# Authors' response to comments from Reviewer #3 on "Stratospheric gravity waves over the mountainous island of South Georgia: testing a high-resolution dynamical model with 3-D satellite observations and radiosondes"

N. P. Hindley et al.

## General Comment for all Reviewers

We would like to thank the reviewers for the hard work in preparing their reviews of our submission. Their helpful suggestions have significantly improved the study. Several main improvements are listed below:

- In response to the reviewers' comments, we have significantly improved the way the model is sampled to create the model-as-AIRS dataset in our study. We realised that it is not enough to simply apply the AIRS horizontal resolution to the model: the AIRS horizontal sampling must be considered too. By sampling the model on the AIRS horizontal grid and taking into account the different sampling locations of each overpass, we are able to remove the background temperatures in exactly the same way for the AIRS and model-as-AIRS temperatures (we no longer use the nSG model runs for this). This ensures that our analysis steps allow for the spectral range of GWs visible to the AIRS and model-as-AIRS to be consistent.
- We also apply specified AIRS retrieval noise to the model-as-AIRS, which is characterised from a realistic AIRS granule. By applying the noise to the model-as-AIRS, we can separate out the effects of retrieval noise. This is important for the area-averaged results upwind and downwind of South Georgia.
- We now keep GW results measured in the full-resolution model very separate from the comparison between the AIRS and the model-as-AIRS. GW momentum flux in the full-resolution model is now calculated using wind perturbations, rather than from down-sampled temperature perturbations as before, and no comparison is made between GWMF in the model and the model-as-AIRS. This an important distinction because it is not possible to apply consistent horizontal sampling and background removal methods to both datasets, so no fair comparison can be made.
- The above steps have greatly improved the agreement between the AIRS GW measurements and the model-as-AIRS. As a result, the paper has been substantially reduced in size from 16 figures to 11 with a $\sim$20% reduction in text. Inconclusive or superfluous results and discussions have been removed, and a new Fig. 11 showing a case study of a short-$\lambda_H$ GW event has been added.

## Response to Reviewer #3: Major Comments

1. *L169-176: The vertical resolution applied in the model is extremely coarse related to the horizontal resolution. The vertical grid-spacing is 0.6-2 km in the stratosphere, versus a horizontal grid-spacing of 1.5 km. This vertical grid-spacing in the stratosphere in not even sufficient to simulate a self-induced QBO in GCMs. More importantly, GCMs with explicit simulation of GWs (e.g., Watanabe et al., 2008, JAS: General aspects of a T213L256 middle atmosphere general circulation model) employ a vertical level spacing of 300-600 m throughout the middle atmosphere while the resolvable horizontal wavelengths in these models are of the order of 200 km. The necessity for a small enough vertical grid-spacing derives from the fact that the GWs resolved by the horizontal grid must not be spectrally biased in the*

*vertical to too large vertical wavelengths. Indeed a too coarse vertical resolution artificially prevents the GWs from reaching dynamic or convective instability and thus being dissipating by the model's turbulent diffusion scheme.*

This is probably the reviewer's main point, so we will break down our response to it below. While we agree and acknowledge that high vertical resolution is very important for accurate GW modelling in the stratosphere, we argue that the vertical and horizontal grids used in our model are more than sufficient to accurately resolve stratospheric GWs over South Georgia.

Sensitivity tests for vertical grids of 70, 118 and 173 vertical levels were performed by Vosper (2015). They found that the resolved zonal GW momentum fluxes from the surface to altitudes near 40 km for the 118 and 173 level simulations were highly similar. Both of these exhibited more realistic GWMF values in the lower stratosphere than the 70 level simulation. As can be seen from Fig. 2b of Vosper (2015), the 70 level simulation exhibits increased GWMF above 25 km. This is highly indicative of the issues relating to coarse vertical grids that the reviewer highlights here. Very little further improvement was found going from 118 vertical levels to 173, so the 118 level configuration was selected to reduce the computational load and permit the use of a fine horizontal grid over the island. We apologise that we neglected to mention this explicitly in the original submission. This is updated in the revised manuscript.

It is clear from our study however that a high horizontal spatial resolution is essential for accurate simulation of orographic GWs from the small mountainous island of South Georgia. It would always be nice to have more vertical levels, but we are limited by computational resources of what was feasible when the model was run.

*"This vertical grid-spacing in the stratosphere in not even sufficient to simulate a self-induced QBO in GCMs."*

The reviewer's comment here about the QBO is not relevant for our study. We do not need to simulate a realistic QBO over South Georgia near 54°S because (a) it is a primarily a tropical phenomenon and (b) we are only considering month-long time periods whereas the QBO has periods near two years.

It is worth mentioning however that the Kanto model of Watanabe et al. (2008) that the reviewer mentions did indeed resolve a QBO signal, but the period was close to 15 months rather than 28 months. Clearly this is far from perfect, despite the high number of vertical levels.

Ideally of course, one would always have many more vertical levels, but our point is that the 118 vertical levels used here are more than sufficient to resolve a realistic orographic GW field over South Georgia (Vosper, 2015). Once the correct horizontal sampling and resolutions were applied to the model (see revised paper), we actually found the agreement between the model and observations to be quite good. This suggests that the vertical grid issues described by the reviewer do not significantly affect our results.

