Authors' response to comments from Reviewer #2 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-λ_H GW event has been added.

Response to Reviewer #2: Major Remarks

1. *Why don't you use a similar 4th-order polynomial fit as for the AIRS data to determine temperature perturbations that contain orographic and non-orographic GWs? Using SG and smoothed nSG to determine perturbations seems nice way to get the contribution of the orographic GWs in the model but doesn't immediately sound like the best choice for a comparison AIRS (unless you can show that 4th order polynomial leads to similar perturbations as the procedure described here and/or your results are not very sensitive to the background removal) [Moreover, I realized that contributions of non-orographic waves seem to be important in several sub-sections later in the manuscript. SG and nSG could be used to separate and quantify some of the non-orographic contributions in the model simulations*

(not all the analyses need to be done but for some quantities it could strengthen the findings and conclusions with respect to the non-orographic GWs).]

We'd like to thank the reviewer for their helpful and perceptive comment. They are correct that the different background removal methods needed to be revised. As mentioned above, all model-as-AIRS data has now been re-processed in an improved way that takes into account the horizontal sampling of the AIRS measurements. Because we now sample the model-as-AIRS directly onto the grid of the closest AIRS overpass, the same 4th-order polynomial fit can be applied to extract GWs from both datasets. We also apply specified AIRS retrieval noise to the model-as-AIRS for comparability. This means that the background removal method is now thoroughly consistent between the two datasets. As a result, we find a significant improvement in the agreement between AIRS and the model-as-AIRS (see revised paper). The nSG model runs are no longer used. We'd like to thank the reviewer for prompting us to consider this issue. The study is now substantially improved.

Regarding the separation of OGWs from NGWs in the model, because we cannot do the same thing for AIRS measurements, we cannot make a fair and quantitative comparison of OGWs/NGWs in the model to GWs observed in AIRS. Therefore we are not sure how useful such a separation of waves in the model would be. In any case, upon further consideration, transitory NGWs from outside the local-area domain are not expected to be well-simulated around South Georgia. This is due to the coarse resolution of the global simulation and it is unclear how realistically GWs are "transmitted" into the local-area domain. Our current method of "upwind" and "downwind" boxes either side of the island provides a crude but consistent metric for assessing the relative quantities of OGWs and NGWs in the two datasets.

As mentioned above, GW results in full-resolution model are now well-separated from the comparison of GWs in AIRS and the model-as-AIRS. This is because we cannot extract GWs from the full-resolution model in the same way as the AIRS and model-as-AIRS using the 4th-order polynomial across track fit. Instead, a polynomial fit in the zonal direction is used to extract GWs in the full-resolution model for reasonable consistency, but we stress in the revised paper that the two approaches are not directly consistent. Also, the model GWMF is now calculated using wind perturbations. This further separates these results from the AIRS and the model-as-AIRS, which are now the focus of our study.

2. *Structure of the paper:*

    - *strictly separate "data and methods" and "results"*
      This would also mean help the reader to already know by the end of Sec. 2 what to expect in the result section of the paper.
      - First part of Sec 3 and 3.1 describing the data processing is better moved to Sec 2 (which could then be called Data and Methods)
      - First part of Sec 3.3 Should be moved to Sec 2
      - First part of 3.3.2 should be moved to Sec 2.
      - Sec 4.3: Gini coefficient can be introduced in Sec. 2
      - First part of Sec 4.4. should be moved to Sec 2.
      - Sec. 2.4 already presents results and could be moved to Sec. 3 (or create new Sec 3 with only content of Sec. 2.4)

    - *separate results and discussion*
      - L397: This sentence can be left for later discussion.
      - L403-407: This sentence can be left for later discussion ...and so on

Thank you for this suggestion. The revised paper has be significantly reformulated, and there is a much better split between data/methods/results/discussion. This has significantly improved the readability and removed repeated discussions.

3. *Subjective and expletive words like "overwhelmingly" or "very" (>20 occurrences) can be reduced without losing information. There is also a large amount of speculation (some contradictions, some repetitions) in the paper (>25 occurrences of likely and >30 could explanations) that lack quantification. Just some examples: "...should result in simulated conditions over South Georgia that are very close to reality for the given time periods." "...time separation is very small and the local wind conditions can be expected to be very close to reality." "A small fraction of this distribution is likely to be measurement error, but the results may still be significant.", "Since these are clear mountain wave structures, it suggests that this could be due to errors in the speed and direction of the background wind in the model.") I recommend looking through the paper and deciding if such expressions/sentences are essential for the main content/message of the paper and if they can be justified or quantified. If not, they could be removed. Instead of listing every possible explanation for some of the observed differences between model and observations, the explanations could be limited to the one or two most relevant ones.*

   Thank you for this comment. The reviewer is absolutely correct that the original submission contained a lot of unnecessary and subjective discussions. The revised paper is more concise, with a ∼20% reduction in text. One reason for this reduction is that the revised methods mentioned above have greatly improved the consistency between AIRS and the model-as-AIRS results.

