Authors' response to comments from Reviewer #1 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 #1: Specific Comments

- *General Comments: Some figures have been placed uncomfortably far from their first citation*

  Thanks, currently the figures are placed automatically, but we will address this in the pre-print stage once we have uploaded our latex and figure files to ACP.

- *Introduction: As stated above, I believe that some significant work in relation to GWs studied with satellite data and numerical modelling in the Southern GWs hotspot should be included in the context of this section.*

This study is not specifically focused on the southern GW hot spot, but we agree it can be a motivation. We have included an additional list of relevant GW studies. Such studies are useful for context so we are happy to include them. It is difficult to know which studies the reviewer would like us to include since they did not specify any further, but we hope that we have included the studies they had in mind. They are welcome to contact us about any relevant studies that we may not be aware of yet.

- *l.134-135 and l.147 There is a need for clarification about some definitions. If vertical resolution is at best 7 km (uncorrelated or independent successive data), then the 3 km sampling is just some kind of interpolation and no wavelengths shorter than 14 km may be detected, which would have a disastrous effect on the following results. Also l.353-355.*

  The AIRS vertical resolution is determined by the vertical weighting functions of the spectral channels used in the retrieval (at best $\sim$7 km). The 3-D AIRS retrieval of Hoffmann and Alexander (2009) is calculated in 3 km vertical steps. Apologies for the confusion. We have updated the text for clarity.

  Because the vertical resolution due to the kernel functions of the selected AIRS channels (7-14 km) is always larger than twice the vertical spacing of the vertical grid (Nyquist resolution, 2×3=6 km), the vertical resolution of the former dominates our sensitivity to GWs in the vertical. We have updated this description is the revised paper.

- *l.138 If temperature uncertainty may be up to 1.5 K you need to justify how you rely on results of 1K amplitude or even less.*

  The AIRS retrieval noise is assumed to be uncorrelated pixel-to-pixel variations. If there are coherent wave cycles of a 1 K amplitude wave over many pixels, then the 1 K will still be measurable beneath the 1.5 K noise (like a noisy sine wave). But the reviewer raises a good point. It is not perfect, but in practice we find that retrieval noise is often better than estimated in Hoffmann and Alexander (2009), and such waves are measurable using our 3DST method Hindley et al. (2019).

- *l.143 Please mention the artefacts that may remain.*

  We recently found that, in some situations, minor artefacts can arise in the use of the 4th-order polynomial method (which is the main method used in the community for extracting GWs from AIRS observations). These artefacts are very small, with amplitudes less than 0.1 K, manifesting in artificial phase fronts in the along-track direction. But these are only visible if the method is applied to data that contains no clear GW features or retrieval noise. We found no evidence of such artefacts in the AIRS or model data used in our study here, so further discussion is not relevant for this paper.

- *Section 2.2: More details of the simulation characteristics are needed. You should mention the type of sponge and its intensity and the timestep that was used. Did numerical instabilities arise during the initial steps? If so, how did you handle them? Did you assess the model spin up? Did simulations exhibit alterations with slightly earlier or later initial time? In addition, your operational analyses have a 46 km resolution, whereas your simulation domain has a 1.5 km grid. Have you evaluated the possible effect of this factor of 30? Wouldn't it be advisable to use a smaller ratio? May this fact be responsible for the model not being able to adequately represent the non-orographic GWs (l.403-406, l.972-974)? How reliable are simulations if such a large structure is not "transmitted" from the forecast to the local area model?*

  The description of the model set up has been revised in the resubmission. Details regarding the model set up, including of sensitivity tests that justify the model grid configurations can be found in Vosper (2015) and Jackson et al. (2018).

  The sponge layer damping increases exponentially with height above 58.5km to the upper boundary. No indications significant spurious wave reflection from the sponge layer was found. No model instabilities were encountered. The

timestep used was 30s. The results were not sensitive to this choice and this is expected because of the second-order accuracy of the Met Office Unified Model ENDGame time integration scheme.

Model spin up: the integrations presented are long simulations (1-31 July etc), in which the local-area domain is driven by lateral boundary conditions from a sequence of 24 hourly global forecasts, which are linearly interpolated onto timestep of the local-area domain. Spin-up issues in the high resolution domain are therefore eliminated after only a few hours of the experiment due to realistic boundary conditions. The boundary conditions are re-initialised every 24 hours. We did not encounter any discrepancies in the wind or GW field over the island that would relate to initialisation issues.

The reviewer is correct here, we do not expect any GWs in the global forecast to be realistically "transmitted" into the local-area model due to the global to regional grid spacing ratio. This is now discussed in the revised paper.

We did not observe numerical artefacts near the boundaries. This nested high-resolution setup is a common configuration of the Unified Model for providing local forecasts, and the model is specifically designed to be realistic in this case.

- *L.236-237 You should also compare the numerical model with radiosonde temperature (not only wind validation) as you have it at disposal, but you should not use it for GWs as you clearly stated in l.254-256.*

  We did compare this, but we did not include it because it was not especially relevant to our comparison of stratospheric GWs.

  The results of our comparison of model and radiosonde temperatures, performed in the same way as in the paper, is shown in Fig. R1 here. We find that, on average, the model and radiosonde temperatures agree quite well, although there appears to be a systematic positive bias of around 2 K in the radiosonde measurements at most altitudes. This could indicate a cold bias in the model, or it could indicate a systematic temperature bias in the radiosonde instrumentation itself. More investigation is needed to explore these issues, which is not the focus of the present study. For this reason, we did not include the temperature results in the paper since the wind comparison is more useful for GWs. The wind results are derived from the GPS position of the balloon, so systematic biases such as this are not likely to occur.

  The radiosondes are planned to be used in a future study comparing various model/radiosonde parameters, such as comparing model vertical velocity to radiosonde ascent rate. Any systematic temperature biases may be discussed then.

- *l.298 There may be significant positional errors? How large can they be?*

  In l.298 we said that the errors were not significant. We are referring here to geolocation errors in the radiosonde position during each balloon flight. However, since these positional errors are expected to be much smaller than the model grid (metres rather than kilometres), they are not expected to be significant. The sentence has been removed in the revised paper.

- *l.449-453 The redundancy of the method should be shortly discussed or cited as it is strongly related to the reliability of the calculated GWs amplitudes. This is especially important in the context of some notable amplitude discrepancies below among AIRS and the model.*

  It is not completely clear what the reviewer means (in this context) by the redundancy of the S-transform method, but we suspect that they are referring to its robustness in response to unreliable, anomalous, featureless or noisy data, or systematic differences in how different datasets are analysed.

  The 3DST method we apply here is very thoroughly tested and validated in Hindley et al. (2019). It is based upon the 2DST method of Hindley et al. (2016), who also tested it thoroughly. Synthetic wave fields with simulated noise,

in addition to AIRS measurements, are analysed in both studies for a complete evaluation. We are as confident as we can be that what comes out of the analysis is a fair representation of what went in, whatever the dataset.

Regarding the measured amplitude differences (which are smaller in the revised paper due to better sampling), the exact same 3DST analysis method is applied to AIRS and the model-as-AIRS. The software package we use, and the method that it follows, are entirely independent of what kind of data is being analysed. Thus, we are as confident as we can be that any amplitude discrepancies that may arise are inherent to the different datasets, not to the analysis method. We have updated the text to make this clearer.

- *l.490-491 You should check your hourly output for stationary phases and increasing wind speed with height.*

  The increasing model wind speed with height can be seen in the new Fig. 2. We have inspected the hourly output of the model temperature perturbations during both modelling periods as video animations. We find that the phase fronts forming over the island during this time are stationary with respect to the ground, highly indicative of mountain wave activity. We have updated the text to make this clearer.

- *l.643-645 You should use your model simulations to test this argument.*

  This sentence was not clear, so it has been removed. This section has also been significantly revised in the resubmission. It is expected that the full-resolution model simulates significant GWMF that is below the resolution and sampling limits of AIRS due to the fine horizontal grid of the simulation. The new Fig. 11 highlights this. The full-resolution model results are also not compared to the AIRS observations in the revised paper, because we realised that a fair comparison is not possible.

