first observed. We placed these local mechanisms within the larger context of planetary-scale organization by monsoons and the MJO. And lastly, we explicitly identified jet-front systems as the likely source for the energy observed at the subtropical edges of our domain.

(lines 249-255)

This ratio (Fig. C1 in Appendix C1) exhibits significant geographic variability, which can be linked to physical mechanisms that create wave anisotropy. For instance, over regions like the Indian Ocean, the ratio is relatively low (~1.5), suggesting a predominantly zonal orientation of wave energy. This is physically plausible, as persistent surface winds like the trade winds can influence the tropopause-level wave field through two main processes. Firstly, flow over orography can preferentially generate zonally-oriented waves (Kruse et al., 2023). Secondly, the background wind profile itself acts as a directional filter, selectively allowing waves propagating in certain directions to reach the UTLS while attenuating others through critical-level interactions (Plougonven et al., 2017; Achatz et al., 2024).

(lines 377-384)

The presence of hotspots, represented by distinct shapes in the Ek patterns, is expected in regions with prevalent convective activity. These can be attributed to multiple powerful wave generation mechanisms occurring at the scale of individual storms. One primary mechanism is thermal forcing, where the pulsatile nature of latent heat release in a convective updraft acts like a piston on the surrounding stable air, generating a broad spectrum of gravity waves (Beres et al., 2005). A second, complementary mechanism is mechanical forcing, where the body of the strong updraft itself acts as a physical barrier to the background wind. The flow forced over this "moving mountain" generates large-amplitude, low-phase-speed waves that are stationary relative to the storm (Corcos et al., 2025; Wright et al., 2023). The intense kinetic energy observed by Aeolus is likely the signature of both mechanisms operating within active convective systems.

(lines 317-322)

A clear seasonal cycle is evident in both Aeolus and ERA5, consistent with the migration of major tropical convective systems. During Boreal summer (JJA), enhanced Ek is prominent over Central Africa and the Indian Ocean. This corresponds to the active phases of the African and Indian monsoon systems, which provide a persistent, large-scale environment favorable for the development of organized, deep convective systems known to be efficient gravity wave generators (Forbes et al., 2022).

(lines 563-566)

The relation between OLR and the MJO has been used before; It is a reliable index for analysis (Kiladis et al., 2014), hinting towards the possibility for the active phase of the MJO to generate the observed hotspots through its convective activity. Recent work has provided direct observational evidence that the MJO modulates GW activity and momentum transport from the tropics to higher latitudes (Zhou et al., 2024).

(lines 327-336)

It is also necessary to clarify the interpretation of the wave activity observed at the subtropical edges of our analysis domain (near 30°N/S). While our study focuses on convectively generated waves [...] the kinetic energy measured in the subtropics is likely dominated by different, local sources. The strong subtropical jets and associated frontal systems are potent generators of inertia-gravity waves through mechanisms of geostrophic

adjustment and shear instability (Plougonven and Zhang, 2014; Achatz et al., 2024). Recent case studies have confirmed that such jet-merging events can produce significant, large-scale GW fields (Woiwode et al., 2023). Therefore, the enhanced energy often visible near 30°N and 30°S in our seasonal maps should be interpreted as stemming primarily from these midlatitude dynamical processes [...]

2.8 Likewise the meaning and source of the strong, anisotropic (more meridional) “waves” (isotropy ratio >2) on the midlatitude edges could be thought about more deeply. Are all subweekly meridional wind fluctuations (squared), from 4km below the tropopause, really UTLS gravity waves, or was that just a tidy story for students? *

The reviewer is right to question our overly simplistic interpretation. The energy at the subtropical edges of our domain is indeed unlikely to originate from the same tropical convective sources. We have added a dedicated paragraph to the results section (Section 3.1) to provide a more nuanced and physically sound interpretation, attributing this energy primarily to local, midlatitude dynamical processes such as jet-front systems, which are potent generators of inertia-gravity waves that fall within our detection window.