Minor Comments

- *L9: "high" instead of "very high"; you may want to add "without gravity wave parametrization" over South Georgia.*

  Added, thanks.

- *L23: please specify which scales are meant by short and long*

  Added, thanks.

- *L40: not all but "a large amount of these short vertical and horizontal scales are too small to be resolved even in recent GCMs"; pls add more recent citation (e.g., Plougonven et al 2019, How does knowledge of atmospheric gravity waves guide their parametrizations?)*

  Added, thanks. The Plougonven et al. (2020) paper is an excellent inclusion.

- *L43: In some cases? Isn't it rather the norm than the exception?*

  Agreed, sentence revised.

- *L84: GWs can propagate large horizontal distances, and from this point of view the Andes are not too far at all. (compare L705: The island lies only 2000 km east of the southern tip of South America, a region associated with the largest stratospheric mountain wave activity observed anywhere in the world)*

  Agreed, the sentence has been revised. In our response to Reviewer #1 we briefly discussed the study of Ehard et al. (2017), who showed that some mountain waves at these latitudes during winter could propagate several 1000s of km from their sources due to meridional gradients in the zonal wind.

  For discussion, we suspect that some of the large NGWs found in AIRS measurements near to South Georgia may have originated over the southern Andes and Antarctic Peninsula and could have propagated downwind via this mechanism. But we do not expect such waves to be well-simulated in the local-area model due to the coarse resolution of the global forecast that provides the lateral boundary conditions (and the time interpolation applied between these hourly forecasts on to the local-area model timestep).

Discussion of these factors is minimised in the revised manuscript because more further observations and investigation are required (which may be the focus of a future study, but is certainly beyond the focus of the present study).

- *L88: range of scale sizes? Please clarify.*

  This phrasing is confusing and has been removed. We meant that because of the different measurements involved (i.e. the radiosondes, satellites, models) a range of GW scales have been studied.

- *Fig. 1: Why are the soundings of January 2015 shown here? They are not relevant for the content manuscript. (see also comment on L190)*

  They are included to put the wintertime radiosonde measurements in context. They highlight the strong wintertime winds in the troposphere and lower stratosphere. We think they are useful for context for readers new to the field. We want to mention that the data set exists in case these measurements are useful for future researchers.

- *L137: "...to study gravity waves." Not the whole spectrum of gravity waves is small scale.*

  Agreed, the sentence has been revised.

- *L140: Is the fit applied horizontally or vertically?*

  The fit is performed horizontally in the across-track direction. We have revised the sentence to make this clearer.

- *L176: ...much finer than the 3 km vertical grid of the AIRS retrieval: this is kind of a change in the objective of the paper. "when a model is allowed to run at very high spatial resolution over South Georgia, how realistic are the simulated gravity waves compared to observations?" vs how realistic are simulated gravity waves in the observational window of AIRS? Moreover, can the vertical grid spacing of the model be directly compared to the vertical grid spacing of the retrieval? At least in the horizontal effective resolution is more like 5-10 times the grid spacing.*

  We agree. The discussion of this aspect has been significantly revised in the resubmission. In any model-observation comparison paper we can only compare gravity waves within the observational window of the measurements used, but this was not clear in the original abstract.

  We also agree about that we cannot infer model resolutions from model grid spacings alone. The text has been revised to reflect this. Vertical transport however is typically better represented in the Unified Model than in the horizontal, so we expect gravity wave perturbations to be better represented over a few vertical layers than the same number of horizontal grid cells. Even so, using the range of 5–10 times the grid spacing, this would lead to a model vertical resolution of 3–5.5 km at 20 km altitude and 6.5–15 km at 45 km altitude (25–45 km is the altitude range considered in the revised paper). These values are quite comparable to AIRS average vertical resolutions at these altitudes, so our comparison is still valid.

  Sensitivity tests for this model configuration with vertical grids with 70, 118 and 173 levels were performed by Vosper (2015). They found no significant differences in the resolved GWMF over South Georgia between 118 and 173 level simulations, suggesting that the vertical grid spacing used here is sufficient to resolve the dominant components of the mountain wave field.