- *l.650-651 The model simulations should give you a clue for upward or downward phase propagation.*

  Because the same procedure must be applied to AIRS and the model-as-AIRS, we cannot use supplementary model information to constrain the upwards/downwards waves in the model-as-AIRS, because we cannot do the same for the AIRS measurements. Therefore, it would not be a fair comparison.

  There is no way to independently break the upwards/downwards ambiguity from the AIRS temperature measurements alone, so to make a fair comparison we must use the same approach for both the AIRS and model-as-AIRS. We accept any small directional errors in order to ensure consistency.

- *l.678 Does this imply that AIRS amplitudes are typically the double of the model? If so, explain.*

  Please see the new results in the revised paper. Area-averaged GW amplitudes in AIRS are typically larger than the model-as-AIRS, likely due to the presence of NGWs away from the island. As mentioned above, NGWs away from the island are unlikely to be well-simulated in the local-area model. This is discussed in the revised paper.

- *l.701-702 The presence of the jet, the polar vortex, storms and fronts can all be probably checked from your operational analyses in order to verify the support to your argument.*

  Indeed, this is a good suggestion. We have included these possibilities in the revised paper, but Fig. 11 has been removed for brevity because it did not provide a particularly quantitative comparison.

  The discussion of these issues has been revised in the resubmission. As mentioned above, it is not expected that NGWs would be realistically "transmitted" through the boundary conditions of the local-area model.

- *l.706-714 Please use your hourly simulations to verify at least partially in the mentioned geographical domain your detachment or moving secondary waves argument.*

  This argument has been removed from the revised paper, but it is a legitimate possibility.

  The detachment or 3-D advection theory for mountain waves is described and discussed in the modelling study by Sato et al. (2012). Further, Ehard et al. (2017) used forward ray tracing analysis of mountain waves in lidar data

over New Zealand to show that it was possible for such waves to propagate several 1000s of km horizontally in the stratosphere due to the strong meridional shear of the zonal wind. They found examples where the original mountain wave structure over the mountains had dissipated but the advected part of the wave structure still continued to propagate far away from its source. Their results were supported by reanalysis and AIRS observations.

Figure R2 shows four snapshots from an animation of gravity wave temperature perturbations at 30 km altitude in ERA5 reanalysis during July-August 2012 over the southern hemisphere. This is an animation we prepared previously a conference presentation. Gravity wave temperature perturbations are extracted from the background temperature via a zonal planetary wave fit.

Here we can see that mountain wave structures that form over the Andes are found to extend far out over the Southern Ocean to the east. In several cases, we found that these perturbations extended all the way to South Georgia. The size and orientation of these perturbations is consistent with some of the apparent non-orographic wave activity seen in AIRS observations in our study over the ocean around the island.

We appreciate the reviewer's point. These kinds of events not typically expected under classical mountain wave theory in the 2-D case, because the meridional gradient of the zonal wind speed is not considered. Further investigation into this aspect is expected in future studies, but it is not the focus of the present study.

- *Table 1: To check if differences are significant it is necessary to include uncertainties with the averages.*

  The reviewer has a good point, but uncertainties are not straightforward to calculate for these measurements. Further, we are confident that the sources of measurement error are the same for each data set. This is especially so in the revised paper, where specified AIRS retrieval noise is also applied to the model-as-AIRS for consistency.

  Regarding the values in Table 1, standard error values could be included but for these average values they are not helpful for the following two reasons:

  1) The average wave amplitudes and GWMF are all area averages over a large number of time steps. The estimated standard error of the mean of a distribution is given by $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$, where s is the sample standard deviation and n is the number of sample points in the distribution. Our measurement grid between 25km and 45km altitude is 90x120x14 points. There are 87 AIRS overpasses and over 1300 model timesteps. For AIRS and the model-as-AIRS this yields $n > 10000$ sample points that go into the Table 1. This yields a standard error of order $\sim 10^{-3}$ for these average values, which is not helpful.

  2) The same is applied to AIRS and the model-as-AIRS. As a result, any systematic errors will propagate through the analysis in the same way for each data set. Thus, any relative differences in the area-averaged and time-averaged values shown in Table 1 are related to real differences between the data sets over the two regions.

  Errors in the measurement of individual wave amplitudes and wavelengths, from which these values are derived, are harder to define. For this, we would point towards the testing and validation of the 3DST measurement technique in Hindley et al. (2019). However, any systematic errors these measurements would be identical for each dataset (or random errors would average out in the large sample size). Therefore, any mean differences in our comparison are likely to be due to genuine differences in the datasets.

- *Figure 14d,h: Please discuss the parts where the difference is larger than the absolute fluxes.*

  We have gone back and checked our data, and there are no regions in Figs. 14(d,h) where the difference between the AIRS and model-as-AIRS shown (AIRS minus model-as-AIRS) is larger than the absolute AIRS fluxes. Note that these are differences in the absolute flux between the AIRS and model-as-AIRS data and do not imply a direction, and that panels (a,e) and (d,h) use different colour scales.

  In any case, Fig. 14 has been revised and simplified in the resubmission (new Fig. 9).

- *Figure 16: How can you define a unique amplitude if it can change by a factor over 4 from 25 to 55 km?*

  We do not define a unique amplitude, we simply show the average of all measurements between the two altitudes. Indeed, we are aware that the exponential increase of wave amplitudes with altitude will bias the average value to be more representative of values at higher altitudes, but the vertical resolution of the AIRS and the model-as-AIRS is actually significantly better at lower altitudes below around 40 km (see Fig. 5 of Hoffmann and Alexander (2009) or Fig. 2 of Hindley et al. (2019)), so there are a range of aspects involved here.

  In the revised paper, we now only consider averages between 25 and 45km to address this issue.

- *l.896-897, 913-916 and 919-920 This could indicate that AIRS may be omitting an essential contribution to GWs momentum flux and its later parametrization in global models. This fact merits a quantification of the above effect due to the possible discrepancies of simulations or observations in this work with the real atmosphere.*

  As mentioned above, Fig. 16 has been removed in the revised paper due to an error in the bin-width normalisation

  Of course, the observational filter of AIRS means that it can only observe GWs within a spectral range. We do not claim that AIRS can observe all the GWs required to constrain GW parameterisations in global models, but to our knowledge these 3-D observations are the only global dataset that that can independently measure directional GWMF, so they do have value.

  As the reviewer suspects, the sampling pattern of AIRS means that GWs at short-λ_H directly over the island are not observed in AIRS unless the viewing geometry is favourable (see new Fig. 11 in the revised paper). This is discussed in the revised paper.

  A quantification of the GWMF that is not visible to AIRS can be found in Table 1, where GWMF in the full-resolution model is also included. It is shown in the revised paper however that when the AIRS sampling and resolution is correctly applied to the model, the agreement in GWMF between the observations and the model-as-AIRS is reasonably good.

- *l.936-940 Can you give a reference where this effect has been quantified? How likely is it that this high frequency wind variability exists in that zone? Can you draw conclusions from the individual radiosonde profiles?*

  This is a good suggestion, but this argument is not discussed so much in the revised manuscript due to the better agreement between AIRS and the model-as-AIRS, after the improved sampling and resolution methods are applied.

  For the sake of discussion, our radiosonde comparison in Fig. 3 supports this argument somewhat. The model winds evaluated along the radiosonde flight path exhibit less small-scale variability than the radiosonde measurements themselves. But this is no longer a significant consideration in the revised paper.

- *l.951-957 Again, another possibility is that AIRS is missing these GWs.*

  Yes, see our point above and the new Fig. 11 in the revised paper. Some of the characteristic mountain waves over South Georgia occur at scales around λ_H $\sim$30–40km, which are only visible to AIRS when the viewing geometry is favourable. This means that such waves may be underestimated in recent global climatologies such as Hindley et al. (2020). This is a key point in the revised paper.

- *l.993-1000 What was the expectation for GWs amplitudes in your simulations according to the timestep you have chosen? Was it in agreement with your results?*

  The timestep used was 30 seconds, and the GW amplitude results were not sensitive to this choice, provided that it was a sensible value (e.g. of order 30 seconds to several minutes etc). This was expected because of the second-order accuracy of the Unified Model ENDGame time integration scheme.