*More importantly, GCMs with explicit simulation of GWs (e.g., Watanabe et al., 2008) employ a vertical level spacing of 300-600 m throughout the middle atmosphere while the resolvable horizontal wavelengths in these models are of the order of 200 km.*

Despite their fine vertical grids, such models with coarse horizontal grids cannot be used for our study. The reviewer mentions the Kanto model of Watanabe et al. (2008), and later the model of Becker and Vadas (2018). The orography of South Georgia would be, at most, equivalent to one or two horizontal grid points in these spectral models with T213 and T240 respectively, if even resolved at all. Despite their high number of vertical levels (which is very good), they would be unable to realistically resolve orographic GW generation and propagation over the island at the short horizontal scales necessary for our study.

In the vertical, the global models mentioned actually have quite coarse vertical grid spacing in the troposphere, which is a problem for accurate GW simulation over orography. The Becker and Vadas (2018) model for example has a vertical grid spacing of 600m and doesn't even go down to the surface, instead stopping at the boundary layer. The local-area model used here has a 10m vertical grid at the surface, and only begins to exceed 600m above 20 km altitude. This grid configuration is more practical for low-level wind flow over orography and realistic mountain wave generation and propagation.

The local-area model configuration of the UM used here, with a horizontal grid of 1.5 km, will out-perform the resolution of these global spectral models, which is a key benefit of non-spectral models because local refinement is possible. The horizontal grid used is around 20 times finer than the global spectral models the reviewer mentioned, so if we wanted the same ratio between horizontal and vertical grids as these global models (to avoid spectral biasing), we would have to increase our number of stratospheric vertical levels by a similar amount. This is clearly impractical, and well beyond what was feasible when these simulations were performed due to computational limitations. The model runs used here are computed on a $800 \times 600 \times 118$ grid. A simultaneous run on a 750 m horizontal grid was also performed on a $1600 \times 1200 \times 118$ grid. These runs were highly computationally intensive. An increased number of vertical levels in the stratosphere would of course be advantageous, but the trade-off here is necessary to investigate the effect of fine horizontal grids, while remaining practical to run.

We acknowledge that no model is perfect, but some are useful. The sensitivity tests and assessment of simulated GWs in previous studies demonstrate that our chosen configuration can be useful for our study of mountain waves over South Georgia (Vosper, 2015; Vosper et al., 2016; Jackson et al., 2018).

*The necessity for a small enough vertical grid-spacing derives from the fact that the GWs resolved by the horizontal grid must not be spectrally biased in the vertical to too large vertical wavelengths.*

We agree with this comment, but we believe that it is not relevant for our study. Of course, a high vertical resolution is important for all GWs, but it is especially important for inertia GWs with relatively short vertical wavelengths.

However, our study is focused on stratospheric mountain waves over South Georgia during winter, where strong zonal winds at southern high latitudes can refract GWs to relatively long vertical wavelengths in the stratosphere (e.g. $\lambda_H \sim$12–25 km for zonal winds 40–80 m/s). These GWs can also have short horizontal scales $\lambda_H \lesssim$50–100 km. The aspect ratio of these GWs is far from those of inertia GWs, and can be considered mid-frequency or perhaps even high-mid frequency GWs.

One could easily argue the reviewer's point but for horizontal resolution in the Kanto model or the Becker and Vadas (2018) model. In those models, waves from small sources (like South Georgia) will be spectrally biased to long horizontal wavelengths because the horizontal grid is too coarse to accurately simulate them. Here, we accept any limitations of our model grid and have explicitly discussed the reviewer's concerns in the revised manuscript.

*Indeed a too coarse vertical resolution artificially prevents the GWs from reaching dynamic or convective instability and thus being dissipating by the model's turbulent diffusion scheme.*

We agree with this point and have added this into the revised manuscript.

2. *L176-178: I do not find this statement very conclusive. The grid-spacing of a model as such does not say anything about the scales that are reliably resolved. It is the dynamical core (spatial resolution, numerics) combined with the subgrid-scale diffusion (either explicit or implicit) that determines the reliable scales of a model.*

   We agree, this was phrased badly. This has been corrected in the revised manuscript.

3. *L193-196: See my 2 previous major comments and consider reformulation.*

   See our responses above. The paper has been significantly revised to make this clearer.

4. *L137-348: When the model data are interpolated to a 15 km grid, the Fourier components with horizontal wavelengths shorter than 30 km must be filtered out beforehand to avoid aliasing errors from the scales below the 15 km grid. Did the authors apply this spectral filtering before re-griding the model data (for model and model-as-AIRS)? If yes, please mention this point in the text for the sake of clarity. If not, the resulting aliasing could be an explanation for*

*the high power in the GW amplitudes and in the MF at horizontal wavelengths of 30-40 km (e.g. Fig. 16a). In that case you might consider a substantial revision and re-submission of the paper.*

Firstly, we must apologise. As the reviewer correctly suspected, Fig. 16 had an error with the normalisation of the amplitude-horizontal wavelength bin widths (which were different sizes) which caused anomalously high power at short $\lambda_H$. When the bins are correctly normalised for their width, this anomalous high power is removed. We are grateful to the reviewer for spotting this. The figure has been correctly revised, but in the end we decided not to include it in the revised paper for brevity. Also, as a result of the new model-as-AIRS processing mentioned above, the full-resolution model cannot be fairly compared to the AIRS and model-as-AIRS.