(lines 327-336)

It is also necessary to clarify the interpretation of the wave activity observed at the subtropical edges of our analysis domain (near 30°N/S). While our study focuses on convectively generated waves originating from the deep tropics, the kinetic energy measured in the subtropics is likely dominated by different, local sources. The strong subtropical jets and associated frontal systems are potent generators of inertia-gravity waves through mechanisms of geostrophic adjustment and shear instability (Kruse et al., 2023; Plougonven and Zhang, 2014). These jet- and front-generated waves typically have sub-weekly periods and significant wind perturbations, meaning they fall within the detection window of our filtering methodology (Achatz et al., 2024). Therefore, the enhanced energy often visible near 30°N and 30°S in our seasonal maps should be interpreted as stemming primarily from these midlatitude dynamical processes, rather than from the poleward propagation of the equatorial convective waves.

2.9 Figure 8: at last a variance ratio, but only to be taken at face value (with no credible-interval estimation from the F test) in light of some vague gestures at theory whose linearity is considered an easy target. Line 478 says “Fig. 8 presents a detailed analysis” but literally it is just a data plot, with no analysis at all. Too much sales and not enough product for this reader.

The reviewer's critique is fair. The language was promotional, and the analysis was incomplete. We have revised the text to be descriptive rather than declarative (e.g., changing "presents a detailed analysis of the ratio" to "illustrates the longitudinal and temporal variations of the Ek/Ep ratio"). We have added a layer of statistical analysis to the observationally derived ratio plot (now Figure 7). The black stippling indicates regions where the ratio is statistically significant based on an F-test.

2.10 Figure 9: Panel a: Here might be the misinterpretation of convective outflow again, further exaggertated by the false color scheme. Panel b: what is a “dominant” wavelength? Anyone who looks at spectra knows that peak detection is far from trivial and every spectrum is always broad and usually red (more variance falls in eachoi of the wider bins at the low frequency end). What does geometric wavelength really signify over a layer whose stratification goes from upper tropospheric (almost neutral) to 22km (highly stratified)? How does the range here (6500-12000m) relate to the filter which supposedly excludes >9km? Is the spectrum basically red like all geophysical spectra, such that the widest bin near the longest permitted wavelength at the edge of the filter's passband has the most variance? Does that deserve the word "dominant"? Is this figure worth including, or just a thesis figure looking for a place? Is the red exaggeration here the source of the "especially Indian Ocean" trope repeated several times? It's not exactly over the Indian ocean. The authorial prose should reflect a close look, as a reader brings.

We agree with the reviewer's critique of the wavelength retrieval analysis. The methodology was not robust, and the interpretation was flawed. We have removed this figure and the associated analysis entirely. In its place, we have added a paragraph to the Discussion (Section 5) that explains the inherent challenges of performing a meaningful wavelength retrieval with Aeolus data, thereby turning the limitation into a point of scientific circumspection.

(line 665 – 684)

Understanding the vertical wavelength of convective GWs is an essential element for characterizing their dynamics. However, Aeolus is inherently limited in retrieving accurate vertical wavelengths due to its design. The placement of range bins was fixed at the time of observation, introducing inconsistencies in vertical resolution that affect the precise identification of wave peaks and troughs. Additionally, the N/P parameter, which controls the number of accumulated measurements (N) and pulses (P) per cycle, introduces variability in the horizontal resolution of Aeolus data. Changes to this setting, such as the transition from N=30 to N=5, improve horizontal resolution but exacerbate the misrepresentation of vertical wave structures. Furthermore, any spectral analysis of a finite vertical profile is inherently constrained. For geophysical spectra that are typically having more variance at longer wavelengths, a simple peak-finding method would likely identify a dominant wavelength that is an artifact of the analysis window or filtering choices. Given these limitations, we limit our analysis to the vertically-integrated energy within a defined passband (vertical wavelengths < 9 km), which is a more robust quantity.