- *l.1024-1027 Please check if further analysis in the previous sections produces any modifications.*

  As discussed above, the paper has been substantially improved by the reviewers suggestions. The analysis has been re-formulated and the revised paper has been significantly improved.

Minor Technical Corrections and Comments

All corrections made, sentences rephrased/deleted and references added, thank you.

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.

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.

L. Hoffmann and M. J. Alexander. Retrieval of stratospheric temperatures from Atmospheric Infrared Sounder radiance measurements for gravity wave studies. *J. Geophys. Res.*, 114:D07105, 2009. doi: 10.1029/2008JD011241.

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.

K. Sato, S. Tateno, S. Watanabe, and Y. Kawatani. Gravity Wave Characteristics in the Southern Hemisphere Revealed by a High-Resolution Middle-Atmosphere General Circulation Model. *J. Atmos. Sci.*, 69:1378–1396, 2012. doi: 10.1175/JAS-D-11-0101.1.

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.

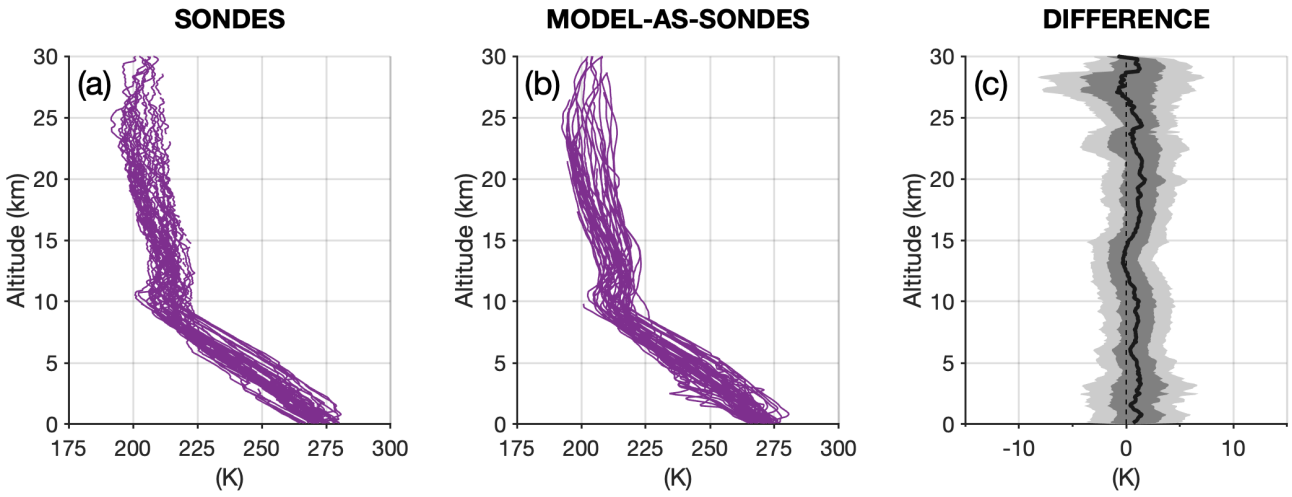

Figure R1: Temperature measurements in (a) the local-area model and (b) coincident radiosonde observations during June-July 2015. Temperature in the model are evaluated along the radiosonde flight path in 3 spatial dimensions and time. Panel (c) shows the difference (Sondes − Model) between them.

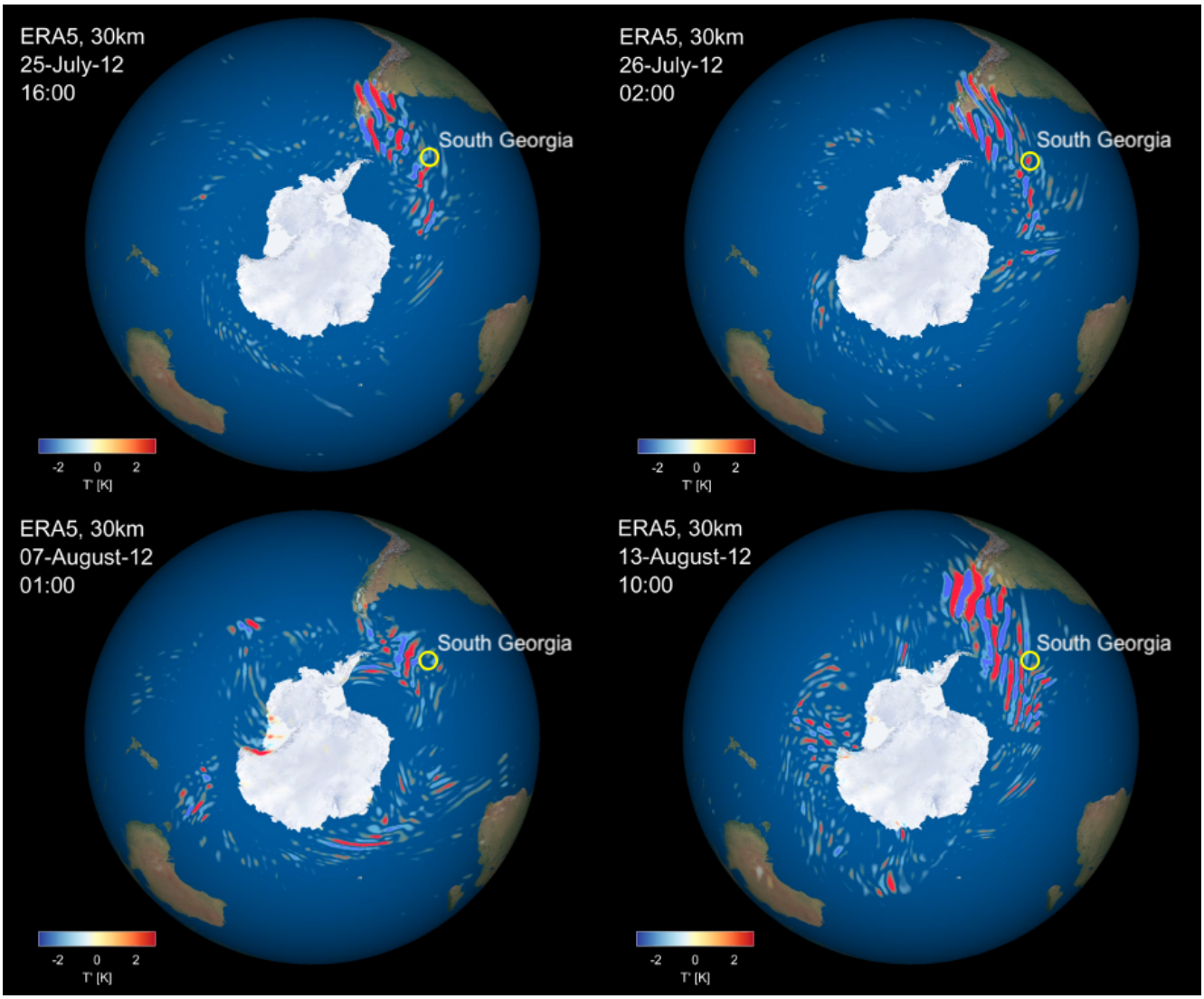

Figure R2: Selected snapshots from an animation of ERA5 reanalysis temperature perturbations over the southern hemisphere at 30 km altitude during July-August 2012.

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.

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-λ_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 800×600×118 grid. A simultaneous run on a 750 m horizontal grid was also performed on a 1600×1200×118 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. λ_H ∼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 λ_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 λ_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 λ_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 (λ_H $\sim$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 λ_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 ∼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 λ_H ∼30–40 km are simulated and observed directly over the island in the model and AIRS observations. The measurement of these short-λ_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.