Perhaps more importantly however, the reviewer's comment made us think about the effect of horizontal sampling. We realised that it is not enough to simply apply the AIRS horizontal resolution, but the horizontal sampling pattern must also be applied to the model to ensure a fair comparison. This led to a major overhaul of the model-as-AIRS processing to accommodate realistic horizontal sampling of the AIRS instrument, including the different sampling locations during different overpasses. Once this aspect was correctly applied, the agreement between AIRS and the model-as-AIRS was significantly improved (see revised paper). This is major improvement, and we are grateful to the reviewer for highlighting it - even though that maybe was not their intention!

With this is mind however (although this is not relevant any more), the reviewer is not correct that Fourier components with horizontal wavelengths shorter than 30km should be filtered out here to avoid aliasing problems. When AIRS samples the atmosphere, the horizontal sampling pattern samples where it samples, including any effects of aliasing. There is no post-hoc removal of Fourier components in AIRS when it samples the real atmosphere, so it would be inconsistent to apply such things to the model. For a fair comparison, we should simply sample at the same locations as AIRS and allow aliasing effects to take their course in both datasets.

Finally, we would also like to direct the reviewer to the new Fig. 11, where GWs with very large amplitudes with $\lambda_H$ ~30–40 km are observed in AIRS and simulated in the model directly over the island. These waves can only be resolved in the model due to the fine horizontal resolution, and in this example they are essentially validated by AIRS observations due to favourable viewing geometry.

Even after the original Fig. 16 was fixed, the figure did not make the cut for the revised manuscript, because its results were not very useful. Instead, the new Fig. 11 shows that although the high power at $\lambda_H$ ~30–40 km in the original Fig. 16 was in error, mountain waves with large amplitudes can be found at these short scales directly over the island, if the resolution is high enough to support them.

We should also mention that the improved sampling approach has also benefited the AIRS results. In the original submission, we applied a 3×3 horizontal boxcar filter to the AIRS data to suppress any spurious pixel-scale noise, as per the approach of previous studies (e.g. Wright et al., 2017). But after close inspection we found that some GWs directly over the island at the pixel-scale in AIRS were actually realistic (see new Fig. 11), so it was a mistake to smooth these out. In the revised paper, we do not apply this which results in a much improved agreement between GWMF in AIRS and the model-as-AIRS directly over the island.

5. *L399: The authors should not only mention that model-as-AIRS produces too small amplitudes compared to AIRS, but also that the GW-phases of the MWs over the Island differ significantly in the two data sets (Figs. 4 and 5). Moreover, the slopes of the phase lines from x=100 km to 600 km in Fig. 4 differ in sign(!); that is, these GWs must propagate in different directions when comparing model-as-AIRS to AIRS. Please mention and discuss these dissimilarities.*

   Figures 4 and 5 show different overpass times 14 hours apart. This was stated in the caption, but we have clarified it in the main text to help to make this clearer. Also, the new revised figure (now Fig. 5) now shows the horizontal area around the island which makes this non-orographic wave (NGW) clearer to see. The slope of the phase lines that the reviewer refers to is identified as a part of a transitory NGW, and not related to the mountain wave field over the island. 14 hours later in the next overpass, this wave is no longer present. This non-orographic (or at least, clearly not from South Georgia) wave does not appear in the model-as-AIRS. The apparent under-representation of NGWs this is one of the results discussed in the revised paper.

6. *L429-430: See my comments above: The horizontal structures in model-as-AIRS and AIRS are at best qualitatively similar over the mountain; they are dissimilar farther downstream. Please describe your comparison of results from model-as-AIRS and AIRS consistently with your high-quality figures.*

   See response above. These overpass times are 14 hours apart.

   We should mention though that in the text of the revised paper we now make a clearer distinction between qualitative and quantitative comparisons as a result of the reviewer's point, so this has been very constructive. We are also grateful for the reviewer's complement about the quality of the figures, we hope they will find the revised figures equally good.

7. *Fig. 7: How did you apply averaging over the GW scales when calculating the MF. Furthermore, the regions of phases going upward with increasing x in Fig. 4c and f should give rise to a reversal from westward to eastward MF in Fig. 7c. Please clarify.*

   The GWMF values estimated via Eqn. 1 are assumed to be averaged over one GW wavecycle (Ern et al., 2004). We have GW wavelength measurements for the dominant (largest spectral amplitude) wave at every location in the 3-D volume from the 3DST (Hindley et al., 2019), so we have an estimated GWMF value everywhere. The isosurfaces then show cuts through these values.

   We acknowledge that this is not ideal, but this is standard practise for estimating GWMF from measured GW amplitudes and wavelengths. Later (and in the revised paper), we take the area average over a well-defined 3-D volume, which again is not ideal but provides a reasonable average over GW scales, as the reviewer later suggests.

   Regarding the regions of upward sloping phase in the model-as-AIRS, recall again that Figs. 4 and 7 show different overpass times 14 hours apart. We also direct the reviewer to the revised figure in the new Fig. 5. Once the horizontal sampling and resolution is correctly applied, we can see that this upward sloping phase with increasing $x$ is no longer apparent. Again, we are very grateful to the reviewer for prompting this revision.

