Response to Reviewer 1:

General Answer:

Thank you very much for your helpful comments; they have made this paper much more readable and have improved the scientific results. We have updated the structure of the introduction and methods sections to introduce the methods used here earlier and more logically. We have also updated our results by applying the masks from the neighbourhood method onto the smoothed input data. This makes it more comparable to other studies that have used this technique and removes the bias from line 335.

As mentioned above, our changes involved smoothing the input data onto which the binary masks are applied. This was a critical step for ensuring comparability with Hindley 2020 and was mistakenly omitted from the initial draft. Implementing this adjustment has resulted in meaningful changes in some results, including sharper directional distributions and an improved amplitude cutoff detection method. We believe our method still outperforms this other method, thus significantly strengthening our argument. We thank the reviewer for their diligent comments, which helped us identify this issue.

General comment:

The authors of this paper present a new methodology for identifying GWs in AIRS observations in the stratosphere. They use a well-known method based on the 3D Stockwell transform and extend it by adding a filter that requires the horizontal wavenumber to remain approximately constant over a defined area. The application of this filter enables the detection of GWs with amplitudes lower than the thermal noise of the instrument and excludes wave amplitudes that are associated with atmospheric noise in an outdated approach. The authors conduct statistical analyses over a period of five years and compare the new methodology with the outdated one. It becomes clear that the new methodology offers many advantages. Regarding methodology and results I would recommend to publish this work. However, several issues regarding the explanation of methods and contextualization of results need to be resolved, which warrants a major revision.

My main points of criticism are:

1. Clarity and precision

The manuscript contains ambiguities in many places. Often, general phrases are used, unclear links are made, or the sentences are overly complex. This makes it difficult to understand the text. Unfortunately, the references are also often not chosen appropriately.

2. Definition of key terms

At the core of the paper, three different methodologies for determining GW amplitudes, momentum fluxes, and vertical wavelengths are compared. Unfortunately, the three methodologies are not described in sufficient detail. A brief comparison, for example using a table, would enhance the paper.

3. Treatment of measurement uncertainties

While it is clear that the new methodology is more robust against the Gaussian noise of the instrument, it is not explained how the measurement uncertainty affects the spectral analysis and hence, determination of wave amplitudes, wavelengths, and ultimately the momentum flux.

In the following you will find my detailed review. I hope for a good discussion and look forward to seeing the results soon in an improved and clarified form that does justice to the hard work that has been put into this.

1. Introduction

Short but good. Quickly funneling down from "gravity waves are important" to "We present a novel method to identify GWs in AIRS data". However, often times it is not clear what the authors are getting at. So, the introduction could need some polishing. In that sense, have a look at my suggestions below.

Line 18: Please add at least one reference for every mentioned impact of GWs.

We have now added Olafsson 2009, Gardner 2024, Garfinkel 2018, and Smith 2005 as references for each impact.

Line 19-20: Technically not wrong, however, I suggest to rephrase the sentence in the following way: "GWs propagate throughout the atmosphere, but have their largest effects at higher altitudes, as according to linear theory conservation of energy and the exponential drop in density enforce their amplitudes to grow exponentially with height."

This is a useful suggestion, this has been rephrased according to your suggestion.

Line 21: I do not agree with the statement that "by far the most important" mechanisms are the excitation by orography and deep convection. For one, it depends on the altitude range one is interested. Second, these are "by far the most studied" mechanisms. Does that mean, they are the most important mechanisms? I'd suggest to weaken that statement a bit or find a proper reference quantifying the dominance of these mechanisms over the other.

This has now been changed to weaken it by swapping “by far the most important” with “such as”, as we agree with this comment.

Line 21: The Smith and Lyjak (1985) reference is not a good one for orographic GWs. There are plenty better ones, like the very first works by Kuettner from the 1940s, Dörnbrack et al. (2002), Ehard et al. (2017), Hecht et al. (2018), Pautet et al. (2021), etc.

We have removed the Smith and Lyjack (1985) reference and added Dörnbrack et al. (2002) in its place.

Line 23: Please add a more specific reference like Fritts and Luo (1992).

Fritts and Luo (1992) has been added.

Line 24: Please add a reference for non-linear wave-wave interaction.

Schlutow et al. (2020) has been added as a reference for non-linear wave-wave interaction.

Line 25-29: That is a bit irritating. I’d suggest to simply state: “These diverse mechanisms can lead to large differences in the wave properties, with the vast majority of GWs having horizontal wavelengths in the range of a few kilometres up to thousands of kilometres (Choi et al., 2012; Kalisch et al., 2016; Trinh, 2016; Hájková and Šácha, 2023) and periods from hours to days (Dunkerton, 1982; Baldwin et al., 2001; Ern et al., 2021).”

We agree with this comment, so this has been changed according to your suggestion, with a slight change from “periods from hours to days” to “periods from minutes to days” in accordance with Reviewer 2.

Line 35: Please change that to “This allows for 3D measurement of temperature and hence, detection of GWs.”

This has been changed to your suggestion.

Line 37: I suggest to rephrase this line in the following way: One problem previous studies were dealing with is the proper filtering and identification of GW signatures in AIRS temperature data.

We have changed this to your suggestion.

Line 37-40: At this point it’s not clear what an “AIRS temperature perturbation footprint” is. Maybe it's more clear to say: "Hoffmann et al. (2013) calculated the temperature variance in a running fashion over regions of 100km radius and compared the result with a predefined threshold in order to identify GW signatures."

We agree with this comment, and have changed the text to reflect this.

Line 40-42: I suggest to say: “Hoffmann et al. (2016) improved on this by taking variance differences between two boxes, one over an orographic hotspot and the other upwind of this hotspot to define a background or reference variance.”