Stratospheric ~~gravity-waves~~ gravity waves over the mountainous island of South Georgia: testing a high-resolution dynamical model with 3-D satellite observations and radiosondes

Neil P. Hindley[1,2], Corwin J. Wright[1], Alan M. Gadian[2], Lars Hoffmann[3], John K. Hughes[2], David R. Jackson[4], John C. King[5], Nicholas J. Mitchell[1], Tracy Moffat-Griffin[5], Andrew C. Moss[1], Simon B. Vosper[4], and Andrew N. Ross[2]

[1]Centre for Space, Atmospheric and Oceanic Science, University of Bath, Bath, UK
[2]School of Earth and Environment, University of Leeds, Leeds, UK
[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany
[4]Met Office, Exeter, UK
[5]Atmosphere, Ice and Climate Group, British Antarctic Survey, Cambridge, UK

Correspondence: Neil Hindley
(n.hindley@bath.ac.uk)

Abstract. Atmospheric gravity waves ~~are key drivers of the transfer of energy and momentum between the layers of the Earth's atmosphere. The accurate representation of these waves in General Circulation Models~~ (GWs) play an important role in atmospheric dynamics, but accurately representing them in general circulation models (GCMs) ~~however has proved very~~ is challenging. This is ~~because large parts of the gravity wave spectrum are at scales that are near or below the resolution of global GCMs. This is especially relevant for small isolated mountainous islands such as South Georgia (54°S, 36°W)~~ especially true for orographic GWs generated by wind flow over small mountainous islands in the Southern Ocean. ~~Observations reveal the island to be an intense source of stratospheric gravity waves, but their momentum fluxes can be under-represented in global models due to its small size. This is a crucial limitation, since the inadequate representation of gravity waves near 60°S during winter has been linked to the long-standing "cold-pole problem" , where the southern stratospheric polar vortex breaks up too late in spring by several weeks. Here we address a fundamental question: when a model is allowed to run at very high spatial resolution over South Georgia,~~ Currently, these islands lie in the "grey zone" of global model resolution, where they are neither fully resolved nor fully parameterised. It is expected that as GCMs approach the spatial resolution of current high-resolution local-area models, small-island GW sources may be resolved without the need for parameterisations. But how realistic are the ~~simulated gravity waves~~ resolved GWs in these high-resolution simulations compared to observations? ~~To answer this question, we present a 3-D comparison between satellite gravity wave observations and a high resolution model over South Georgia. We use a dedicated~~ Here, we test a high-resolution ~~run~~ (1.5 km horizontal grid, 118 vertical levels) local-area configuration of the Met Office Unified Model over ~~South Georgia and coincident 3-D satellite observations from NASA AIRS/Aqua~~ the mountainous island of South Georgia (54°S, 36°W), running without GW parameterisations. The island's orography is well-resolved in the model, and real-time boundary conditions are used for two time periods during July 2013 and June-July 2015. ~~First, model winds are validated with coincident radiosonde observations . The AIRSobservational~~

filter is then applied to the model output to make the two data sets comparable. A We compare simulated GWs in the model to coincident 3-D *S*-transform method is used to measure gravity-wave satellite observations from AIRS/Aqua. By carefully sampling the model as AIRS, we present the first like-for-like comparison of simulated and observed 3-D GW amplitudes, wavelengths , directional momentum fluxes and intermittency in the model and observations. Our results show that although the timingof gravity wave activity in the model closely matches observations, area-averaged momentum fluxes are generally up to around 25% lower than observed. Further, we find that 72and directional GW momentum flux (GWMF) over the island using a 3-D *S*-transform method. We find that the timing, magnitude and direction of simulated GWMF over South Georgia are in good general agreement with observations, once the AIRS sampling and resolution is applied to the model. Area-averaged zonal GWMF during these two months is westward at around 5.3 mPa and 5.6 mPa in AIRS and the "model-as-AIRS" respectively, but values directly over the island can exceed 50 mPa. However, up to 35% of the total flux in the model region is located downwind GWMF in AIRS is actually found upwind of the island, compared to only 5717% in the AIRS measurements. Directly over the island, the model exhibits higher individual flux measurements but these fluxes are more intermittent than in observations, with 90% of the total flux carried by just 22% of wave events, compared to 32% for AIRS . Observed gravity wave fluxes also appear to dissipate more quickly with increasing height than in the model , suggesting a greater role for wave-mean flow interactions in realitymodel-as-AIRS, suggesting that non-orographic GWs observed by AIRS may be underestimated in our model configuration. Meridional GWMF results show a small northward bias (∼20%) in the model-as-AIRS that may correspond to a southward wind bias compared to coincident radiosonde measurements. Finally, spectral analysis of the wave fields suggests that the model over-estimates gravity wave fluxes at short horizontal scales we present one example of large-amplitude ($T' \sim 15$–20 K at 45 km altitude) GWs at short horizontal wavelengths ($\lambda_H \sim 30$–40 km) directly over the island , but under-estimates fluxes from larger horizontal scale non-orographic waves in the region, leading to a lower average value overall. Our results indicate that , although increasing model resolution is important, it is also important to ensure that variability in the background wind vector and role of non-orographic waves are accurately simulated in order to achieve realistic gravity wave activity over the Southern Ocean in future GCMsin AIRS measurements that show excellent agreement with the model. This suggests that orographic GWs in the full-resolution model with $T' \sim 45$ K and $\lambda_H \sim 30$–40 km can occur in reality. Our study demonstrates that not only can high-resolution local-area models simulate realistic stratospheric GWs over small mountainous islands, but the application of satellite sampling and resolution to these models can be a highly effective method for their validation.

1 Introduction

Atmospheric gravity waves (GWs) are a key driver of the atmospheric circulation. These waves a play key role in many important chemical and dynamical processes throughout the atmosphere via the transportation dynamical component of the Earth's atmosphere. Through the vertical transport of energy and momentumbetween atmospheric layers and across great distances (e.g. Fritts and Alexander, 2003; Fritts et al., 2006; Alexander et al., 2010). Gravity waves carry a flux of horizontal pseudo-momentum which is deposited when , these waves are an important coupling mechanism between between atmospheric

layers (e.g. Fritts and Alexander, 2003; Fritts et al., 2006). When they break or dissipate, GWs deposit a horizontal momentum forcing into the background flow, resulting in a drag or driving force ~~on the background atmospheric flow. The large-scale effects of this forcing have significant impacts throughout the whole stratosphere and mesosphere.~~ that drives circulations away from states expected under radiative equilibrium.

~~Global~~ But despite their importance, accurately representing GWs in global circulation models (GCMs) used for numerical weather and climate forecasting ~~must therefore include the effects of gravity waves. Indeed, it is now recognised GCMs must have a well-resolved stratosphere that includes realistic dynamics, including resolved gravity waves, in order to deliver accurate seasonal weather forecasts, predict long-term climate change and predict the future of the ozone layer (e.g. Baldwin et al., 2018).~~

~~However, gravity waves are notoriously difficult to represent, even in state-of-the-art numerical models.~~ has proved challenging (Alexander et al., 2010; Plougonven et al., 2020). One reason for this is that ~~they have a very wide spectrum of physical scale sizes, ranging from hundreds of metres to a few tens of kilometres in the vertical and from tens to many hundreds of kilometres in the horizontal. These short vertical and horizontal scales are too small to be resolved in GCMs so their effects are instead represented by parameterisations, where a momentum-forcing term is applied the background flow (e.g. Warner and McIntyre, 1996)~~ a large fraction of GWs and their sources lie at physical scales that are below the spatial resolution of GCMs. The momentum forcing of these sub-grid waves on the background flow must instead be simulated by parameterisations. (e.g. Warner and McIntyre, 1996; K .

