

## **Authors response to reviewers – Reviewer 1**

We thank the reviewer for their helpful and detailed review. We believe the changes made in responses to these comments have improved the paper. Here is our response with the relevant reviewer comments included in grey italics.

### *Main comments*

*I have some questions on how gravity waves characteristics are derived from AIRS observations. You use a fourth-order polynomial in the across-track direction to obtain temperature perturbations (l. 216). AIRS' swath is 1,800-km wide (l. 90). This polynomial thus typically excludes waves with wavelengths longer than 450 km in the across-track direction. On the other hand, there seems to be no similar filter applied in the along-track direction. Is there therefore an inherent anisotropy in the way you derive the temperature perturbations?*

Yes, the reviewer is correct that the fourth order polynomial typically excludes waves with wavelengths  $\sim 500$ km and that there is no along-track smoothing in our method. The across-track background removal is a standard technique with AIRS data and previous work has shown along-track smoothing to be ineffective. Hoffmann et al., 2014[1] experimented with along-track smoothing but they found that problems were introduced where there were strong latitudinal gradients in the temperature field (e.g. the polar vortex). The reviewer is correct that this lack of along-track smoothing would suppress waves with phase fronts oriented E-W (because AIRS has a near-polar orbit). We have adapted the manuscript text to make this limitation clearer.

*In line 224, you make a reference to the "dominant" wave. Could you be more specific here? Does it mean that you only retain the maximum coefficient of the 2-D horizontal Stockwell decomposition in each grid node? If yes, could you quantify the fraction of temperature variance that is discarded in this process?*

Yes, only the largest-amplitude wave properties at each location are retained in the Stockwell transform for this analysis. The full output of the 3DST is a 6 dimensional object and is extremely complicated computationally. Early work is undergoing within the research group to investigate the impact on overall momentum flux when considering overlapping waves in AIRS data (i.e. more than just the dominant wave). For this study however, only the dominant wave is considered. This method is applied to all datasets and does not bias the comparison between models and observations.

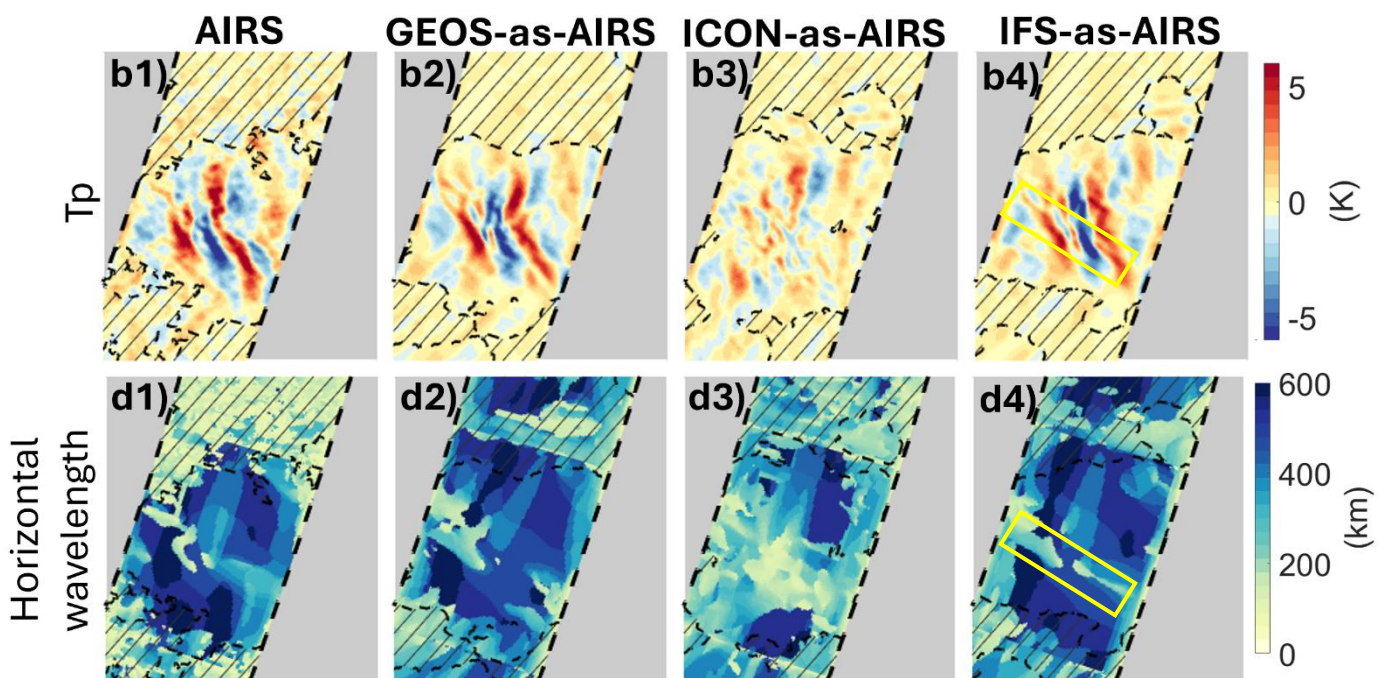
*Last, Figure 3 looks very nice, but I am not sure to know which of the method shown in this figure (if any) is used in the paper. Furthermore, the wave packet shown in this Figure has occurred in 2008, i.e. outside of the time period addressed in this study. I would thus favour a figure that specifically applies to the methodology and time used in the paper (and the accompanying text could discuss the differences/pros and cons of this methodology with respect to previous ones).*