- *L183: Does "no gravity wave parametrization" also mean no non-orographic parametrization?*

  Yes, we have revised the text.

- *L190: I would expect that there wasn't much mountain wave activity at all in the stratosphere in summer, so I think January 2015 can be omitted.*

  Indeed, very little GW activity was found in the AIRS measurements during January, so a comparison is not included here. The weaker stratospheric winds in summer do not typically refract mountain waves to long vertical wavelengths visible to AIRS. As with the summertime radiosonde measurements, we wanted to briefly mention that the summertime modelling data exists in case it is useful for future researchers.

- *L194: Can you revise this sentence being more specific and naming the simulated conditions you are interested in, i.e. gravity waves. Then a large part of the wave spectrum can be expected to be close to reality but not the small scales.*

  Agreed, sentence revised.

- *L196: This sentence can be omitted.*

  Agreed, the paragraph has been revised.

- *L213: How can this have an effect at all on the data above 20 km? Is this due to the analysis performed later on?*

  Sentence removed. The reviewer is right, it doesn't affect our results at all. We were finding that, in the lower troposphere, passing synoptic systems could sometimes manifest as temperature perturbations to our background fit, so we didn't want to include them. For the model, the spectral analysis method is applied to the whole vertical range, so we didn't want to risk any spectral contamination from these features.

  Fortunately however, these considerations are no longer important for the revised study due to the consistent sampling and background removal method used for AIRS and the model-as-AIRS (see revised paper).

- *L235: Was the radiosonde data assimilated in the operational analyses? This should be mentioned here.*

  No, we have now mentioned this in the text.

- *L244: Do you mean a wind reversal in the meridional wind? Meridional wind direction is also changing at 30 km on 21st of July and end of July 2013.*

  Yes. Sentence removed, it was unnecessary. We only mentioned the 10th July 2013 case because it was the most significant meridional wind reversal in the data.

- *L271: Measurement errors and artifacts should be removed from the measurement data before doing the comparison. They are not physically meaningful and are too obvious in the profiles (especially in Fig. 3b, d but also in Fig. 3g above 15 km). Moreover, it would probably help to filter the small scale fluctuations in the sounding data that are well below the vertical resolution of the model data. Fig 3b, e would then look smoother and easier to compare to 3c, d.*

  This is another very useful comment, thanks. Once we removed measurement errors from the radiosonde error by hand for each flight individually (there were more than we had realised), we found that this had a significant improvement on the resulting agreement between the zonal winds in the model and the sondes. This also made the southward wind bias in the model clearer, which has helped to reaffirm our results. The figure and associated text has been updated.

- *L286-290: "slight southward directional bias", "more northward": please revise this paragraph. The wording is very circuitous. It's easier to just say that the model tend to slightly overestimate (underestimate) the southward (northward) winds in the mid-stratosphere. Because the mean difference is zero for the zonal component, this then not only tends in a small directional bias but also in a bias in the horizontal wind speed.*

  Fixed, thanks. We have used the reviewer's wording.

- *L287: the initial and boundary conditions*

  Fixed, thanks.

- *L293-L299: In my view, this paragraph is too speculative and can be omitted. Moreover, real time forecast of one to multiple days is different from short-term forecasts of up to 6h used here. Positional errors larger than then the horizontal grid-spacing of the model (everything smaller than that does not really an influence) are hopefully not contributing to the spread because they do not occur (or rarely occur and should then be removed from the sounding data before doing the comparison).*

  We agree about the unnecessary speculation and the positional spacing errors. We have updated the paragraph to remove them in the revised paper.

  The global forecasts that supplied the lateral boundary conditions were run forward from midnight on each day for each 24-hour period, providing hourly forecasts for that day (24 in total), so the time gap between boundary conditions was 1 hour rather than 6 hours as the reviewer stated. We have updated the text for clarity.

- *L301: Comparison is concluded and then starts again with discussing the surface winds. They are already included in L281 and local topographic effects are mentioned as possible reason. So L301- L306 can be removed. Detailed discussion of topographic differences between model and reality would include a comparison of the model topography to high-quality elevation data of the island. I don't think it's relevant for the rest of the paper.*

  Agreed, some this paragraph has now been incorporated in to the discussion above, and the rest has been removed.

- *L227: Can you specify what scales are meant? Vertically it's clear to me from Sec. 2 (8-9km) but not horizontally (3 times footprint size, e.g. > approx. 80km?).*

  We agree. The sentence (line 327) was confusing, so we have removed it.