8. *L489-493: The wave refraction argument can be applied for either upward propagating GWs (negative vertical wavenumber) or downward propagating GWs (positive vertical wavenumber). Here you apply this argument even though the longer vertical wavelengths that you expect for a westward MW in an increasing stratospheric eastward jet show up in your plot with reversed sign. How do you explain the reversal from negative to positive vertical wavenumber at 20-30 km in Fig. 6d? Why is there a noisy mixture of positive and negative vertical wavenumbers in Fig.6h? These wavenumber (wavelength) results need to be revisited.*

   We agree with the issues highlighted in this comment. Note however that when we try to measure GWs with long vertical wavelengths, only a small amount of horizontal directional error is required to flip the horizontal direction because the phase fronts are aligned so near to the vertical.

   We thought it might be useful to discuss these changes in sign of the vertical wavenumber, just in case they were physical, but as the reviewer points out they are probably simply due to horizontal directional error in measurement of very long vertical wavelengths.

   In the revised figure (new Fig. 6), we do not discuss positive or negative vertical wavenumbers and instead we accept these regions as experimental error, and have clearly mentioned this is the text. Our revised results in later sections (see new Fig. 8) however indicate that the area-average GWMF results are not significantly affected by this error.

9. *L510-515 and L528-L532: This discussion relates to my previous comment. Please give a hint on why you possibly have positive vertical wavennumbers in AIRS. One possibility is that the background wind in the lower atmosphere shows accelera- tion/deccelerations which can cause the phase lines of MWs sloping upward/downward in time-height cross-sections. Another possibility is the generation of secondary GWs from MW breaking causing downward propagating GWs (which are no longer MWs). See also Vadas and Becker (2018, JGR Atmos.: Numerical Modeling of the Excitation, Propagation, and Dissipation of Primary and Secondary Gravity Waves during Wintertime at McMurdo Station in the Antarctic), as well as Vadas et al. (2018).*

   See response above. Both of these suggestions are possible, and in the original submission we wondered if we could be could measuring secondary GWs or some kind of reflection. But upon reflection, the data do not fully support an investigation into this, so in the revised paper we instead accept any directional errors as measurement error, rather than discussing the possibility of 2GWs here. It is something that could be considered in future, but is beyond the scope of what can be addressed in this paper.

10. *L599: Note that the wind in the lower troposphere is crucial for MW generation, while the wind at higher altitudes facilitates propagation (strongly eastward) or dynamical instability (weakly eastward or westward). Again, it is unclear how dynamical instability (including critical levels) are handled by the model, given its coarse vertical level spacing in the stratosphere and the lack of information about subgrid-scale processes.*

    See our response above regarding vertical level spacing in the model and the sensitivity tests conducted in Vosper (2015). We respectfully disagree with the reviewer here. Of course it would always be good to have more vertical layers, but we do not agree that the model has a "coarse" vertical grid spacing that could significantly affect our results in this specific study. We have however included these possible issues in the text for discussion.

    The model configuration used here is well-described in Vosper (2015). A comprehensive description of the dynamical core of the Met Office Unified Model is provided in Wood et al. (2014) and citation therein. Please consult these descriptions for information about subgrid-scale processes.

11. *L619-623: This is another example of a very speculative discussion about suspicious features in the model data. Are stationary, non-orographic GWs indeed present around the island in the global model? Are these waves artificial? Please clarify.*

    Agreed. This discussion, and other speculative discussions like it, have been removed in the revised paper. NGWs are not expected to be well-simulated in the local area model due to the coarse resolution of the global forecast. Even if there were realistic NGWs in the global forecast, it is not clear how well these waves would be "transferred" into the local area domain due to the 1hr time integration used for the lateral boundary conditions. These aspects are described in the revised paper.

12. *L638-645: How is the simulated very large MF at scales close to the horizontal grid scale possibly related to the coarse vertical level spacing and, in addition, to insufficient parameterization of dissipation processes in the stratosphere below the sponge layer? Your model results would imply that the vast majority of MW momentum flux resides at horizontal scales not even observable by AIRS. Hence, according to your model results, observations from AIRS are essentially useless to estimate the orographic GW MF from small Islands that is missing in global models? Please clarify.*

    Sensitivity tests by Vosper (2015) suggest that stratospheric vertical grid spacing in the 118-level configuration chosen here has no significant effect on the resolved GWMF, because increasing the number levels to 173 made no significant difference.

    The large GWMF at short horizontal scales ($\lambda_H$ ~50 km) found in the full-resolution model is because the vast majority of GWs in the model are mountain waves from the small island of South Georgia, which is less that 37 km across. The characteristic horizontal wavelengths of mountain waves are primarily determined by the horizontal size of the obstacle, as the reviewer later mentions. Largest mountain waves amplitudes occur directly over the island, where these short wavelengths are found. This results in very large GWMF measurements via Eqn. 1. We direct the reviewer to the new Fig. 11 in the revised manuscript, where the large amplitudes of GWs at short $\lambda_H$ are found and validated by AIRS measurements in one example with favourable viewing geometry.

*Hence, according to your model results, observations from AIRS are essentially useless to estimate the orographic GW MF from small Islands that is missing in global models? Please clarify.*

No single instrument can observe the full GW spectrum, but no other instrument can yet provide global 3-D measurements needed to constrain GWMF. We can only compare models to the observations we have.

Regarding the "usefulness" of the 3-D AIRS measurements, the horizontal scales of GWs from South Georgia can be resolved equally well or better in 3-D AIRS measurements than they can be resolved in the Kanto model of Watanabe et al. (2008) or the model of Becker and Vadas (2018), so these measurements are useful.