This suggestion in this comment is more explanatory and concise than my text, so we have changed the text to reflect that.

Line 42-43: Please erase "This was used to identify whether orographically generated GWs exhibited larger variances downstream of the mountain than upstream."

This has been removed.

Line 46-47: Please change to: "This resembles a continuous wavelet transform except for the fact that the complex phase is kept constant."

This has been changed.

Line 48-49: Please erase "Such an analysis assumes that a wave is present at all locations in the data, and so will fit wave properties even if there is no wave."

This has been removed, but "When no wave is present in a volume, the ST fits low amplitude waves to the AIRS pointwise thermal noise in that volume" has been added in its place, as this is directly relevant to the method developed in the study.

Line 49-54: Please rephrase these lines. The term "voxel" was not properly introduced yet and it remains unclear what it refers to. Also, the "amplitude-cutoff method" is not properly introduced. Is that the "amplitude-cutoff approach" by Ern et al. (2017)? Please make more clear how the ST separates between GW and noise features. It seems like the convolutional nature of the ST considers multiple independent data points and therefore the noise floor is lowered, right?

References to the "amplitude-cutoff method" and to "voxels" have been removed, and are explained further on in the text.

The ST does not distinguish between gravity waves (GWs) and noise—this is a key motivation for our paper. It is a spectral analysis technique that provides localized spectral information but does not identify the source of the energy. While the reviewer is correct that the convolutional nature of the ST involves multiple data points—effectively averaging spectral properties over a local area and lowering the noise floor—this alone is insufficient. Noise can still produce high spectral energy in small regions, sometimes exceeding that of large-scale GWs. In our approach, we select the largest spectral peak at each X, Y, Z location, which often corresponds to noise unless a strong GW signal is present. Other methods attempt to identify multiple peaks at each location, but this is computationally intensive and the peaks are often poorly defined. Our proposed neighbourhood method addresses this limitation by using the characteristic differences between coherent GW signals and random noise to improve detection.

Line 55-56: S-transform was already introduced. Please abbreviate it with "ST".

This has been changed.

Line 56: Please be more concrete here and tell the reader what kind of wave properties you have in mind.

The wave properties have been added.

Line 58: I suggest to use “constant” instead of “stable” in that context.

This has been changed.

2. Data

Perfect as is.

This is much appreciated.

3. Methods

3.1. Preprocessing

Very well written. Please highlight that a granule is a 3D data array with 128 x 135 x ??? pixels and introduce the term voxel in that context. Also, the bias mentioned in Section 4.2 surprised me that late in the manuscript. Since you mention the sensitivity to GWs with wavelengths in the range 30/80km to 600km, I would like to see a more precise statement maybe about the sensitivity of zonal and meridional wavenumbers separately.

The “128 x 135 x 13” voxels have been added. Yours and the other reviewers' comments highlighted issues with the processing chain, which have since been fixed, as mentioned in my general answer above. This has fixed the bias problem.

The sensitivity to zonal and meridional wavenumbers depends on the latitude, as the instrument sensitivity is tied to the cross-track and the along-track directions. These directions vary as a function of latitude, especially as it goes over the poles. The fourth-order polynomial detrending is in the cross-track direction, which changes with latitude, so the sensitivity of the zonal and meridional wave numbers cannot be said explicitly.

3.2. S-Transform Method

This paragraph is too short and lacks clarity. Please recall what the S-Transform is and state the equation, how to compute the ST. How is the wavelet normalized? Do the results differ significantly when using another wavelet normalization?

ST-based methods have been used widely since 1996; accordingly, we believe a detailed description is not necessary. The citations within the paper explain the ST and how it is used for GW science in detail. However, we have now added a short brief on the ST.

Line 104: Please explain, how you characterize the properties of dominant waves in each voxel. Why is the ST applied to two and not one granule at a time?

We have changed “characterised” for “determined”, as it more aptly fits what was meant. The ST is applied to two granules, as it reduces the number of edge truncation effects that can occur at the edges of the granules. It should be noted that a granule is an arbitrary number of scans long, so this number of granules was chosen empirically for the most efficient processing and least edge effects.

Line 106: Be concrete. What is meant by “the full three-dimensional dataset”?

This has been updated with a description of what is being S transformed.

Line 107-108: How exactly do you constrain the 2D+1ST approach using only those spatial frequencies with the greatest spectral magnitude in the 3DST output? Is that documented in another publication?

The method is based on Wright (2021), with a minor modification. Instead of applying a 2D ST at each level to estimate horizontal wavelengths, we use a 3D ST on the entire granule and extract the horizontal wavelengths from each level. At each X, Y, Z location, we select the wavelength corresponding to the peak spectral amplitude, then apply the phase shift method as described in Wright et al. (2021). For a detailed overview of the ST approach, see also Hindley (2019).

Line 114: It might be worthwhile to mention that Reichert et al. (2021) found that 20% of the detected GWs are apparently downward propagating in the stratosphere over the Southern Andes.

This has now been mentioned.

Line 117-119: These two sentences seem contradictory. Please rephrase. Also, this is not a good place to mention “neighborhood method” for the first time in the manuscript since it is not explained yet.

This has removed, as it was confusing for the reader.

3.3. Wave Detection

Line 121-122: Not a good opening. The reader doesn’t know what the “neighborhood method” is, yet. State the problem and how you are gonna solve it.

This section has been overhauled. A subsection outlining the problem includes a brief description of the methods used here before moving on to each method in detail.

Line 128-129: It is suggested to rephrase these lines for clarity: “The ST method is applied to the concatenated granules providing wavenumbers along-track, cross-track and in the

vertical as functions of space. After that, the satellite track referenced wavenumbers are projected into zonal, meridional, and vertical wavenumbers k, l, and m.