- *L330: I cannot follow the reasoning of this sentence. Is this because the model runs without GW parametrization or why/how does the generation of long scale waves depend on the smallest scales?*

  The sentence was badly phrased and is not needed so it has been removed. We meant that, particularly for the case of mountain waves from small islands, the large horizontal-scale wave structures observed by AIRS (10s to 100s of km) originate from small-scale perturbations induced by topography (1s of km and lower).

- *Figure 4: It is probably better not to show the model data above 58 km where the damping layer is located. With the saturated amplitudes and vertical phaselines, it distorts the visual perception.*

  We have included a dashed line showing the model damping layer above 58km, as was done in Fig. 15. This was an oversight, we should have added this. For this example however, we still prefer to show the model data in Fig. 4 up to the model top for completeness. We have stated this in the text. In the revision, all subsequent analysis results are presented for measurements that are well below this damping layer (altitudes less than 45 km).

- *L464: applied to the*

  Fixed, thanks.

- *L470ff: Can you provide some values for a more quantitative comparison? For example, max. amplitude (and later on horizontal and vertical wavelength) at 20 and 40 km above the island and the downstream values you are referring to for both AIRS and model.*

  Agreed, the paragraph has been revised to be more quantitative.

- *L523: Really "measurement error of AIRS" or rather an uncertainty in the analysis and determination of the sign of m?*

  The reviewer is correct, we meant uncertainty in the determination of the sign of m in the analysis. We have updated the text to reflect this.

- *Sec 4 "results": Section 3 contains already plenty of results. 4.1 could be just labelled 3.4 and so on*

  Agreed, fixed in response to major comment 2.

- *L643: I cannot follow. Isn't this a conclusion resulting from comparing model to model as AIRS? There is clearly more MF in the model outside the observational window of AIRS.*

  Agreed, the sentence was badly phrased and has been removed.

- *L979: It would be interesting to repeat the analysis with the output of the UKMO global configuration in the near future. Or was something similar already done in the past? If yes, you could add the reference here.*

  This is a good suggestion. We don't know of any studies comparing resolved gravity waves in the UKMO global model to 3-D AIRS observations (or similar) yet. The closest studies we can think of are probably Preusse et al. (2014), who analysed resolved waves in the IFS model and Holt et al. (2017), who compared GWs globally in a high resolution GEOS-5 simulation to 2-D AIRS observations.

  The key step in our study is the sampling of the model using the horizontal sampling and vertical weighting functions of the observations, which eliminates observational filter differences between them (Wright and Hindley, 2018). This is perhaps not done as routinely it should be in model-observations comparisons for GWs. Our study points to a way forward for direct like-for-like comparisons of observed and simulated GWs in these high-resolution configurations used to investigate GW generation from sub-grid scale orography. We have a planned study to compare resolved waves globally in re-analysis (probably ERA5) to 3-D AIRS observations, but this could be tricky because ERA5 assimilates AIRS radiances, so the comparison might not be straightforward.

References

B. Ehard, B. Kaifler, A. Dörnbrack, P. Preusse, S. D. Eckermann, M. Bramberger, S. Gisinger, N. Kaifler, B. Liley, J. Wagner, and M. Rapp. Horizontal propagation of large-amplitude mountain waves into the polar night jet. *Journal of Geophysical Research: Atmospheres*, 122(3):1423–1436, 2017. doi: 10.1002/2016JD025621.

L. A. Holt, M. J. Alexander, L. Coy, C. Liu, A. Molod, W. Putman, and S. Pawson. An evaluation of gravity waves and gravity wave sources in the southern hemisphere in a 7km global climate simulation. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2481–2495, 2017. doi: 10.1002/qj.3101.

Riwal Plougonven, Alvaro de la Cámara, Albert Hertzog, and François Lott. How does knowledge of atmospheric gravity waves guide their parameterizations? *Quarterly Journal of the Royal Meteorological Society*, 146(728):1529–1543, 2020. doi: 10.1002/qj.3732.

P. Preusse, M. Ern, P. Bechtold, S. D. Eckermann, S. Kalisch, Q. T. Trinh, and M. Riese. Characteristics of gravity waves resolved by ECMWF. *Atmos. Chem. Phys.*, 14, 2014. doi: 10.5194/acp-14-10483-2014.

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.

C. J. Wright and N. P. Hindley. How well do stratospheric reanalyses reproduce high-resolution satellite temperature measurements? *Atmos. Chem. Phys.*, 18(18):13703–13731, 2018. doi: 10.5194/acp-18-13703-2018.