The manuscript text has been edited to further contextualise this figure and highlight that panel (e) is the method used throughout this study. We agree that this was omitted in the previous version. We believe this figure adds useful context firstly to observations of gravity waves in 3D by AIRS and to different methods that have been used to analyse momentum flux from AIRS measurements. This figure shows how differing methods can produce different estimates of momentum flux even when starting from the same initial data.

- Fig 4: The y-axis in subfigures c), d) and e) is neither explicited in the subfigures nor in the caption. This seems to be a normalized count, but this is only a guess. In particular, I would like to know whether this is only a count or if there is some e.g. amplitude-weighted average in d) or e)?

It is a normalised count with no amplitude weighting, this is now included in the figure description and y-axis label of each subplot.

I am somewhat suprised by the horizontal wavelength figures. For instance, Figure d4 shows a narrow band of wavelengths shorter than 200 km in the center of the wave packet, and otherwise wavelengths longer than 400 km. This does not look so obvious in Figure b4: in particular, I am not able to distinguish a clear along-track gradient in wavelength as the one shown in d4. I have a similar comment on figure b1/d1 and b2/d2. Could you comment on this?

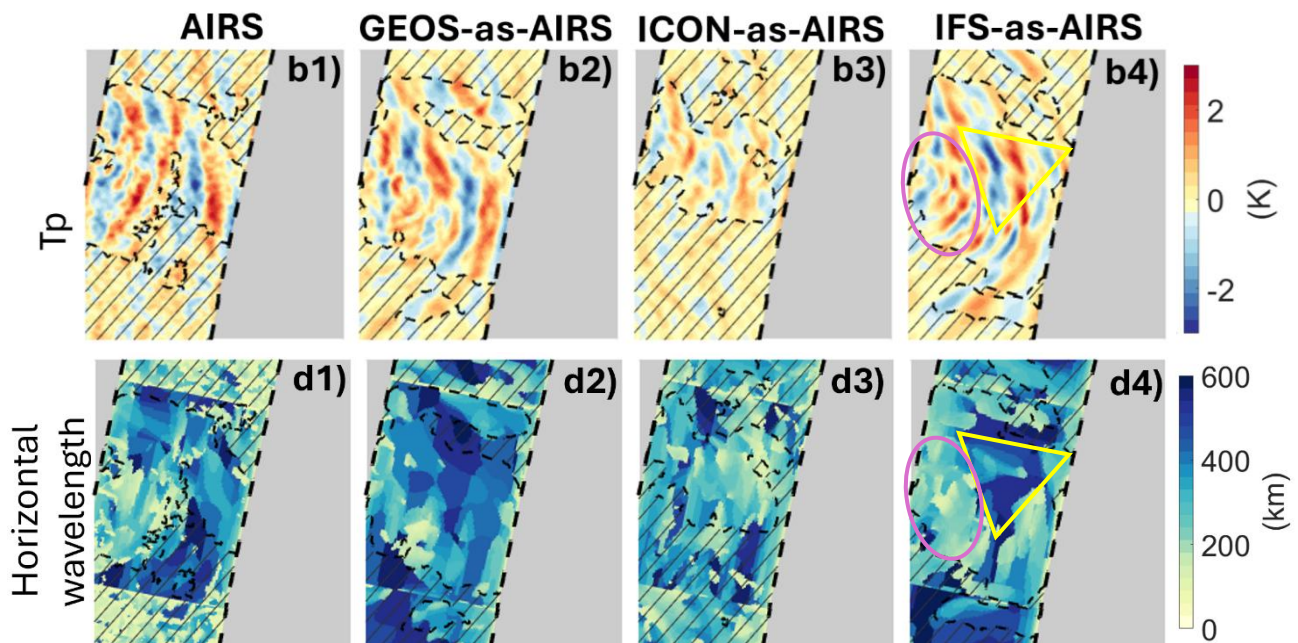


This figure is cropped from Figure 4 in the manuscript from the first orographic example over Iceland. Please note that the results for these two examples have changed subtly from the previous version due to a previous error in the pre-smoothing in the two

examples. This does not impact any of the results in the manuscript. If I understand correctly, in d4) the reviewer is referring to the narrow band of shorter wavelengths that appear almost in an across track direction mid-way through the granule. I have highlighted this in a yellow box in panel d4). I have then copied the box across to the original temperature perturbations from panel b4). This yellow panel lines up with the region of the wave where we do see wave structure in the temperature perturbations that has a lower horizontal wavelength and I believe that