This has been rephrased.

Line 136: Please change as suggested: "Regions where wavenumbers are very similar in a 5x5 neighborhood, which is consistent with our requirement for an extended GW field.

This has been rephrased.

Line 147: Please use either "point", "pixel" or "voxel" consistently.

This has been updated.

Line 147-155: Please create a subsection for the computation of momentum flux.

Appendix A. of Ern et al. (2004) derives this equation, and has been referenced in text, from which the output of the ST was passed into it as the variables.

Line 153: g=9.69m^2/s^2 at 39km altitude. (Just for physical correctness, I am not asking you to redo the analysis.)

This has been fixed.

Line 153: Is N measured or assumed?

N is assumed here.

Equation 1: This is a simplification of Ern's formula. Please argue why you can use this simplified version.

We can use the simplified version because the GWs analysed here are all within the mid-frequency range, which we have checked by using the wave dispersion relation to find the extrema of the intrinsic frequencies. These lay between N and f, as is necessary for the mid-frequency. We have included this in the text.

4. Results

Please change the section title to "Results and Discussion".

This has been updated.

Line 159: Please change the citation style.

This has been updated.

Line 161-164: Please erase "Specifically, it shows the mean amplitudes of detected GWs during local winter months, November-February (NDJF) for the Northern hemisphere and June-September (JJAS) for the Southern hemisphere. The difference (g,h,i) and ratios (j,k,l)

between the two methods are also presented." This improves readability since this information is in the caption of the Figure.

We believe this is a stylistic choice, as some people do not read the figure captions.

Line 168: I would expect the mean of the noise to be zero. Do you mean the mean of the noise amplitude?

Yes, this has been updated.

Line 173: I wouldn't use "especially" here, since both "GW belts" nicely show up in the maps.

This has been updated.

Line 175-176: The statement makes sense. However, please deliver more proof that wave activity is more persistent over the Antarctic peninsula. Figure 12 in Wright et al. (2017) shows an enhanced Gini coefficient for southern hemispheric winter months over the Antarctic peninsula, indicating not constant but intermittent GW activity! Therefore, I would expect a stronger deviation between the amplitude cutoff and the neighborhood method.

This has been updated to better reflect what was meant: high-amplitude waves with a clearly defined wave structure. Consistent was the wrong word to use here, and it has been removed.

Line 180-186: Concerning Figure 3: Please provide a linear color scale to enable a fair comparison with Figures 6 and 7 from Hoffmann et al. (2017).

This figure has been created and placed in the supplementary documents, a log scale is needed to see the lesser detected areas, such as convective regions.

4.1. Histograms of Latitude Bands

Line 188-189: You never properly introduced the amplitude cutoff method. (I'm still assuming it is the amplitude-cutoff approach by Ern et al. (2017)) Now a third method is mentioned without introducing it, the ST. Please provide proper descriptions of the three different methods in the methods section.

As stated in our General comment, we have restructured the methods section to now definitively state what each method is, and have mentioned the methods used in the introduction. The third method is the ST with no additional detection method, the text has been updated to reflect this.

4.1.1. Amplitude (A)

There is not a single reference in this subsection. Please provide some contextualization. The novel neighborhood method retrieves smaller wave amplitudes in the tropics than the

old amplitude-cutoff method. Is that closer to reality? Is that in better agreement to other studies?

Previous studies using AIRS have in general had extreme difficulty measuring waves at all in the tropics, largely due to the instrument noise issue. An example of this is Ern et al. (2016) where momentum fluxes are near-zero at the equator. Therefore, it is challenging to identify comparator studies.

4.1.2. Zonal Momentum Flux (MFx)

How is the wavelength and hence momentum flux derived in the case of the amplitude-cutoff method?

The amplitude cutoff method essentially creates a binary mask of wave/no-wave, which can be applied onto any of the spectral outputs of the ST, including onto the wavelengths. This has been clarified in the methods section.

Line 209: A bit irritating, maybe just say: "When the GW's horizontal phase speed equals the horizontal wind speed […]" and cite Lindzen (1981).

This has been updated.

Line 216: instead of "the Hindley study" use \citet{Hindley_reference}.

This has been updated.

Line 222-224: That is not clear. I understand that the 3DST samples discrete angles, however, the 2D+1ST computes the vertical wavelength from a phase shift. So, technically, an infinite vertical wavelength would be retrievable. What do you mean by "uncertainty bounds" in that context? Please elaborate. In case of very long vertical wavelengths, the Ern formula (equation 1) does not hold anymore since the wave's intrinsic frequency approaches N and the momentum flux drops to zero.

We agree with the reviewer that a zero-phase shift between two vertical levels would imply an infinite vertical wavelength. By "uncertainty bounds," we refer to phase shift errors propagating into vertical wavelength estimates. To avoid unphysical values, we constrain vertical wavelengths to the range 6–50 km, excluding cases where the phase shift is too small. Our inspection shows all retrieved waves fall within the mid-frequency regime, which we have included in the text, so the Ern formula remains valid, though we acknowledge the reviewer's point.

Line 226: GW activity is not a vector quantity and cannot be directed eastward/westward. Please change that.

This has been updated.

Line 227: Please change "the Hindley et al. (2020) study" to \citet{Hindley_reference}.

This has been updated.

Line 228: Please be more specific.

As mentioned in our General answer, this has been removed due to the updated results.

Line 231-233: Please change that to "The low percentage of data close to 0mPa can be attributed to the fact that the neighborhood method cannot identify waves with arbitrary small amplitudes due to the noise floor of AIRS. Another explanation might be that the waves mostly propagate in zonal direction and hence, k is never close to 0." (Although we will later find out that there is a bias problem with zonal wavenumbers.)

This has been updated.