13. *Fig. 15: This is a very nice figure (like most of the other figures)! I cannot see the grey lines mentioned in the caption. My comment is this: The AIRS curves nicely indicate wave dissipation from about 25 km on. This wave dissipation is not reflected by the model results. Therefore, this figure supports my major concerns about the model: Too large vertical level spacing combined with possible shortcomings in subgrid-scale parameterization leads to insufficient dissipation.*

    Thanks! We have made the grey lines thicker.

    We have included these possible issues in the discussion of the results in the revised paper.

14. *L858-868: Ditto.*

    Ditto above. We have included these possible issues in the discussion of the results in the revised paper.

15. *Fig. 16a: This figure suggests that you would get a reversed power spectrum of the wave amplitude with respect to the horizontal wavenumber, i.e., increasing (instead of decreasing) power with increasing wavenumber? Please check. If this is so, this would imply that the model results at these small scales are not reliable at all.*

    Once again we apologise for the error in the bin-width normalisation in this figure. The revised figure did not feature such a prominent distribution, but we have not included this analysis in the paper because it was not useful. The average GWMF at a given wavelength and amplitudes is more or less determined by Eqn. 1, and the differences between the AIRS and the model-as-AIRS in the revised manuscript were less significant. This analysis has instead been replaced with the new Fig. 11, which provides a good indication of some of the smallest horizontal scales that occur over the island.

    But we should consider that the largest GWMF values for mountain waves over South Georgia do indeed occur at short horizontal wavelengths (large wavenumbers) up to near the characteristic size of the island ($\lambda_H \sim$30–40 km).

    We should also remember that the model configuration used here is not some global model with a full spectrum of resolved GWs in the middle atmosphere and a well-behaved GW power law spectrum. This is a small regional model (with under-represented NGWs) over South Georgia in which the single largest source of GWs is flow over mountainous orography of the island. This creates a spectrum of GWs that is unique to the physical size and characteristics of the island. If we applied the same analysis to a large oceanic region, we would expect a much better behaved power law spectrum.

16. *L927: Ditto*

    Ditto above.

17. *L978-981: As mentioned above, it is not just horizontal grid-spacing (and model numerics, as you mention in L992) that determines how well a model simulates GWs. You have to consider the vertical grid spacing as well. Most importantly, inviscid fluid dynamics cannot handle GW breakdown and wave-mean flow interaction. You need an explicit dissipative process for non-transient wave-mean flow interaction (see the non-acceleration theorem, Lindzen's GW saturation theory, or the classical McFarlane paper about orographic GW parmameterization). That is why the parameterization of subgrid-scale processes (turbulent diffusion) is very important in any GW-resolving circulation model (e.g., Becker and Vadas, 2018).*

    Agreed, thanks. We have added this to the revised paper. See our points above relating to the vertical grid spacing.

# Minor Comments

- *L73-75: I agree with this statement. However, the authors miss the opportunity to put the orographic GW momentum flux from South Georgia into the context of the general circulation in SH winter.*

  The introduction has been revised in the resubmission.

- *L92: Please point out that the model used in this study is a real-date regional model that is forced by a global forecast model via lateral boundary conditions. Therefore, this regional model is not "essentially free running".*

  Added, thanks for the correct terminology! This is an important distinction that allows the model to be compared directly to observations.

- *L135: Please be specific whether the vertical resolution relates to wavelength or grid-spacing.*

  Fixed, thanks.

- *L179: The vertical resolution of the global model is presumably too coarse to represent inertia GWs in the stratosphere. This could be the reason why the regional model misses these waves when compared to the AIRS data.*

  AIRS is very unlikely to see inertia GWs (IGWs) either, due to the deep vertical weighting functions of the AIRS instrument, so this is not likely to affect our comparison.

- *L205-214: This paragraph is hard to follow and distracts a bit from the very good writing otherwise in the paper.*

  Agreed, this has been fixed now.

- *Figure 2: Please plot the zonal wind with the same color coding as the meridional wind (blue for minus, red for plus)? Can you use a nonlinear color scale to make the accelerations and deccelerations of the tropospheric wind visible? Note that the wind in the lower troposphere determines the forcing of orographic GWs.*

  Agreed, the colour scale has been fixed now for consistency. We tried a non-linear colour scale for this figure but we found that, visually, it placed too much emphasis on whether the wind was positive/negative at low speeds and less emphasis on the large wind speeds in the stratosphere, which are important for GW propagation and refraction to long vertical wavelengths visible to AIRS and the model-as-AIRS. The surface winds are reasonably strong for most of the campaign (i.e. reasonable orographic forcing), but the stratospheric wind speeds can have a first order effect on the measured GWMF in AIRS due to GW refraction effects (Hindley et al., 2020), so it is important to highlight this.

- *L245: The radiosonde observations do not provide a horizontal average over the domain covered by the model. Please reformulate correspondingly.*

  We did not do a horizontal average over the model domain, we traced each individual radiosonde through a 4-dimensional model space ($x$,$y$,$z$,$t$) and evaluated the model temperatures along that path using linear interpolation (as stated in l.245-250). These model-as-Sondes paths were then compared to the radiosonde observations.

The wind contours in Fig. 3a are provided for illustration of the local wind conditions only.