it is this that contributes to the region of shorter horizontal wavelength measured in d4). I believe that this shorter wavelength-wave structure is also present in panels b1/d1 and b2/d2. Furthermore, we note that the horizontal wavelengths that result from the S-transform are not smoothly varying as one might expect. The discontinuities in the measured horizontal wavelengths arise because we choosing the dominate wave at each pixel meaning the wave that is measured will change across the spatial extent.

- Fig 5: I have the same comment on the horizontal wavelength here: for instance, I do not understand the dark blue patches in Figure d4).

As before the discontinuous patches arise a result of the ST analysis and the dominant wave. The dark blue patch in d4) correspondes to a wave packet with longer wavelength waves (yellow triangle). In the pink oval I believe we are seeing a wave packet with shorter wavelength waves, possibly superimposed on the larger wave. The ST is picking the dominante wave in this area which in this case is the shorter wavelength one. This creates the apparent discontinuities.



*I also note that the vertical wavelength histogram shown in Figure e) exhibits mostly wavelengths shorter than 16 km, while AIRS vertical resolution is between 7 and 13 km (line 97). Hence, the amplitude of those wave packets are likely significantly affected by AIRS observational filter. This may be stressed in the text.*

This is an important point and we have now included it in the text.

*Finally, the AIRS wave amplitude in this example peaks at about 1K, while it is stated in line 96 that the retrieval noise is between 1.4 and 2.1 K. It may be worth expliciting how the detected wave amplitude could be smaller than the noise level.*

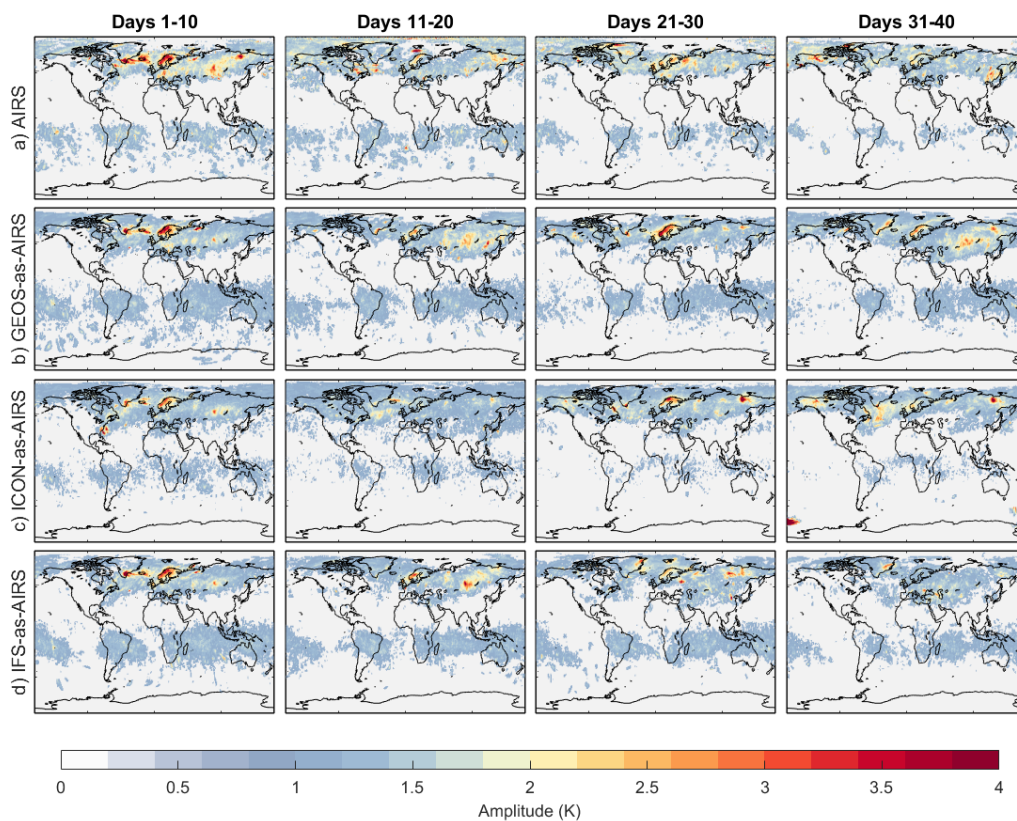
Noise will appear as semi-random temperature perturbations on top of the wave perturbations themselves which means that the wave oscillations can still be visible to eye even when their amplitude is below the noise amplitude. Additionally, our application of 3x3 boxcar smoothing can smooth out a lot of the small-scale random pixel to pixel noise, allowing us to observe waves below this stated noise floor. We have now included the details on this smoothing in the methods section.