4.1.3. Vertical Wavelength

Again, please provide contextualization.

There are very few studies using AIRS that investigate the vertical wavelengths of gravity waves, as this is a very difficult parameter to measure, as shown in Ern et al. 2016, so it is hard to find studies to compare against.

Line 237: Please change "curves" to "distributions".

This has been updated.

Line 241: These winds allow westward propagating or stationary GWs to propagate into the stratosphere. Their stratospheric penetration does not depend on their vertical wavelength.

This has been rephrased for clarity.

4.2. Regional Studies

Line 246: Misleading reference of Zhang et al. (2013). Better use Rapp et al. (2021) and Reichert et al. (2021).

This has been updated

Line 253-258: Erase everything except "Figure 6 shows the annual cycle of GW activity over each region." The other information is in the caption.

This is a stylistic choice, as mentioned above.

Line 261-263: Change to "This is due to the stratospheric wind reversal close to 20km in summer limiting the vertical propagation of orographic GWs."

This has been updated.

Line 272-273: It can be excluded that the signal is due to orographic GWs. But please elaborate on why more stable winds or the Southern Ocean storm belt are responsible for enhanced summertime lower stratospheric wave activity over New Zealand.

The stable winds allow a wider spectrum of GWs with different phase speeds to propagate through. This has been added for clarity.

Line 286: Use \citet{Hindley_reference}.

This has been updated.

Line 290-293: Erase except for "We next consider Figure 7, which shows polar histograms of wave phase propagation angle over the three regions during wintertime." The other information is given in the caption.

This is a stylistic choice, as above.

Line 294: Why do you expect propagation against the background wind?

The majority of waves seen in this paper can be attributed to orographic sources, which, by definition, propagate against the background wind as they have 0 ground-based phase speed (from Nappo 2012, pages 57 – 85).

Line 295: is "eastward" attributed to the wave propagation or the background wind?

The background wind, this has been updated.

Line 315-320: What a kick in the teeth. This issue should be mentioned earlier in the manuscript. Moreover, effects on the momentum flux retrieval should be discussed. Does this lead to an underestimation of momentum flux? Could this problem be resolved using only one granule at a time?

This bias has been resolved with our update to the processing chain and removed, as mentioned in our General answer above.

5. Conclusions

Line 324: I'd prefer "constant" or "homogenous" over "stable" in that context.

This has been updated.

Line 332: Be more specific. How large is the fraction? Give me a number.

We have calculated the number, and added it into the conclusions

Line 333: Be more specific. How much larger are wave amplitudes on average?

This has been removed since updating the conclusions to be in line with the updated results.

Line 335: The color scale ranges from -1 to +1 Kelvin. I cannot follow where the maximum amplitude difference is 3K.

This has also been removed for the same reason as above.

Line 346: What do you mean by "reliable ST output"?

This has been rephrased for clarity

Response to Reviewer 2:

General Answer:

Thank you very much for your helpful comments; they have made this paper much more readable and have improved the scientific results. We have updated the structure of the introduction and methods sections to introduce the methods used here earlier and more logically. We have also updated our results by applying the masks from the neighbourhood method onto the smoothed input data. This makes it more comparable to other studies that have used this technique and removes the bias from line 335.

As mentioned above, our changes involved smoothing the input data onto which the binary masks are applied. This was a critical step for ensuring comparability with Hindley 2020 and was mistakenly omitted from the initial draft. Implementing this adjustment has resulted in meaningful changes in some results, including sharper directional distributions and an improved amplitude cutoff detection method. We believe our method still outperforms this other method, thus significantly strengthening our argument. We thank the reviewer for their diligent comments, which helped us identify this issue.

In response to the reviewer's point about no-wave events, in this study, our primary aim was to characterise the properties of detected gravity waves rather than develop a basis for parameterisation. So, we chose to focus our results and statistics on detected waves. This technique could be used for further study along this line, but it was outside the scope of this paper. We agree this would be an excellent avenue for further work aimed at larger population studies, including parameterisations or assessing momentum flux budgets.

General Comment:

This paper proposes a method for improved noise removal and GW identification from satellite observations, by utilizing the spatial homogeneity of spectral wave characteristics. This method enables the detection of small-amplitude waves that were previously excluded by the existing amplitude-cutoff method to avoid noise contamination. The identified waves also exhibit a realistic seasonality with asymmetry in the direction of zonal momentum flux, which is highly encouraging. I recommend this paper for publication after revision, considering the following issues.

Discrepancy with the literature. The advantages of the proposed method are demonstrated through a comparison with the results from the existing amplitude-cutoff method presented together. However, these results appear largely unrealistic in terms of phase direction. This contrasts with previous studies that have employed similar methods and reported more reasonable results. For more details, please see Specific comments on Figs. 4 and 7 below.

Statistics including no-wave events. It would be valuable to assess the amplitude and momentum flux averaged over all observations including no-wave events (with zero amplitude and momentum flux where waves are absent). However, the results presented in this study are averages only over detected cases. This approach limits its applicability to GW parametrization and makes it less straightforward to evaluate the GW impact on circulation. I recommend including results that incorporate no-wave events. Additionally (and optionally), including no-wave events in the statistics could provide valuable insights into wave intermittency, particularly since the proposed method effectively remove noises without artificially truncating all small-amplitude signals.

There are several instances where the explanations of the presented results lack clarity (see Specific comments below). These should be clarified to ensure better understanding of the findings.

L9-10: Examining the GW impact requires the results integrated over all observations including no-wave events (see General comment #2). In the following sentence (L10-11), it is stated that small-amplitude waves can be additionally detected by the new method, which partly supports the argument in L9-10; however, how much noises were falsely detected as waves in the existing method (leading to an overestimation of GW impact) has not been presented in the current manuscript.