Nevertheless, we can speculate that the high Ek values observed by Aeolus in convective regions are associated with shorter-wavelength waves. This interpretation is consistent with established physical mechanisms which state that waves with high EK are typically generated in regions with strong convective updrafts and downdrafts, where the rapid vertical movement of air masses creates intense small-scale disturbances. These localized and transient disturbances, arising from geostrophic imbalance, generate GWs that carry energy away from the convective region, where strong forcing efficiently transfers energy into the EK spectrum at shorter wavelengths (Waite and Snyder, 2009). The correlation between high EK and shorter wavelengths is particularly pronounced in convective systems, as confirmed in both observational and numerical estimations (Kalisch et al., 2016), especially in tropical regions and cyclones (Chane Ming et al., 2014). A definitive observational confirmation of this from the satellite itself, however, remains a challenge due to the aforementioned limitations.

3. Discussion should be rewritten with care and thought, in light of all the above. A celebration of this amazing dataset, a technological marvel from such long hard efferts by

so many, deserves more science value than a too-easy critique of reanalysis and/or underlying voids in data sources (common over all the equatorial oceans), and some vague words about how nature is not linear. Some senior author voice could help, if a bit of leadership can be mustered from a committee. Congratulations to so so many people contributing to make this possible! Wonderful data.

We fully agree with the reviewer's assessment. We have rewritten the Discussion (Section 5) to move beyond a simple comparison and to extract deeper scientific value from the dataset, as requested. The new discussion is more structured, scientifically rigorous, and forward-looking. The key changes are summarized below by category.

Instead of a simple critique, we now present a more sophisticated, evidence-based argument for why ERA5 underestimates kinetic energy. We show that since ERA5 successfully assimilates potential energy, the discrepancy points specifically to the assimilation system's inability to generate the divergent wind component of GWs in the absence of direct wind observations.

(lines 587-605)

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... This sharp contrast strongly suggests the problem is not a wholesale failure of the model's physics [...] 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 [...] these statistical relationships are [...] less effective at specifying the smaller-scale, divergent component of the wind field [...]

We now frame our findings within the broader context of tropical dynamics, explicitly linking the observed kinetic energy patterns to the organizing influence of the Madden-Julian Oscillation (MJO).

(lines 558-571)

"The slow eastward propagation of these energy maxima suggests that the underlying wave sources are not random, but are organized by planetary-scale atmospheric patterns. Indeed, the relation between OLR and the Madden-Julian Oscillation (MJO) has been used before [...] and recent work has provided direct observational evidence that the MJO modulates GW activity [...] The structures observed by Aeolus are therefore highly consistent with the kinetic energy signature of gravity waves generated by [...] the large, organized convective superclusters of the MJO."

We have added a new sensitivity analysis to directly address and refute the potential misinterpretation of our signal as tropospheric outflow, thereby providing stronger evidence that we are observing stratospheric gravity waves.

(lines 572-581)

"Another consideration [...] is whether the large Ek values [...] could be an artifact of misinterpreting non-wave tropospheric outflow [...] Our sensitivity analysis (see Fig. B1 and B2 in Appendix B) directly refutes this concern... The fact that a strong, structured signal remains provides compelling evidence that we are observing vertically propagating gravity waves [...] This validates our central conclusion [...] "

We have deepened the discussion on the Ek/Ep ratio, using our unique multi-instrument comparison to show not just that nature is non-linear, but where and why it deviates most from linear theory.

(lines 638-645)

" [...] The observed comparison in Fig.4 of the Ek/Ep ratios from ERA5, Aeolus, and GNSS-RO confirms that the characteristics of gravity waves vary significantly across time and space [...] The frequent observation of ratios exceeding unity, aligning with trends identified in previous studies, suggests that a substantial portion of the waves' energy is contained in kinetic form, often indicative of non-linear behavior [in convectively active regions]."

We conclude the discussion with a new subsection that thoughtfully addresses the challenges and future pathways for using these novel kinetic energy measurements to constrain momentum fluxes, the ultimate goal for model improvement.

(lines 685-699)

"Looking forward, a critical application for such observations is the constraint of gravity wave momentum fluxes [...] However, deriving momentum flux estimates directly from single-component wind measurements [...] presents significant theoretical and observational challenges [...] Therefore, while Aeolus does not directly measure momentum flux, its unprecedented global measurements of kinetic energy provide an additional observational constraint [...] a critical prerequisite for developing and testing the more complex, multi-instrument techniques [...]"