The wave detection method is a machine learned method (Okui et. al., 2025) which is based on an original detection method, known as the Neighbourhood method (Berthelemy et al., 2025). This method uses regions of spatially consistent zonal and meridional wavenumber to define a wave. This means that the wave detection method does not rely on wave amplitudes.

*- In several instances, the study claims that gravity waves are underestimated in high-resolution simulations when compared to AIRS observations. I am not sure that this is fully supported by the material shown in this paper. This statement for instance appears in Section 4.3 (namely around line 370 and following ones), and is associated with Figure 8. Focusing on the first ten days of simulation (left column), this figure indeed shows higher gravity-wave amplitudes in observations mostly over Greenland, Iceland and Eurasia than in simulations. Yet, the previous section (especially Fig. 6) has shown that AIRS observations are biased toward higher-amplitude waves than the simulations. Hence, the conditional (wave-packet) mean, which is used to produce Fig. 8, seems to average both a mix of low- and high-amplitude waves in the simulations, while it averages mostly high-amplitude waves in the observations. I would be for instance curious of what would be the comparison if the same 1-K threshold that was used in Section 4.2 is applied here. The amplitude histogram in Figure 4c furthermore illustrates my point: AIRS seems to compare well with simulations for higher amplitude waves, but underestimate lower amplitude ones.*

This is an interesting point and we have investigated further. We have reproduced the conditional mean amplitude figure (Figure 8) but with both the masking and the 1K threshold applied. Please note the extended colorbar to 4K compared with Figure 8 (0 to 3K). This figure does shows better agreement between the AIRS observations and the

models than the original Figure 8, suggesting that (as the reviewer states), AIRS agrees well with simulations for higher amplitude waves and the bias is arising as a result of the low amplitude waves. We have included this figure in the supporting information. Note we do not include the unconditional-mean amplitude and momentum flux figures in the appendix as they are unaffected by the additional cut off and are almost identical to those figures shown in the main text body. This is likely because the conditional mean amplitudes contain many near-zero values and assigning them to zero does not cause any consequential changes. We have also made changes to the manuscript to remind the reader that there are more low-amplitude waves in the models and this will be influencing the time-averaged amplitudes.



*I also note that the unconditional mean comparison (Figure 9) does not show a systematic difference in wave amplitudes between observations and simulations. Yes, agreed. This is now noted in the text*

*- The sentence that starts in line 392 seems to announce that diagnostics as those shown in Plougonven et al. (2017) will be shown. On the other hand, the following lines consist in a discussion based on several previous figures and relying on the reader's eye to compare them. This discussion is not very easy to follow (e.g. the sentence on line 401 looks odd). Wouldn't it be easier to show the pdf of wave amplitudes conditional to wind speed?*