We agree that GW impact cannot be presented in this study, and so that line has been rephrased. Since reprocessing the data, we have changed the abstract to reflect the updated results, but have kept this point in mind while writing.

L13 "consistent with surface-levels winds": I found several places in the main text where the stratospheric winds are discussed, but could not find a discussion with surface-level winds.

This has been updated and removed.

Fig. 1: The unmasked regions in panel (d) are slightly larger than those in (c). Has the neighbourhood been unmasked in (d) ? If so, this information should be included somewhere around L141. If this is not the case, please provide the reason for the different sizes.

The final mask had been slightly smoothed, which enlarged it slightly. The methodology now specifies this.

L132: It should be clarified whether the absolute difference is in wavenumber vectors or in each of k and l.

This clarification has been added.

L135 "C": I would expect that a more broadly effective tolerance might be relative to the horizontal wavenumber (e.g., C ~ kh / 10, or so) rather than a constant for any waves. Is this constant approach because AIRS-detectable GWs have a limited range of wavelength spectrum ?

In this proof-of-concept study, we aimed to see if this detection method would work. A relative tolerance has been discussed, and it will be applied in future work using this method.

Eq. (1): This seems to omit the minus sign on the right-hand side. (m < 0 in the authors' convention: L114)

This has been updated.

L172: From here onward, the detected GW activity is often discussed in relation to the polar jet. At least once here, please provide a brief description of how the GW detection is linked to the jet. Does this refer to upward refraction by the jet, which increases vertical wavelengths, making them more detectable by AIRS ?

Yes, as seen in Wright et al. (2015). A brief section has been added explaining this.

Fig. 4: It is very encouraging to see the realistic asymmetry in the direction of zonal momentum flux derived using the new method. On the other hand, in the results of the amplitude-cutoff method, it is surprising to see that eastward/westward momentum fluxes are rather symmetric in the extratropics. This seems to contradict previous studies that have shown the realistic direction of zonal momentum flux using the amplitude-cutoff method (e.g., Hindley et al., 2020). What am I missing ? In the authors' work, the advantages of the new method are demonstrated by comparing it with the amplitude-cutoff method shown together. If the latter is not fully representative of others' results in the literature in some way, that should be explicitly stated.

As mentioned in our General answer above, we have reprocessed our results to better compare with the other literature, such as Hindley et al. 2020. These updated results are now much better aligned with those found in others' work, so we are more confident in our own results now.

L203-204 "during austral winter, …": What is discussed here seems to be true in both winters.

This has been updated.

L214: Please clarify the levels: near-surface or stratospheric winds ?

This has been updated, stratospheric winds.

L229: (1) Because the period of 2010–2014 fully covers two QBO cycles, I would not expect to see the QBO effect in the statistics integrated over this period if this effect simply refers to QBO-phase dependence. Or, if it refers to another aspect (e.g., phase asymmetry or some seasonality?) of the QBO, this should be clearly described to justify the attribution to short vertical wavelengths. (2) Again, it was nice to see the realistic asymmetry in the direction of momentum flux with seasonal dependence. Therefore, if it is also possible to include a result showing GW responses at ~39 km (or any other level) to the QBO, that would be great. Comparing any given month in 2012 to the same month in 2013 (or, even the 2012 annual mean to the 2013 mean) might be interesting, as the months in 2012 are dominated by stronger easterlies in the QBO domain, compared to those in 2013. However, this suggestion is optional, as such a result could be subject to uncertainty due to the limited statistics.

We have removed the sentence about the QBO to reduce confusion, but we agree that it would be very interesting to look into these data. We believe this is out of scope for this paper, but a follow-up paper could investigate this thoroughly.

L239-241: There is an inconsistency in logic between these two sentences. The first sentence attributes the long vertical wavelengths to the strong winds. However, the second sentence suggests that such waves with long vertical wavelengths already exist and they need that favorable wind condition (strong winds) to reach the stratosphere. Please re-write these sentences.

These have been corrected and updated.

Fig. 6 caption "Histograms for each day": Does this mean daily-mean values counted into monthly bins or shorter-term (instantaneous?) values into daily bins ?

These are monthly-moving means for each day of data, this has been clarified in the text.

L263: "This also indicates that these regions have similar drivers of GWs": This argument seems to be logically weak, as different sources could exhibit similar seasonality (e.g., being more active in winter). Additional evidence or reference would be needed to support the claim that the GW sources are similar across these regions.

This statement has been weakened, with references added to support it.

L268-270 "On average, ...": This information, which refers to the average rather than local or seasonal features, would align better with Fig. 4. I suggest moving it to Sect. 4.1.

This should refer to the local features, and has been rewritten for clarity.

L273: While the cited reference (Chapman et al., 2015) explained the interaction of the Southern Ocean storms with topography, I could not clearly find an explanation for why such an interaction is weaker with the Rocky Mountains.

We have added a reference (Shaw et al., 2022) to clarify that the Northern storm belt is weaker than the Southern storm belt.

L278 “as discussed in Figure 4”: I though that the lack of near-zero flux discussed in Fig. 4 (L231-233) was for the results using the neighbourhood method (where very small amplitude noises were filtered). However, the current paragraph explains the amplitude-cutoff results. The lack of near-zero flux in the amplitude-cutoff results shown in Fig. 4 (although I am not sure whether it was discussed) may be due simply to the cutoff. Consider clarify or adjust the text to avoid potential confusion.

We agree this could cause confusion, so have clarified both by Figure 4 and in this paragraph.