~~However, these parameterisations remain poorly constrained by observations. In some cases, there is a risk that gravity wave parameterisations are so poorly constrained that they begin to take on a role as simply a tuning parameter that is adjusted in order to reproduce realistic large-scale dynamics in reanalyses, such as zonal-mean zonal winds and a realistic quasi-biennial oscillation (QBO) (Alexander et al., 2010; Wright and Hindley, 2018), rather than being adjusted to match observational measurements.~~

~~At larger gravity wave scales of hundreds of kilometres, current operational GCMs can resolve waves directly in the stratosphere. Given the advances in computer power, GCMs are likely to operate at ever-finer resolution in the coming years, which will enable them to resolve more and more of the gravity wave spectrum. Indeed, offline high-resolution simulations are already being used to help tune gravity wave parameterisations in operational GCMs(e.g. Vosper, 2015; Vosper et al., 2016, 2020) . A question then arises, as posed by Preusse et al. (2014): in the future, will ever higher spatial resolution in GCMs remove the need for gravity wave parameterisationsaltogether?~~

This ~~question~~ is especially significant for small ~~, isolated mountainous islands whose physical scales of a few tens of kilometres are at or near the spatial grid size of current GCMs. In terms of orographic drag parameterisations, these islands~~ mountainous islands in the Southern Ocean. Observations reveal intense "hot spots" of stratospheric GW activity over these islands during austral winter (Alexander and Grimsdell, 2013; Hoffmann et al., 2013, 2016; Hindley et al., 2020), but due to their small size islands like these lie in the "grey zone" of orographic GW parameterisations, where they are neither fully resolved nor fully ~~subgrid (Vosper et al., 2016). As such, gravity wave generation by flow over~~ parameterised (Vosper, 2015) . Thus, orographic GW drag from small mountainous islands ~~may be significantly underestimated in GCMs. Several recent~~

studies can often be inaccurately simulated in GCMs, which can in turn result in a significant underestimation of GW momentum (McLandress et al., 2012; Vosper et al., 2016; Garfinkel and Oman, 2018).

These islands also lie beneath a "belt" of intense wintertime GW activity at latitudes near 60°S, which also includes the well-known hot spot of GW activity over the Southern Andes and Antarctic Peninsula. Gravity wave activity in this region, and the surrounding 60°S belt, has been explored in numerous observational and modelling studies in the past two decades (Eckermann and Preusse, 1999; Jiang et al., 2002; de la Torre and Alexander, 2005; de la Torre et al., 2006; Hertzog et al., 2008; Alexand .

Recent studies have suggested that under-represented gravity wave momentum flux from isolated islands in the Southern Ocean "missing" GW momentum flux near 60°S may be a significant contributing factor to the wintertime "cold-pole problem", one of the largest and most a significant and long-standing biases common to virtually bias in nearly all major weather and climate models (Scaife et al., 2002; Butchart et al., 2011; McLandress et al., 2012; Alexander and Grimsdell, 2013; Garfinkel and Oman, 2018). This is in addition to significant non-orographic wave activity in the 60°S belt (Hendricks et al., 2014; Hindley et al., 2015, 20 .

The cold-pole problem refers to a simulated wintertime stratospheric polar vortex that , when compared to observations, is too cold by around 5 to 10 K, has winds that are too strong by around 10 ms^{-1} and persists for some breaks up around two to three weeks too long into spring before breaking up late into spring compared to observations (e.g. Butchart et al., 2011). This bias dynamical bias also causes difficulty in simulating chemical systems such as the stratospheric ozone cycle (e.g. Garcia et al., 2017), global chemical transport (e.g. McLandress et al., 2012) and surface climate change in the Antarctic (Thompson et al., 2011). Poor simulation of GWs in this region has other important implications. For example, GWs play an important role in the dynamical balance of the stratospheric jet (Choi and Chun, 2013), which in turn affects surface storm track locations at mid- and high latitudes (Perlwitz, 2011). Gravity wave modulation of background winds can also affect the formation of polar stratospheric clouds (Höpfner et al., 2006), which can significantly reduce ozone concentrations, a principal driver of Antarctic climate change (McLandress et al., 2011; Garcia et al., 2017). There is thus a critical need to establish the sources, fluxes and variability of GWs from small islands near 60°S in order to determine their contribution to this region and so guide the development of GCMs.

Regarding the question

At larger horizontal scales of a few hundreds of kilometres, GWs can usually be directly resolved in current operational GCMs. To resolve GWs at fine horizontal and vertical scales, dedicated offline simulations are needed, which have provided encouraging results in recent years (e.g. Watanabe et al., 2008; Sato et al., 2012; Holt et al., 2017; Becker and Vadas, 2018). High-horizontal resolution offline simulations can also be used to help improve GW parameterisations for sub-grid scale orography in operational GCMs (e.g. Vosper, 2015; Vosper et al., 2016, 2020).

Future advances in computing power will likely result in ever-finer horizontal and vertical grids in operational GCMs, which will enable the resolution of a large part of the GW spectrum. A question then arises, as posed by Preusse et al. (2014), it is possible that if models are run at sufficiently high spatial resolution , gravity wave fluxes : in the future, will ever-higher spatial resolution in GCMs remove the need for GW parameterisations altogether? For orographic GWs from small mountainous

islandswill be accurately represented and the cold-pole bias may be reduced. This raises a follow-on question, which is the focus of our study: when a model is allowed to run at very high spatial resolution over a small mountainous island in the Southern Ocean, , where spatial resolution is a key limiting factor in their representation, this seems to be a realistic possibility. But how realistic are the simulated gravity waves compared to observationssimulated GWs in high-spatial resolution simulations over these islands?

In this study, we address this question for one such island: South Georgia (54°S, 36°W) . South Georgia is a classic example of an isolated mountainous island in the Southern Oceanthat experiences intense surface wind conditions during winter, so is an ideal subject for our study. The island is approximately . Despite being only around 170 km long, 25 km wide and lies more than 600 km from any other major islands and more than 2000 km from the nearest continent of South America. The island South Georgia is entirely mountainous and has interior peaks that reach heights of with interior peaks exceeding 3000 m. The sharpness and During winter, the abrupt orientation of the topography relative to the prevailing wind, combined with the great distance from other potential orographic sources, make the island a ideal natural laboratory for orographic gravity wave generation . Previous studies of gravity waves strong prevailing wind provides favourable conditions for orographic GW generation and vertical propagation. Previous observational and modelling studies over South Georgia and other small islands in the Southern Ocean that have been conducted using a diverse range of observations and models (e.g. Alexander et al., 2009; Hoffmann et al., 2013; Vosper, 2015; Vosper et al., 2016; Hoffmann et al., 2016; Moffat-Griffin et al., 2017; G have revealed intense gravity wave activity at a range of scale sizes have revealed intense wintertime GW activity in the troposphere and middle atmospherestratosphere over the island (e.g. Alexander et al., 2009; Hoffmann et al., 2013; Vosper, 2015; Vosper et al., 2

.

Here, To investigate simulated GWs over South Georgia, we use a dedicated high-resolution local area local-area configuration of the UK Met Office Unified Model (1.5 km grid, 118 vertical levels)is run for the region around South Georgia. No gravity wave parameterisations are applied in this local area model , so the dynamics are essentially free-running. In this study use data from two runs: one . The local-area model is nested in a real-date configuration for two time periods during July 2013 and another during June-July 2015using model configurations as described by Vosper (2015) and Jackson et al. (2018) respectively. Gravity waves in these simulations are compared to co-located , where lateral boundary conditions are provided by a global forecast, which ensures that simulated conditions are close to reality. No GW parameterisations are applied in the local-area model. After validating the model winds with coincident radiosonde observations, we compare simulated GWs over South Georgia to observed GWs in co-incident 3-D satellite observations from NASA AIRS/Aqua . A specialised 3-D temperature retrieval for AIRS is used that has superior spatial resolution over the standard AIRS product. After validating the model wind with co-located radiosonde observations, the observational filter for the same time periods. By applying the vertical resolution, horizontal sampling and retrieval noise of the AIRS retrieval is applied measurements to the modeloutput so that we can compare the , we are able to make a direct like-for-like comparison of observed and simulated wave fields fairly. This novel approach allows us to address the question of how realistic the simulated gravity waves in the high-resolution model are when compared to observationsGW amplitudes, wavelengths and directional momentum fluxes over the island.

In Sect. 2 we describe the model, satellite and radiosonde datasets ~~that we use~~ used in this study. ~~We also validate the model winds with the co-located radiosonde observations here to ensure they are realistic. In Sect. ??~~ In Sections 3 and 4 we validate background winds in the model using the radiosonde observations and inspect the simulated GWs. Then in Sect. 5 we apply the AIRS ~~observational filter~~ resolution, sampling and retrieval noise to the model ~~output fields so that we can compare gravity waves in the two data sets fairly and we apply a~~ to make a fair comparison of GW measurements in AIRS and the model-as-AIRS. A 3-D S-transform analysis method for measuring ~~gravity wave properties . In Sect. 7 we present simulated and observed wave~~ GW properties is described in Sect. 6, after which we present a comparison of measured GW amplitudes, wavelengths and directional momentum fluxes over South Georgia ~~during July 2013 and June-July 2015. We investigate the distribution and intermittency of these gravity wave properties with respect to time, altitude, distance from the islandand spectral properties~~in the model and satellite observations in Sect. 7. In Sect. 8 we investigate a case study of large amplitude GWs at short horizontal wavelengths over the island. These results are ~~then~~ discussed in Sect. 9, and ~~our conclusions are presented~~ we draw our conclusions in Sect. 10.