- *L266: Figure 3 is very well composed. However, Fig. 3g illustrates that the simulated winds are not in good agreement with the radiosonde data. Rather, the agreement is only reasonable. The mean meridional wind in Fig. 2d is predominantly southward from 30 to 60 km and is of the order of a few -10 m/s. The corresponding wind in Fig. 3d shows a bias of about 10 m/s.*

  Thank you for spotting this. We assume the reviewer means Fig. 3g. Reviewer #2 also mentioned that this needed reformulating and you are both right. We have since removed obvious anomalies from the radiosonde measurements and the resulting comparison provides a much clearer result.

  There is indeed (on average) a southward bias in the model winds compared to the radiosonde observations. This now forms one of our key results of the paper, since a small corresponding northward bias in simulated stratospheric GWMF is also found later.

- *L279: Short-timescale variability would average out when comparing time-averaged wind profiles. I suggest to accept these discrepancies and to discuss the possible implications for orographic forcing and vertical propagation of GWs in the model.*

  The discussions regarding short-timescale variability have been removed in the revised paper. As the reviewer suggests, we instead accept these discrepancies and focus on what we can say with the time-averaged results. The implications for a persistent southward bias in the model may be a resulting northward bias in the simulated GWMF, which is discussed in the new results section.

- *L284-L290: See my comment with respect to L226 above.*

  Thanks, this discussion is revised now. We assume the reviewer means L266 above.

- *L300-306: The differences between model and radiosonde data are not minor. Invoking the "climatological level" of simulated wind in case studies of orographic GWs, which are subject to extreme intermittency, does not sound conclusive.*

  Agreed, we now consider the discrepancies between the model winds and the radiosonde observations more carefully in the revised paper. Note however that because the zonal wind is (usually) so much stronger than the meridional, a meridional bias in wind speed only corresponds to a small directional bias, and the magnitude of the mean wind in the model and the sondes is reasonably close. This is what we were trying to say (albeit badly), but this has been thoroughly revised now in the revised paper.

- *L360-363: These sentences are hard to understand (e.g., "vertical resolution for that vertical layer"). Please reformulate.*

  In this sentence, we meant horizontal layer, sorry. The AIRS vertical resolution changes with altitude. So for a given layer in the retrieval, this layer will have its own vertical resolution that must be applied to the model. The description of the model-as-AIRS process has been substantially revised for clarity.

- *L377: This statement is not conclusive. What about model errors?*

  We meant in terms of time separation between the AIRS overpass and the model timestep. The text has been improved in the revised paper.

- *L470-471: A "reasonable apparent similarity" is not observed when considering the dissimilarity of individual phase lines between the two data sets in Fig. 6a and e.*

  See above. As discussed, Figs. 4 and 5 show two different overpasses 14 hours apart.

- *page 25: Why is this new section called "Results". The previous Section 3 contained plenty of results, not just methodology.*

  Agreed, the sections were not well arranged in the original submission. This was a major comment from Reviewer #2. This has been substantially revised in the resubmission, and more care has been taken to separate methods from results and results from discussions.

- *L535-538: This description of secondary GW generation from MW breaking does not seem consistent with the aforementioned papers by Vadas and coauthors.*

  The sentence has been removed.

- *L539-542: This sounds very vague. I recommend to simply discard speculations of this kind. Furthermore, if you want to discuss secondary GWs in your model, then you need to consider how the model simulates dynamical instability and dissipation of resolved GWs and, hence, the necessary body forces for secondary GW generation. As discussed earlier, the very coarse vertical resolution of the model combined with the lack of knowledge about the built-in (presumably implicitly numerical) dissipation casts doubts on whether the model reliably simulates body forces from GW dissipation in the stratosphere.*

  The sentence has been removed, and we have included in the model description the following:

  "It should be mentioned that this although this vertical grid spacing is sufficient to resolve wintertime orographic waves over South Georgia, the vertical grid spacing of around 1.5–2 km in the upper stratosphere is unlikely to accurately simulate body forces under wave breaking that are necessary for secondary GW (2GW) generation (e.g. Becker and Vadas, 2018)."

- *L553: Note that this equation holds strictly only for a monochromatic GW or, at best, for a narrow spectrum of GWs. As soon as you have a broad spectrum, the wavelengths to be used at the rhs become arbitrary. More importantly: I am missing the Reynolds-type average of (T')**2 (see my comment on Fig. 7 above). Please clarify.*

  Ern et al. (2017) showed that this equation is valid for GWs visible to AIRS, which is all we apply it to in the model and the model-as-AIRS in the revised paper. This relation is not perfect, but the mid-frequency approximation on which it is derived is certainly not a "narrow spectrum" of GWs regarding those visible in satellite observations.

  We are well aware that it is valid for a monochromatic GW only. This is why we only apply it to the dominant (largest spectral amplitude) wave at each location, before taking the area-average. This is standard practice in observational GW studies. We do not pretend that this is an ideal method, but for observational of GWs where only temperature perturbations are available, to our knowledge there exists no other reliable method to estimate GWMF from measured GW temperature amplitudes and wavelengths.

  The terms in Equation 1 (where $T'$ is defined as the temperature perturbation amplitude of a GW) are consistent with previous studies involving estimates of GWMF from temperature perturbations (e.g. Ern et al., 2004). The 3DST method of Hindley et al. (2019) delivers spatially-localised phase-invariant "packet" amplitude for a GW at a given length scale (or wavelength here) which is equivalent to the average perturbation amplitude usually describe for wind perturbations (see new Eqn. 1 in the revised manuscript for the model wind perturbations).