L284-287: (1) Logically, the use of a cutoff cannot be the reason for detecting more waves of low amplitudes (as it cuts off). The neighbourhood method filtered the low-amplitude noises, leading to relatively larger mean fluxes. (2) I guess MFx in this line refers to the net zonal momentum flux, as this is the quantity presented in Hindley et al. (2020). The authors have showed the predominance of westward momentum fluxes over eastward momentum fluxes in winter (e.g., ~88% over the Rocky Mountains), while this was not the case for the amplitude-cutoff based results (please also see my comment on Fig. 4 above). If this was also the case in Hindley et al., where the amplitude-cutoff method was used, the cancellation between the nearly symmetric eastward and westward fluxes could be the primary reason for the 10 times smaller MFx.

As mentioned in our General answer above, we have reprocessed our results, which has led to a rephrasing of the first part of this comment, and the second part is no longer applicable, as the results from the amplitude cutoff method are now more comparable to those found in Hindley et al.

L289 “lower proportion of low value MFx”: Does this mean that the low-flux part is relatively smaller in this region than in the other regions when the eastward-momentum fluxes in the summer season is scaled (normalized) in each region ?

This has now been rephrased since reprocessing the results.

Fig. 7: In the amplitude-cutoff results, the meridional propagation is surprisingly predominant over zonal propagation. In L315-320, this is explained as a bias due to several factors, including the across-track fourth-order polynomial filtering. This factor is said to be a standard. Does this mean that results of previous studies in the literature were also affected by this bias ? I am confused because in those studies, the zonal flux was found to be much larger than the meridional flux (e.g., Fig. 3 in Hindley et al., 2020).

Fixing this bias was one of the main reasons for reprocessing the data, and as such, it has now been fixed. The majority of the flux is in the zonal direction.

L306: How can the African easterly jet account for the westward propagating waves ? (In many parts of the manuscript, observed waves have been thought to propagate against the winds.)

This has now been removed since reprocessing the data.

L340 "~40× greater": This specific information has not been mentioned in the Result section.

This result has been removed since gaining updated results from the reprocessed data.

L343-345 "… consistent …" / "and consistent …": (1) The former part seems to be repetitive if it refers to the global morphology of flux directionality, which was already mentioned in L341-342. If it refers to another aspect that is consistent with the linear theory, please specify it. (2) For the latter part ("and ..."), its meaning is unclear. Please rewrite this.

This bullet point has been removed and incorporated into the other bullet points.

Figs. A1-2: I would suggest showing the masked wave fields (like those in Fig. 1d) rather than just the masks.

We believe that showing only the masks makes it clearer as to what is changing from panel to panel.

[Technical comments]

L7-8: "defined": This would be unnecessary.

This has been deleted.

L20: pressure → density ? (The latter may be the most directly relevant variable.)

This has been updated.

L28: hours → minutes ?

This has been updated.

L45-46: The citation of these three references should be placed at the end of the sentence.

These references have been moved.

L46: "with with"

This has been updated.

L48-50: Please rephrase the sentences. For example, "a wave is present at all locations in the data" → "any signal in the data is projected to a wave" ?; "assign low amplitudes":

Please provide the reason (noise?); "which is why …": I could not understand this logically in the context.

This section has been rewritten for clarity, and in accordance with Reviewer 1.

L53: "and as such, a method ..." (comma) ?

This has been updated.

L58: stable spatial → spatially stable ?

This has been changed to "constant spatial", in accordance with Reviewer 1.

L94: "large": Please clarify if this is for the spatial scale or amplitude.

This has been clarified as spatially.

L118: "other levels" commonly stated in the two consecutive sentences seem to indicate different groups of levels: for the former, the levels with similar resolutions and noise magnitudes to 39 km; and the latter, any levels with different characteristics. If this is correct, I suggest avoiding repetition of the same term "other levels" for different objects.

This has been removed.

L119: "variables within the neighbourhood method": This method has not yet been introduced (except being mentioned in the abstract), so discussing some variables used in the method may be premature at this point.

This has been removed, and the whole methods section restructured to fix this.

L124: wave properties → spectral properties (as this is for noise)

This has been updated.

L142-143: The difference cutoff has been repetitively mentioned here (L142; L143). Please rephrase the sentences.

This paragraph has been removed, and incorporated into an earlier point.

L154: amplitude → temperature amplitude

This has been updated.

L184-185: "but they each hotspot … in Figure 3.": Please rewrite the sentence.

This has been updated.

Fig. 4 caption: "that the axis for vertical wavelength (c)"

This has been updated.

L200: data → results ?

This has been updated.

L211: break → dissipate (as wave breaking often refers to a special process, not related to the critical level)

This has been updated.

L226: “same”: I agree that this result is consistent with Hindley et al. (2020, Fig. 2b). But I suggest rephrase this word, as they presented (net) zonal momentum flux. The same pattern as in Fig. 4b cannot be observed (but inferred) there.

This has been updated for clarity

L239: “from 90°N to 90°S”: Please revise this. Waves were not observed around 90°S in boreal winter (Fig. 3a).

This has been revised.

L277: and → an

This has been updated.

L319 “… is the standard, identifying … remains …”: (1) no conjunction for the two sentences. (2) References may be necessary for the standard.

This paragraph has been removed since the reprocessing mentioned in our General answer.

L353: average → wavenumber (the latter may be more informative)

This has been updated.

L356 “… have no extraneous noise” / “or that no noise is …”: Do these two have different meanings ?

We believe they do, but we have rephrased this for clarity.

Response to Reviewer 3:

General Answer:

Thank you very much for your helpful comments; they have made this paper much more readable and have improved the scientific results. We have updated the structure of the introduction and methods sections to introduce the methods used here earlier and more logically. We have also updated our results by applying the masks from the neighbourhood method onto the smoothed input data. This makes it more comparable to other studies that have used this technique and removes the bias from line 335.