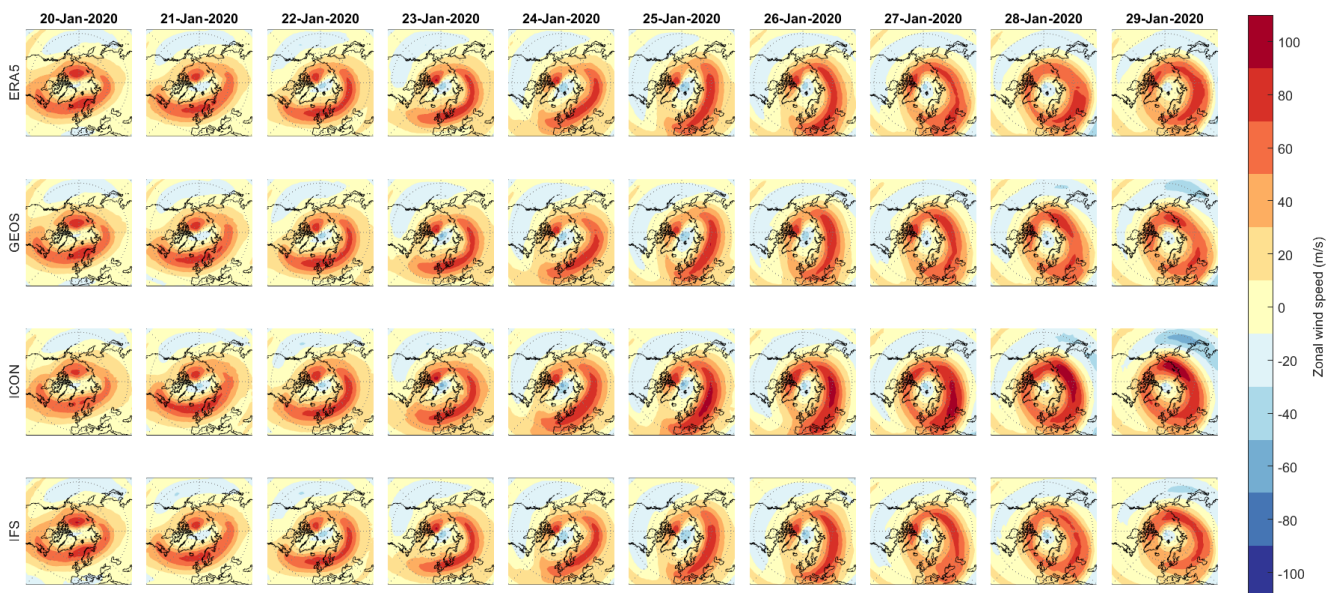
We have rephrased the discussion for clarity and now also included Figure (Fig. 11) in the manuscript which captures the relationship between the wave amplitude and zonal wind speed.

- line 445-446: *I do not agree with this sentence: the lowest values associated with the NH peak in Figure 11e are those of AIRS and GEOS (not ICON), and I do not see the factor 3-4 difference between the 4 datasets.*

The reviewer is correct, this 3-4 times factor is referring to the SH peak. This has now been corrected.

- line 453, last sentence of Section 4.5: *Yet, a small meridional displacement of the jet center in models with respect to the reanalysis might explain the feature discussed here. Maybe worth double checking.*

It was worth a check, thank you. We now include the figure below in the supplementary material for the manuscript. This figure shows the daily average winds at 33km altitude for each of the first 10-days in ERA5 and each of the models. Around the 28<sup>th</sup> January the stratospheric polar vortex is positioned further south than any of the models in the vicinity of the discussed region. This could possibly explain some of the difference in meridional flux over this region. We have rephrased the text to reflect this.



- line 469-471: *I am not sure to fully agree: waves approaching their critical level would see their vertical wavelength decreasing. In numerical models, this could result in*

*decreasing amplitudes as they become sub-grid scale in the vertical. And those waves would soon break and thus contribute to the general circulation!*

This is a fair point, thank you for bringing it to our attention. We have removed the statement that these waves would not be important.

#### *Specific points*

- line 175: *Rephrase the sentence that starts here.* done
- line 289: *"Outside the main wave signal": do you mean in the hatched region? Yes, text rephrased to make this clearer.*
- line 408: *Rephrase the sentence that starts here.* Rephrased for clarity
- line 411: *wavesm -> waves* done
- line 414: *What does the second "this" of this sentence refer to?* Rephrased for clarity
- line 448: *What do you mean by an "overall balanced flux"?* This has been rephrased to 'overall flux'

#### References

[1] Hoffmann, L., Alexander, M. J., Clerbaux, C., Grimsdell, A. W., Meyer, C. I., Roessler, T., and Tournier, B.: Intercomparison of stratospheric gravity wave observations with AIRS and IASI, *Atmos. Meas. Tech.*, 7, 4517–4537, <https://doi.org/10.5194/amt-7-4517-2014>, 2014.