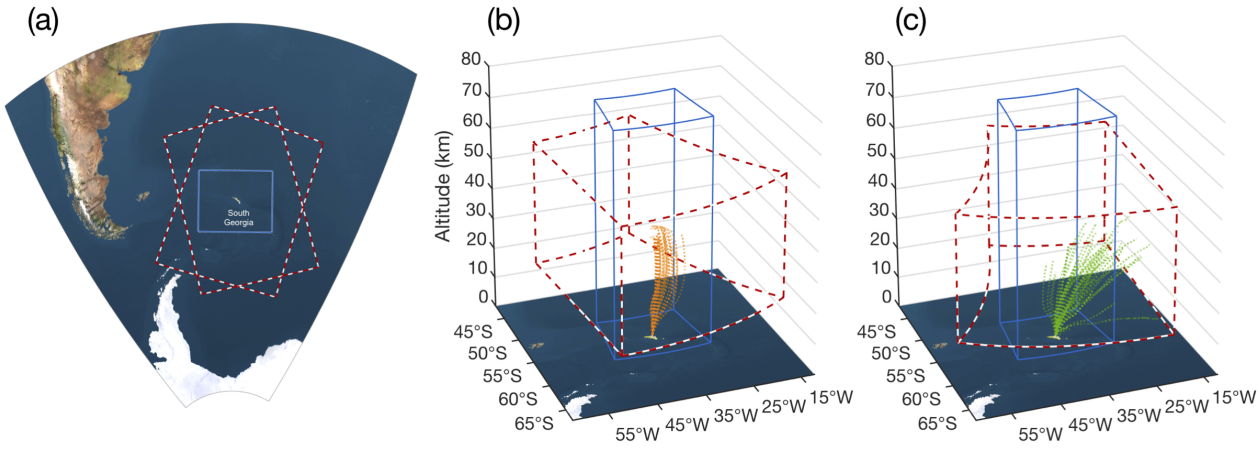

Figure 1. Maps showing the horizontal and vertical extent of the ~~local area~~ local-area model (blue lines) ~~around~~ over the island of South Georgia and two ~~typical~~ examples of ~~satellite observations from~~ the typical extent of AIRS satellite measurements (red and white dashed lines) used in this study. Panel (a) shows a map of the local region around South Georgia, plotted on a regular distance grid. Panels (b) and (c) show the vertical extent of the model on a latitude-longitude grid. The vertical extent of usable temperature data from the 3-D AIRS retrieval scheme of Hoffmann and Alexander (2009) is shown in red dashed lines for both an ascending (b) and descending (~~b~~c) overpass. Orange (green) lines show the trajectories of radiosondes launched from the island during a summer (winter) campaign in January (June-July) 2015.

2 Data

Three atmospheric datasets over South Georgia are analysed in this study: 1) ~~3-D satellite observations from AIRS/Aqua; 2)~~ modelling simulations in a local-area domain centred on the island; 2) 3-D satellite observations from AIRS/Aqua; and 3) radiosonde observations launched from the British Antarctic Survey (BAS) base at King Edward Point (KEP)~~during June-July 2015. Comparisons of the satellite observations and modelling simulations make up the main results of this study, while the radiosonde observations are used to validate the modelling simulations.~~ .

The ~~geographical locations and~~ spatial extent of ~~the three datasets are~~ these three datasets is shown in Fig. 1. South Georgia is located ~~near the centre of Fig. 1a, lying~~ around 2000 km east of South America and the Antarctic Peninsula in the Southern Ocean. The 1200 km × 900 km local-area ~~domain of the modelling simulations~~ modelling simulation over the island is shown by the light blue box in Fig. 1a, while the two red-and-white dashed boxes show two example overpasses of the AIRS instrument (one during ~~ascending~~ an ascending node orbit and one during a descending node)~~over the area. The~~ . Note that the exact location of each of the overpasses varies with each orbit, as discussed below. Figures 1b and 1c show 3-D views of these domains~~. Also shown are~~ , through which the trajectories of ~~radiosonde balloons launched from KEP on South Georgia during January (green)~~ radiosondes launched from the island during January and June-July ~~(orange)~~ 2015. 2015 are shown by dashed orange and green lines respectively. Note that the June-July radiosondes travelled much further downwind due to stronger stratospheric zonal winds during austral winter, and many of these travelled so far east that they exited the local area model domain.

2.1 ~~AIRS satellite observations~~

~~We use satellite data from the Atmospheric Infrared Sounder (AIRS) on NASA/Aqua (Aumann et al., 2003). Aqua has a ~100-minute near-polar sun-synchronous orbit, with an ascending-node equator-crossing local solar time of 1:30pm. It has been collecting data since August 2002, with only minor interruptions since that date.~~

~~AIRS is a nadir-sensing instrument that makes measurements in the across-track direction at scan angles between ±49.5° from the nadir. Radiances in 2378 spectral channels are measured in a continuous 90-element, ~1800 km-wide swath along the scan track. The horizontal spacing of these elements varying from around 13.5 km × 13.5 km at nadir to 41 km × 21.4 km at track-edge. In the along-track direction, the scan track is split into arbitrary 135-element along-track sections, referred to as granules, which correspond to 6 minutes of data collection. The spatial extent of these 135-element granules is shown in Fig. 1a.~~

~~In this study we use 3-D temperature measurements from AIRS observations derived using the retrieval scheme described by Hoffmann and Alexander (2009). This retrieval uses multiple 4.3 and 15 μm CO_2 spectral channels to produce estimates of stratospheric temperature that have a significantly higher vertical resolution than can be achieved using single channel radiances. Retrievals are carried out for each individual satellite footprint independently, improving the horizontal resolution of this retrieval by a factor of 3 in the along and across-track directions compared with AIRS operational data. Temperatures are retrieved on a 3 km vertical grid at a vertical resolution that varies between 7 — 14 km in the stratosphere (Hindley et al., 2019, their Fig. 2)~~

~~, compared to only $\gtrsim$ 14 km for single-channel operational AIRS data at 41 km altitude. This makes the 3-D temperature retrieval well suited to the study of small-scale processes such as gravity waves. At the latitudes and altitudes studied here, uncertainty in temperature measurement is typically $\lesssim$ 1.5 K (Hoffmann and Alexander, 2009; Hindley et al., 2019). Validation of the retrievals is described by Hoffmann and Alexander (2009) and Meyer and Hoffmann (2014).~~

~~To extract gravity-wave temperature perturbations, a 4th-order polynomial fit is subtracted from each across-track scan to remove slowly-varying background signals due to large-scale temperature gradients or planetary wave activity. This approach has been widely used in previous work (e.g. Wu, 2004; Alexander and Barnet, 2007; Hoffmann et al., 2014; Wright et al., 2017; Hindley et , although we acknowledge that some artefacts may remain, its use here ensures consistency with previous studies.~~

~~The sensitivity of our final AIRS temperature perturbations product to GWs is thus defined by the combination of both the retrieval averaging kernels and the detrending method. The full processing chain results in data that are sensitive to GWs with wavelengths longer than about 8 to 9 km in the vertical. In the horizontal, the sensitivity cutoff for short horizontal wavelengths dependent on the footprint size, which varies between roughly 15 km at nadir and 40 km for the outermost tracks. For longer horizontal wavelengths, sensitivity drops below 90% at horizontal wavelengths of 730 km and below 10% at 1400 km; thus, longer horizontal wavelengths will be more strongly attenuated than short ones, modifying the observed spectrum. The vertical and horizontal resolutions of the 3-D AIRS retrieval for different atmospheric conditions and sensitivity to gravity waves of varying scales can be seen in Fig. 2 of Hindley et al. (2019), Fig. 5 of Hoffmann et al. (2014) and the supplementary material of Ern et al. (2017).~~