As mentioned above, our changes involved smoothing the input data onto which the binary masks are applied. This was a critical step for ensuring comparability with Hindley 2020 and was mistakenly omitted from the initial draft. Implementing this adjustment has resulted in meaningful changes in some results, including sharper directional distributions and an improved amplitude cutoff detection method. We believe our method still outperforms this other method, thus significantly strengthening our argument. We thank the reviewer for their diligent comments, which helped us identify this issue.

General comments

This is an interesting analysis and study, meeting the publication requirements of novelty and significance. However, I concur with previous reviewers that the manuscript needs to be better organized and the writing needs polishing. Additional comments are discussed in detail below in the specific comments below

More fundamentally, the manuscript needs a more complete discussion of why the two methods applied in this study often give wildly different results. At one point a discrepancy of a factor of 40 is mentioned. There and elsewhere, large differences are simply stated or briefly explained in a sentence or two. Explaining these differences systematically should be THE fundamental goal of the paper in my view. Note the word "Method" in title; clearly critiquing the methodologies discuss here will put the results in better context and help bolster the credibility of the new S-transform method being described. The Conclusions section would be the appropriate place to summarize results from the different methodologies.

The manuscript would also be improved by more careful use of terms for methodology. For example, the term "neighbourhood method" is used five times in the abstract, implying this is the novelty mentioned in the study title. However, when "neighbourhood" is used for the first time in the main text near the beginning of Section 3.3, line 120, the Wright et al. (2017) study is cited, implying the method is not novel. This discussion also follows the S-transform discussion around line 45, and further S-transform descriptions in Section 3.2. Line 55 even states that the S-transform is the new method. Only in the first sentence of the Conclusion is the methodology concisely described. **A sentence or two like that in the Abstract and early in the manuscript would be very helpful.** The details of the methods should be discussed in the Methods section rather than scattered through the text.

The abstract needs to be expanded to include more quantitative results and clearly state some of the discrepancies between the two methods considered, as mentioned above.

Line 35 forward. Is this a description of the amplitude cutoff methods? If so, please mention explicitly. Even more helpful would be a brief discussion here, then provide a

more detailed description to Section 3. Having all the methods described carefully in one place will make a much more readable paper.

This section has been rewritten to more accurately capture the past techniques for detecting gravity waves in AIRS. A brief discussion of the techniques has now been added in this section, and a more structured description of the methods has been created in Section 3, which we believe creates a much more readable paper, as you said in your comment.

Line 45. The discussion of wavelets and gaussian window should be moved to Section 3.2 since they are fundamental to the method. See earlier comment about a concise description of the method being needed before Conclusions.

A very brief description of the S-transform is written here, and has been modified according to Reviewer 1, and then expanded on in the methods section.

Line 47. Should "frequency" be "wavenumber"? Frequency implies time information.

Yes, this has been updated.

Line 49. The sentence essentially says that no. waves detected by the ST method is why the amplitude threshold method works. Please re-write.

This was commented on by Reviewer 1 as well, and has been rewritten in accordance with both of yours' comments.

Line 63. Delete "specifically" as the items are listed very specifically.

This has been deleted.

Section 2. This would be an opportunity to discuss the considerable gravity wave literature supported by the AIRS data. This field of study is noteworthy since GW studies were not anticipated to be important at the beginning of the AIRS mission.

We believe this would be relevant when introducing AIRS as an instrument for analysing GWs in Section 1, and as such many studies have been referred to that utilise AIRS for GW research there.

Line 73. Mention here that a granule is based on 6 minutes of time. That's buried elsewhere in the manuscript. Also, please give the along-track dimension of a granule. The scan width is mentioned on line 69, but only the cognoscenti will know that the along track scale of a granule is 13.5 x 135 km.

The duration and length of a granule has been added, with more explicit references to the dimensions and scales along-track added too.

Line 79. How is vertical resolution defined? Does this mean that vertical wavelengths of ~7 km can be detected, or that points 7 km apart are independent of one another, or something in between? Line 111 makes this clear, but somewhere around line 79 is the

place to be explicit. A summary of key findings of Hindley et al. (2019) might be helpful, but only a sentence or two.

Content has been added to make the vertical resolution more clearly defined in this context.

Line 89. What are the spatial dimensions of the 128x135 pixel region? Something like "128x135 pixels (nnnn x mmmm km)" is short but helpful.

This has been added.

Line 94. As mentioned by another reviewer, the polynomial filter has important implications for later results. This is a good place to mention that this filter may be removing meaningful information. Something like "As will be seen in Section 4.2 below, this filter appears to be removing zonally aligned GW."

As mentioned in our General answer, our reprocessing of the data has removed this bias, so we do believe this must be mentioned.

Line 103. As a first step, this section should clearly define the ST methodology in an opening paragraph. Earlier details given around line 45 should be included here. This paragraph should also explicitly describe the amplitude cutoff method, which is implied in the discussions around line 40, but never described completely; it should also be included in the section title. Then, and only then, should the application of the method be described. Note that the section starts "We then apply…" and the application theme continues throughout. The current structure makes it difficult to discern the actual methods.

We have restructured the methods section so that it is more clear where the novelty lies, along with a better description of the ST, and new subsections for the two detection methods used. Brief descriptions have been added in the introduction section of what will occur later, which adds clarity.

Line 115. The 39 km level should be mentioned in the figure captions where appropriate.

These have been added.

Line 127. The text states "pairs of granules" but Figure 1 shows three granules.

These granules were chosen as good visual examples of how the neighbourhood method could be used.

Line 135. As discussed in Appendix A2?

This has been updated.

Line 142. The discussion is fragmented. The references to the Appendices should be within the appropriate variable discussions. This paragraph should be deleted and the relevant parts incorporated into the itemized discussions immediately above.