## **Authors response to reviewers – Reviewer 2**

These constructive comments have helped us improve the manuscript and we thank the reviewer for their time. Here is our response with the relevant reviewer comments included in grey italics.

*A suggestion on the structure:*

*Section 4 and 5 are termed Results and Discussion, respectively. However, I noticed that the authors contextualize and discuss their results quite extensively already within the Results Section. Therefore, I suggest to call Section 4 Results and Discussion and include lines 491-513 to Section 4.3, lines 513-540 to Section 4.5, and lines 540-555 to Section 4.6. In addition, I suggest to include Section 4.7 termed for instance Limitations on Methods. I suggest to include lines 556-562 to this new subsection.*

We have re-arranged as suggested, however felt that lines 513-540 were more appropriate in the ‘Limitations of methods’ section.

*A few minor suggestions:*

*Throughout the paper, please abbreviate northern hemisphere as “NH”, southern hemisphere as “SH”, gravity wave momentum flux as “GWMF” once and stick to the abbreviation from then on. In the current version it’s not consistent. Fixed*

*Line 4: it should be “horizontally”. - fixed*

*Line 12-13: It would be a bit clearer to say “Agreement is more consistent in the northern hemisphere (where orographic waves dominate) than in the southern hemisphere (where convective waves dominate).” – agreed and changed as suggested*

*Line 82: AIRS was already introduced in line 33. Please erase “(Atmospheric Infrared Sounder)”. - fixed*

*Line 100: I guess the numbers given here for wavelengths AIRS is sensitive to are based on the 90% sensitivity contourline in Figure 2c of Hindley et al. (2019). I would suggest to soften the criterium here a bit and rather stick to the 50% line and state “Being a nadir sounder, AIRS is most sensitive to GWs with short horizontal (<1,000km) and long vertical wavelengths (>16km).” This is simply because your histograms in Figure 4e, 5e, and 12 e-h show a lot of waves with vertical wavelengths shorter than 30km! – good suggestion, thank you, changed as suggested*

*Line 200-207: I’d suggest to erase these lines because they distract the reader. The Lear et al. (2024) study is better placed in the discussion. Text removed as suggested*

*Line 225-226 and Figure 3: I’d suggest to erase this line and Figures 3c and d because you’re not mentioning or referencing to them in the body of text.*

The manuscript text has been edited to further contextualise this figure and highlight that panel (e) is the method used throughout this study. We agree that this was omitted in the previous version. We believe this figure adds useful context firstly to observations of gravity waves in 3D by AIRS and to different methods that have been used to analyse

momentum flux from AIRS measurements. This figure shows how differing methods can produce different estimates of momentum flux even when starting from the same initial data.

*Line 230: I would love to see a bit more discussion on this upward-propagation assumption in the suggested extra subsection 4.7. Dörnbrack et al. (2018) have shown significant wave activity in the lower stratosphere of downward propagating GWs during a minor SSW over Europe. What would be the consequences if an actual downward propagating wave is taken as an upward propagating wave in your analysis?*

We agree that the upwards propagation assumption was not discussed in detail in the text and we have included a small extra discussion on this in the manuscript.

To provide extra context around the reviewer's specific comment, in the Doernbrack study they generally find downwards propagating waves in the lower stratosphere and upwards in the upper stratosphere which they conclude suggests an in-situ stratospheric generation of waves. They predominantly find upward propagating waves except at altitudes of 12-20km, 12-28km (depending on dataset/analysis method). Our study is carried out at 33km altitude above this region of downwards propagation. Hence the upwards propagation assumption that we make at is consistent with the Doernbrack study at regions around 33km. Additionally, the purpose of assuming upwards propagation allows for the ambiguity of wave direction to be made without relying on externally derived winds (e.g. reanalysis which exhibit significant biases at stratospheric altitudes). This was investigated by Hindley et al., 2019. The consequences of labelling a downward propagating wave as an upward propagating wave would be a reversal in the directional momentum flux.