~~There are typically two AIRS/Aqua overpasses per day over South Georgia. However, due to the precession of the Aqua orbit relative to the Earth's surface, the extent of the AIRS overpass swaths do not always cover the same regions every day. For comparison with the model domain around South Georgia, we only select AIRS overpasses where at least three out of four of the corners of the model domain are contained within the AIRS measurement swath during a given overpass, as shown in Fig. 1a. Due to the high inclination of the Aqua orbit, this usually results in 80 to 90% of the model domain being covered, and the area directly over the island itself is usually measured twice per day.~~

~~The two daily overpasses that meet these criteria usually occur at around 0300 UTC and 1700 UTC with the offset in timing usually less than 20 minutes from these times. However, there were several occasions where the AIRS swath at these times did not meet the three-corner criterion for the model domain that we set above, so we do not include these overpasses in our analysis.~~

2.1 Numerical modelling: local-area simulations over South Georgia

~~In this study~~ Here we use model output from specialised high-resolution runs of the UK Met Office Unified Model using the Even Newer Dynamics for General Atmospheric Modelling of the Environment (ENDGame) dynamical core (Davies et al., 2005; Wood et al., 2014). The model ~~is configured in a~~ consists of a nested high-resolution local-area domain 1200 km × 900 km around the island of South Georgia ~~,~~ and is run in a real-date configuration with lateral boundary conditions supplied by a ~~lower-resolution global forecastdescribed below~~global forecast.

The nested local-area domain simulation consists of an 800×600-pixel latitude-longitude grid centred at 54.5°S, 37.1°W, with 118 vertical levels from the surface to altitudes near 80 km. The simulations are run in a rotated-pole coordinate frame in order to provide latitude-longitude spacing that is close to Cartesian. This grid gives a horizontal spacing of roughly 1.5 km × 1.5 km, ~~which is significantly finer than the operational resolution of the Met Office global model. In the vertical, a damping layer is applied above 58.5~~for which the island's orography is well-resolved (Jackson et al., 2018). As described in Vosper (2015), a simultaneous run with a 750 m horizontal grid was also performed, but here we analyse the 1.5 km run due to computational constraints. Jackson et al. (2018) found no significant differences in the dominant stratospheric GW characteristics between these two runs, suggesting that the 1.5 km ~~altitude to suppress reflection effects~~grid is sufficient to resolve the main features of the islands orography.

~~Grid spacing between the model vertical levels~~ The vertical grid spacing of the local-area model increases from around 10 m near the surface to ~~3~~around 700 ~~km at 75~~m at 25 km altitude and 1.9 km at 55 km altitude ~~. There are 34 vertical model levels between altitudes of 20 to 60~~(Vosper, 2015, their Fig. 2). A damping layer is applied above 58.5 km ~~, where the AIRS measurements are most reliable. In this altitude range the model grid spacing increases from 0.6~~altitude to suppress reflection effects near the model top. Sensitivity tests for vertical grids of 70, 118 and 173 vertical levels were performed by Vosper (2015). They found a high degree of similarity between resolved zonal GW momentum fluxes in the 118-level and 173-level simulations from the surface to altitudes near 40 ~~km at 20 km altitude to 2.1~~ km~~at 60~~. Both of these configurations exhibited more realistic values than the 70-level simulation at high altitudes. Therefore, the 118 level configuration is selected to reduce the computational load and permit the use of a fine horizontal grid over the island. 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 altitude. This spacing is much finer than the 3 km vertical grid of the AIRS retrieval, so we expect that gravity waves at vertical scales that are visible to AIRS to be well simulated in the model.~~ 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).

~~Meteorological~~

Meteorological initial and lateral boundary conditions for the local-area domain are provided by a global N512 simulation with 70 vertical levels from the surface to altitudes near 80 km. At latitudes near South Georgia, this global model has a horizontal grid spacing of $\Delta x \approx 46$ km. This simulation is provided by Met Office operational analyses and re-initialised every 24 hours, providing hourly forecasts that supply lateral boundary conditions for the local-area configuration over South Georgia. At the edges of the local-area domain, these hourly forecasts are linearly interpolated in time to each ~~model time step. No gravity wave~~ timestep of the local-area model (30 seconds). As mentioned above, no orographic or non-orographic GW parameterisations were included in the local-area simulations~~, so the resolved wind and temperature fields are effectively free-running~~. Output fields were archived hourly. More information on the configuration of these simulations is described in detail in ~~Vosper (2015); Vosper et al. (2016)~~Vosper (2015), Vosper et al. (2016) and Jackson et al. (2018).

~~In this study we analyse data from two runs during austral winter, one covering the period~~ The model run used here is for two time periods: 1st to 31st July 2013 and ~~another for the period~~ 11th June to 8th July 2015. These austral wintertime

periods were chosen to coincide with the high probability of strong orographic GW forcing and deep vertical propagation due the strong prevailing winds at these latitudes during winter. A third model run for January 2015 was also ~~analysed but,~~ conducted and analysed, but due to the ~~weaker~~ weak stratospheric winds during austral summer, ~~very few gravity waves with vertical wavelengths long enough to be resolved by AIRS are found. Because of this, a comparison is not included here~~too few GWs (orographic or non-orographic) are visible in AIRS measurements for a meaningful comparison. Both model simulations during 2015 were designed to coincide with ~~a~~ summer and winter radiosonde campaigns on South Georgia (Moffat-Griffin et al., 2017; Jackson et al., 2018) that are described below.

~~The high spatial and temporal resolution of the local-area domain, coupled with the application of daily boundary conditions from Met Office operational forecasts should result in simulated conditions over South Georgia that are very close to reality for the given time periods. This presents an ideal opportunity to compare the simulated wave fields with co-located satellite measurements from AIRS.~~

~~To extract gravity wave temperature perturbations from the simulations, two runs were performed in parallel for each time period. In one run,~~

2.2 AIRS 3-D satellite observations

The Atmospheric Infrared Sounder (AIRS) (Aumann et al., 2003; Chahine et al., 2006) flies aboard NASA's Aqua satellite in a ∼100-minute near-polar sun-synchronous orbit. AIRS is a nadir-sounding hyperspectral radiometer that measures radiances in 2378 infrared spectral channels in a continuous 90-element, ∼ 1800 km-wide swath in the across-track direction at scan angles between ±49.5° from the ~~high-resolution topography for the island is included as normal (we refer to this as the SG configuration), while in the other run this topography is flattened to mean sea level (referred to as the nSG configuration). In previous studies, such as Vosper (2015); Vosper et al. (2016), output temperature fields from two runs such as these are differenced (SG − nSG) in order to reveal orographic gravity wave perturbations as a result of flow over the mountains. However, this approach will only reveal orographic gravity wave activity. This is because non-orographic wave activity will exist in both the SG and nSG runs and will thus be removed when the difference is taken. The AIRSobservations contain gravity wave perturbations from both orographic and non-orographic waves, so this is not desirable for our comparison.~~ nadir. The across-track horizontal spacing of these elements varies from around 13.5 km at nadir to 41 km at track-edge.

~~Therefore, in order to compare simulated gravity wave perturbations from both orographic and non-orographic waves in the model, non-orographic gravity wave activity in the nSG configuration must be removed before the difference is taken. To do this, we apply a 2-D horizontal smoothing filter to the output temperature fields of the nSG configuration. We found that a Gaussian smoothing filter with full-width-at-half-maximum (FWHM) equal to 300~~Here we use 3-D AIRS temperature measurements derived using the retrieval scheme of Hoffmann and Alexander (2009). This retrieval uses multiple 4.3 and 15 ~~km~~μm CO_2 spectral channels to produce estimates of stratospheric temperature for each individual measurement footprint on a 3 ~~×~~km vertical grid. For each height level, retrieved temperatures have a vertical resolution related to the kernel functions of the selected AIRS channels used, which varies between $7 - 14$ ~~225~~km for altitudes between 20 and 60 ~~km (one quarter of the model domain in each dimension) was sufficient to produce abroad, featureless background temperature field. This smoothed~~

version of the nSG configuration is then subtracted from the SG configuration to reveal temperature perturbations from both orographic and non-orographic wave activity. This process was only applied to altitudes above 15km (Hoffmann and Alexander, 2009; Hind . The retrieval is optimised for GW analysis, where a balance is achieved between retrieval noise and vertical resolution. At high southern latitudes during winter, temperature measurement error is typically $\lesssim 1.5$ km because it was found that artefacts were introduced in temperature perturbations due to synoptic systems in the troposphere. Since we only compare the model to AIRS observations above altitudes of K (Hoffmann and Alexander, 2009; Hindley et al., 2019). Validation of the 3-D AIRS temperature retrievals is described by Hoffmann and Alexander (2009) and Meyer and Hoffmann (2014).