- *L582-585: This information clarifies my previous comment at least for Fig. 8-10. Given the size of the island relative to the model domain and the GW scales in AIRS and model-as-AIRS, you use the area-average to compute the MF. I think that is the right choice here. How would the resulting MF contribute to the zonal mean parameterized in global models?*

  Agreed, the area-average approach is probably the only reasonable choice we can do regarding GW scales.

  *How would the resulting MF contribute to the zonal mean parameterized in global models?*

A similar approach would need to be taken to that employed by Hindley et al. (2019) and Hindley et al. (2020), who used a latitudinal band approach to show that $\sim$75% of the total GWMF during winter near 60°S was found over the ocean, including over small islands. If we have an area-average GWMF value for one segment of a latitudinal band, we can get it's contribution to the zonal mean by considering the fraction (zonally) of the latitude band that it occupies. A future regional study aiming to break these fractions down further into individual islands to try and constrain the contribution of each is currently planned.

- *L588: I can not see the red markers in Fig. 8-14.*

  We have made these markers bigger for clarity.

- *L681: Again, I disagree that "observed and simulated wave fields are quite similar". As mentioned earlier, there are even qualitative differences.*

  See above. Figs. 4 and 5 are 14 hours apart in time.

- *L700: What about spontaneous emission from the upper tropospheric jet stream? See Plougonven and Zhang, 2014, Rev. Geophys: Internal gravity waves from atmospheric jets and fronts.*

  Agreed, reference added.

- *L827-829: These differences could simply result from errors in the background wind (driven by the global model) in the lower troposphere, leading to errors in orographic forcing of MWs in the model. I believe the authors should discuss this role of the tropospheric winds somewhere in the paper.*

  A southward wind bias in the model is now thoroughly discussed in the revised paper, in particular relating to an observed northward bias in model GWMF, which may be related as the reviewer suggests.

- *L845-849: See my previous major and minor comments regarding the obvious and possible shortcomings of the model.*

  See our responses above.

- *L928-935: It is hard to follow these arguments. Of course, MWs can be forced by non- stationary background winds. Furthermore hourly fluctuations of the background wind would correspond to non-orographic GWs that you force at the lateral boundaries. Your discussion of possible reasons for the model shortcomings (see also L936-940) do not mention the concerns that I raised above.*

  Agreed, these arguments were poor. This discussion has been removed, because the improved sampling method for generating the model-as-AIRS resulted in significant improvements in this regard.

- *L946-947: This sentence seems not logical. Consider reformulation.*

  This whole discussion has been revised for clarity.

- *L949-950: "not so commonly"? Which observations are you aware of that show this feature of very large MW amplitudes in the stratosphere at very small horizontal scales?*

  We direct the reviewer to the new Fig. 11, where very large amplitude mountain waves at horizontal scales $\lambda_H \sim$30–40 km are simulated and observed directly over the island in the model and AIRS observations. The measurement of these short-$\lambda_H$ waves in AIRS is only possible in this example due to favourable viewing geometry of the specific overpass. The measured wavelength agrees well between all three datasets. The fact that the measured AIRS amplitudes agree reasonably well between the AIRS and model-as-AIRS suggests that the wave amplitudes in the full-resolution model ($T' \sim 45$ K near 45 km altitude) may have occurred in reality.

  GWs like this do not appear in global high resolution models like those of Watanabe et al. (2008) or Becker and Vadas (2018) because the horizontal resolution is too coarse to resolve them, however fine their vertical grids.

We did not find them in AIRS observations in the original submission because we foolishly smoothed them out with the 3×3 horizontal boxcar filter. The discovery of this example has shifted the conclusions of the paper considerably.

- *L954-955: Why should intermittency of MW forcing give rise to shorter horizontal wavelengths than stationary forcing? Usually, the structure of the topography determines the spectrum that can be forced.*

  This discussion was weak and has been superseded in the revised paper.

- *L963-964: Now you argue that an "overly-stable wind vector" could give rise to the high power of MWs at very small scales in models.*

  This discussion was weak and has been superseded in the revised paper. We thought that perhaps the MW field was too idealised in the model compared to reality.

  However, the high power of MWs at small scales near 30–40 km in the model has been shown to be realistic in the revised paper.

- *L993-997: I think that here you reveal a misconception about semi-implicit time stepping in circulation models. Semi-implicit time stepping is applied to suppress the artificial generation of very fast anelastic waves and sound waves; otherwise, smaller time steps would be required for numerical stability. In any event, the time step is always small enough to properly resolve the time scales of anelastic GWs that are well described by the representation of the model equations in gridspace.*

  Thanks for the information. Apologies if I have misunderstood, but I'm not sure if this is consistent with the description of the process employed in the Unified Model as described by (Shutts and Vosper, 2011). It's probably my misunderstanding, so don't worry. In any case, I think the use of a relatively short time step here (30s) means that these issues are not likely to be significant for our configuration.

- *L999-1000: Here you finally come up with a critical comment about the lack of dissipation in the model stratosphere.*

  We have added this as a possibility in the revised paper.