In our results we find that directional momentum flux is almost always against the background stratospheric winds, suggesting that the assumption of upwards propagation holds at 33km.

*Line 261: The Hindley and Wright reference should be included via \citep{}}. Changed*

*Line 263-267: I'd suggest to erase these lines because they distract the reader.*

**Removed as suggested**

*Line 283-288: I'm a bit (not to say very) confused by this part. You're stating that the horizontal wavelengths as measured in ICON at 33km on the 21st January 2020 match well with the horizontal wavelengths measured by a completely different instrument (GLORIA) at a different altitude (10-15km) four years earlier? Furthermore, measured wavelengths (I assume in ICON) agree with calculated wavelengths (I'm not sure what the difference is here between measured and calculated wavelengths...) measured by AIRS over South Georgia. Please either erase these lines or rephrase. Admittedly, I'm not sure what point you're trying to make.*

This point has now been rephrased. The intention was to demonstrate the wavelengths measured are similar to those made by GLORIA at the same time of year. Here, 'measured' and 'calculated' mean the same but these sentences have been rephrased in the manuscript for clarity.

*Figure 4: Please add a label to panel a). The a) is missing. Also, there is a tiny leftover title at the very top (red dots and dashes) and the "N" of "70°N" is hiding just a bit behind the image. Maybe you could fix that. Also, regarding panels c, d, and e, you might want to consider to draw bars with four different colors instead of lines. The sometimes overlap and make it hard to read. Regarding panels e1-e4, I'd also suggest to extend the colorbar towards values of 50km since the histograms indicate that waves with vertical wavelengths even larger than 50km are present.*

*Figure 5: Please add a label to panel a). The a) is missing. Also, the "S" in "30°S" is hiding behind the image. Like I commented before, maybe you could work on the readability of panels c-e.*

We have now fixed these formatting errors in Figures 4 & 5. We did experiment with filled histograms for this figure but unfortunately, the colours overlap in the area under the histogram making it difficult to see when one line falls below the other three distributions.

*Line 296: I'd suggest to erase "(note the smaller amplitude colourbar ranges)". Done*

*Line 309: Is it the 28th of February or the 29th? 29<sup>th</sup>, this has been corrected.*

*Line 329: Please add Sato et al. (2012) and Ehard et al. (2017) as references for the oblique propagation of waves into the jet. Done*

*Line 378: It's a bit hard to understand. I suggest to write "[...] wherein no wave is detected is defined as a wave with zero amplitude". Rephrased for clarity*

*Line 381: Please erase "mean". Done*

*Figure 10: I think the y-labels make no sense here. Please correct them. For panels a-d the y-labels are the datasets, this is consistent with Figures 6-10. For panel e) we have included the latitude y-label and longitude x-label on the topography map.*

*Line 411: Please erase the -m attached to wavesm. Done*

*Line 444-446: I think you're mixing up the two peaks in the SH and the NH. The factor of 3-4 in the fluxes is valid for the SH peak and not the NH peak. Please double check. Yes the reviewer is correct, this has now been fixed.*

*Figure 11: Ingenious way to visualize a vector quantity in 2D. Two tiny comments. First, on the colorwheel, please push the "N" and "S" a tiny bit to the right. They are not centered. Second, the y-ticks in panel e) do not allow for unambiguous clarity whether the y-axis is linear or logarithmic. Thank you, we have fixed the positioning of N and S.*

We have clarified in the caption that the y-axis is linear.

*Line 487: Please add Gupta et al. (2024) and Garny (2025). There appear to be 4 Gupta et al. 2024 papers, and we were unable to locate one that was especially relevant to this*

sentence. Please could the reviewer clarify? We have included a reference to Garny et al., 2025.

*Line 496: There is a space missing before “Gong”. Corrected*

*Line 505: Please add Gupta et al. (2024) and Gisinger (2022). Again, please could the reviewer clarify the Gupta 2024 paper? We are not sure if the Gisinger paper is relevant as it looks at GWs in the SH winter over the Andes, please could the reviewer clarify.*

*Line 533: Please erase “(Atmospheric Infrared Sounder)”. AIRS was already introduced. Corrected*

*Line 538: Please add a period after “behaviour”. Corrected*

*Line 551: Please erase “be”. Corrected*

*Line 565: Is it the 28th or 29th of February? Please check again. It is not consistent with line 309. 29<sup>th</sup> February is correct, thank you for noticing the inconsistency!*