This paragraph has been deleted and incorporated into the bullet points, and been edited for clarity.

Line 153. How is the Brunt-Vaisala frequency defined?

The Brunt-Vaisala frequency has been assumed to be 0.02s^-1.

Figure 3 caption. Is this at 39 km, as mentioned in the text around line 115?

It is, this has been updated.

Line 167. Replace "only detects GWs" to "only detects sinusoids". We interpret those as GW.

This has been updated.

Figure 3 caption. What altitude is this?

This has been updated to include the altitude.

Figure 4 caption. Altitude is important here because vertical wavelength may be misconstrued as a property of the entire height of coverage rather a local property derived by the S-transform. Also, the inverse of wavelength is linear in wavenumber. It's not obvious why either version of that final sentence is needed, so it can be deleted.

The altitude has been added, and the final sentence has been updated. This has been kept as it makes clear why the axes have unevenly spaced ticks.

Line 196. Rather than "amplitudes" it should read "normalized occurrence frequency of amplitude". Better still, put the clear definition in the caption and use the text to discuss result. If the caption is clear, then vaguer terms like 'histogram' are fine in the text. The converse of vague terms in the captions that are clearly defined (somewhere) in the text is a recipe for confusion. Please note that this reviewer does not have a clear understanding of what is being plotted. Histograms of counts? Histograms of occurrence per granule within latitude bands? Histograms of normalized counts is the best guess, but only a guess

This has been changed in the text, and the caption has been updated to clarify what is being plotted.

Line 215. The ~10x discrepancy needs further explanation and should include a discussion of equation (1). That equation suggests two possible explanations for the large disagreement: differences in A (or its square), and/or, differences in k/m. This study presents good evidence that the neighbourhood and amplitude threshold methods have roughly the same sensitivity for large amplitude waves. Some questions arise: Are the waves dominating zonal momentum flux of large amplitude? Are smaller waves magnified by k/m more important? Are the two methods equally sensitive to k/m? It would be very helpful if these issues were thought through and discussed by the authors in an updated

manuscript. A paragraph discussing these sensitivity issues would greatly strengthen the study and help support the geophysical plausibility arguments presented elsewhere. Such a discussion should also be included in the Conclusions.

As discussed in our General answer above, we have since reprocessed the data and gained more comparable results with the literature, which has lead to our results being slightly different, including the 10x discrepancy now being incorrect.

We agree that such an addition about sensitivity would strengthen the study, and as such have created a new figure that discusses the sensitivity of both methods to amplitude and k/m.

Line 228. Once again, why the large differences?

Since reprocessing the data, this has been removed.

Line 238. Should "peak location" be "dominate wavelength"? Location implies a point in physical space.

This has been updated.

Line 247. Delete "in this region" since the region was just stated.

This has been updated.

Line 261. Better to use "This is consistent with…" rather than "This is due to…" since causality hasn't been demonstrated.

This has been updated.

Figure 6 caption. Please be more specific about what is being shown. The color bar label of Normalized Counts makes this pretty clear, but it's ok to repeat the definition. Changing the start of the caption to "Histograms of normalized counts per day…" is an easy fix.

This is a very good point, and has been updated.

Line 268 forward. These differences in detection frequency are not unexpected and should be part of an overall discussion/summary of the strengths and weaknesses of the methods used.

We agree, and have summarised these points in the conclusion.

Line 285. As discuss earlier, is this difference entirely due to amplitude effect or does sensitivity to certain wavelengths also matter? The connection to other discrepancies seen in Figure 4 should be noted and discussed in the Conclusions.

The results have been updated since the reprocessing, and are now more comparable. The differences have now been summarised in the conclusion.

Line 297. Panels (a-c) show clear northward and southward directional preference, with very few zonally propagating waves. The lack of the latter is discussed around line 315. However, the text on Line 297 claims a "nearly uniform spread". This should be corrected or explained in more detail.

As mentioned in our General answer, we have reprocessed the data, from which the resulting angles are much more reasonable. We have also updated the text to discuss these findings.

Line 315 forward. This important result is worth discussing in more detail in the conclusions.

This bias has been fixed by reprocessing the data, as mentioned in our General answer.

Section 5. The word "Method" is used in the title. Consequently, a more complete critique of methods is needed.

Another reviewer suggested a table comparing the methods. This could be taken further to include a table broadly comparing the results. How best to organize these ideas is up to the authors, but a comparison of results is needed.

Also, the interpretations presented in the bullets are of two basic types. One type is essentially methodological and would include sensitivity thresholds, resolution, etc., as in the first two bullets on the bottom of p. 15. The second type of interpretation is based on geophysical plausibility. Examples are the remaining bullets in the Conclusions.

Consciously retaining these distinctions between methodology and physical plausibility should help structure the discussion.

We believe that Figure 4 succinctly discusses the differences between the results, and have added a figure that discusses the differences in the sensitivities. Since reprocessing the data, we have changed the conclusions and kept this point in mind when rewriting. We agree that this improves the study.

Line 170. A statement about sensitivity differences should be included in a more complete discussion the Conclusions.

We have added a figure in the results and discussions that discusses the sensitivity and another point in the conclusions.

Technical comments

Line 184. Delete "they". Apparent typo.

Updated.

Line 209. Change "GWs" to "a GW" since the sentence shifts from plural to singular.

Updated.

Line 223. "If it is possible to cut a word out, always cut it out." -George Orwell. Following this advice, please change "…assigned to be angled in some direction relative to the vertical, and if the vertical angle assigned is…" to "…assigned an angle relative to the vertical, and if that angle is…".

This is good advice, updated.

Line 281. Start the sentence "Approximately" rather than with using its abbreviation "~".

This has been updated.

Line 351. Currently reads "Appendix A: Appendix".

This has been updated.