- *L1002: You did not run the model at very high spatial resolution. Your vertical resolution in the stratosphere is much coarser than even in GW-resolving global models run at moderate horizontal resolution (e.g., Sato et al. 2012., JAS: Gravity Wave Characteristics in the Southern Hemisphere Revealed by a High-Resolution Middle-Atmosphere General Circulation Model). Again, your coarse vertical level spacing is certainly not adequate to support your very high horizontal resolution.*

  We meant that we ran the model at a high *horizontal* spatial resolution. As mentioned, the sensitivity tests in Vosper (2015) did not reveal any issues with our chosen vertical grid for mountain wave simulations.

- *L1005: As long as we do not solve the (viscid) Navier-Stokes equations with a resolution of 1 cm in the troposphere, the performance of our circulation models will always depend on how unresolved (subgrid-scale) dynamical processes are parameterized.*

  We are specifically referring here to GW drag parameterisations, such as parameterised GW generation from flow over small sub-grid scale islands, not the accurate parameterisation of GW dissipation processes (which will always need parameterising), but the reviewer's point is fair. We do not claim here that increased horizontal resolution GCMs will be able to remove GW drag parameterisations altogether, we simply suggest that as resolution improves we may be able to reduce reliance on parameterised GW drag for small orographic sources, which are almost impossible to fully constrain by observations.

- *L1022-1023: Yes! See my comments above.*

  We have expanded on this significantly in the revised manuscript.

- *L1030: You did not perform sensitivity experiments using the same model with different horizontal resolutions.*

  Sensitivity tests on this model configuration (and the actual July 2013 run) are described by Vosper (2015). For the runs used here, the same model was run on both a 1.5 km and 750 m horizontal grid was simultaneously. Jackson et al. (2018) reported that the general characteristics of the mountain wave field were the same between each run. See also Vosper et al. (2016), where the balance between resolved and parameterised GW drag for varying horizontal grid resolutions is investigated directly using an identical model set up.

# Typos/Suggestions

All suggested typos and edits have been added, thank you.

# References


E. Becker and S. L. Vadas. Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model. *Journal of Geophysical Research: Atmospheres*, 123(5):2605–2627, 3 2018. doi: 10.1002/2017JD027460.

M. Ern, P. Preusse, M. J. Alexander, and C. D. Warner. Absolute values of gravity wave momentum flux derived from satellite data. *J. Geophys. Res.*, 109:D20103, 2004. doi: 10.1029/2004JD004752.

M. Ern, L. Hoffmann, and P. Preusse. Directional gravity wave momentum fluxes in the stratosphere derived from high-resolution airs temperature data. *Geophy. Res. Lett.*, 44(1):475–485, 2017. doi: 10.1002/2016GL072007.

N. P. Hindley, C. J. Wright, N. D. Smith, L. Hoffmann, L. A. Holt, M. J. Alexander, T. Moffat-Griffin, and N. J. Mitchell. Gravity waves in the winter stratosphere over the southern ocean: high-resolution satellite observations and 3-d spectral analysis. *Atmospheric Chemistry and Physics*, 19(24):15377–15414, 2019. doi: 10.5194/acp-19-15377-2019.

N. P. Hindley, C. J. Wright, L. Hoffmann, T. Moffat-Griffin, and N. J. Mitchell. An 18-year climatology of directional stratospheric gravity wave momentum flux from 3-d satellite observations. *Geophysical Research Letters*, 47(22), November 2020. doi: 10.1029/2020gl089557.

D. R. Jackson, A. Gadian, N. P. Hindley, L. Hoffmann, J. Hughes, J. King, T. Moffat-Griffin, A. C. Moss, A. N. Ross, S. B. Vosper, C. J. Wright, and N. J. Mitchell. The south georgia wave experiment: A means for improved analysis of gravity waves and low-level wind impacts generated from mountainous islands. *Bulletin of the American Meteorological Society*, 99(5):1027–1040, 2018. doi: 10.1175/BAMS-D-16-0151.1.

G. J. Shutts and S. B. Vosper. Stratospheric gravity waves revealed in NWP model forecasts. *Quart. J. Roy. Meteor. Soc.*, 137:303–317, 2011. doi: 10.1002/qj.763.

S. B. Vosper. Mountain waves and wakes generated by south georgia: implications for drag parametrization. *QJRMS*, 141 (692):2813–2827, 2015. doi: 10.1002/qj.2566.

S. B. Vosper, A. R. Brown, and S. Webster. Orographic drag on islands in the nwp mountain grey zone. *Quarterly Journal of the Royal Meteorological Society*, 142(701):3128–3137, 2016. doi: 10.1002/qj.2894.

S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, M. Takahashi, and K. Sato. General aspects of a T213L256 middle atmosphere general circulation model. *J. Geophys. Res.*, 113:D12110, 2008. doi: 10.1029/2008JD010026.

N. Wood, A. Staniforth, A. White, T. Allen, M. Diamantakis, M. Gross, T. Melvin, C. Smith, S. Vosper, M. Zerroukat, and J. Thuburn. An inherently mass-conserving semi-implicit semi-lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1505–1520, 2014. doi: 10.1002/qj.2235.

C. J. Wright, N. P. Hindley, L. Hoffmann, M. J. Alexander, and N. J. Mitchell. Exploring gravity wave characteristics in 3-d using a novel s-transform technique: Airs/aqua measurements over the southern andes and drake passage. *Atmospheric Chemistry and Physics*, 17(13):8553–8575, 2017. doi: 10.5194/acp-17-8553-2017.