There are typically two AIRS/Aqua overpasses per day over South Georgia, but due to the precession of the orbit, the locations of AIRS measurements during each overpass are not at the same geographic locations each day. For our study, we select only AIRS overpasses where the measurement swath covers at least three out of four corners of the local-area model domain, as shown in Fig. 1a. During the model runs in July 2013 and June-July 2015, we found that 39 and 48 AIRS overpasses respectively met this three-corner criterion, giving 87 co-incident 3-D AIRS measurements in total for our comparison. These overpasses occurred within $\pm$20 minutes of 0300 km, this does not affect our results significantlyUTC and 1700 UTC each day and measurements typically cover around 80% to 90% of the local-area model domain due to the high inclination of the AIRS/Aqua orbit.

2.3 Radiosondes

In this study we also make use of We also use wind measurements from a radiosonde campaign radiosonde campaigns that took place on South Georgia during June-July (austral January (austral summer) 2015 and June-July (austral winter) 2015. The details of this campaign and further analysis of these data are described by Moffat-Griffin et al. (2017), together with data from a campaign in austral summer during January 2015, the trajectories details of which are shown by the orange and green lines in Fig. 1b and 1c. Although both campaigns were analysed in this study, only data from the wintertime (June-July) campaign is shown due to the much lower levels of stratospheric gravity wave activity observed during austral summer.

described by Moffat-Griffin et al. (2017). Balloons were launched twice-daily from the British Antarctic Survey base at King Edward Point (54.3°S, 37.5°W), equipped with Vaisala RS92-SGP radiosondes, with additional launches timed to coincide with AIRS overpasses or when forecasts predicted strong winds suitable for GW generation. Meteorological and geolocation parameters are recorded at 2-second intervals during the flight.

The trajectories of the balloons are shown by the orange and green lines in Fig. 1b and 1c. 54 balloons were successfully launched during the wintertime period 13th June to 6th July .2015. Due to challenging local environmental conditions, 10 launches failed to reach the tropopause and only 20 reached altitudes of 25 km or above. It can During summer, nearly all of the 44 balloons launched reached their target altitudes near 35 km during January 2015. It can also be seen in Fig. 1that c that during winter the balloons travelled much further downwind to the east during winter than in summer due to the much stronger prevailing winds, with several balloons exiting strong westerly wintertime winds. Several balloons were blown so far that they even exited the eastern boundary of model domainbefore bursting., 600 km to the east, before reaching their final altitude. Wind measurements from these balloons are used to validate the direction and magnitude of the background wind in the local-area

model to assess conditions for orographic GW generation and propagation. A comparison for both the summer and winter campaigns was performed, but due to reduced stratospheric GW activity in the model during summer, only a comparison for the wintertime measurements is shown below.

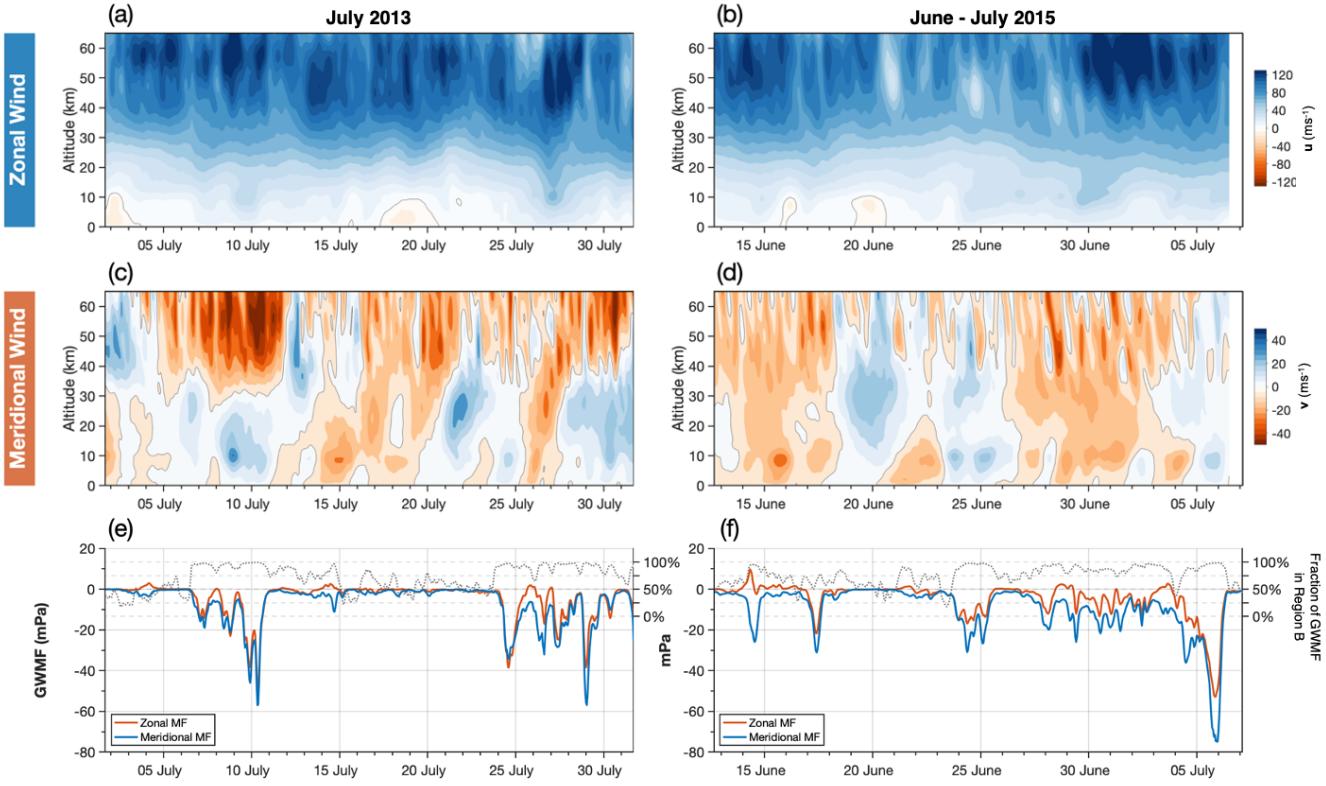

Figure 2. Hourly zonal and meridional wind speeds against altitude in the ~~local area~~ local-area model over South Georgia during July 2013 (a,c) and June-July 2015 (b,d) averaged over a horizontal region 600×400 km centred on the island (region C in Fig. 4). Panels (e,f) show average zonal (blue) and meridional (orange) gravity wave momentum fluxes (GWMF) $\overline{\rho}\left(\overline{u'w'}, \overline{v'w'}\right)$ over the same horizontal region but between 25 km and 45 km altitude. Positive (negative) values indicate eastward (westward) zonal GWMF and northward (southward) ~~directions~~ meridional GWMF. Dotted lines in panels (e,f) show the ~~zonal and meridional winds respectively~~percentage of the total model GWMF (right hand axis) downwind of the island (region B in Fig. 4), which is a strong indication of mountain wave activity.

~~Comparison of observed and simulated wind speeds during June-July 2015 from radiosonde observations and the local area model over South Georgia. Panel (a) shows the absolute model wind speed against height, with launch times and maximum altitudes of the radiosonde observations overlaid in black. Profiles of zonal (blue) and meridional (orange) wind against height for the radiosonde measurements and the model wind evaluated along each sonde trajectory are shown in panels (b,c) and (e,f) respectively. Thick black lines in panels (d,g) show the mean difference (Sondes — Model-as-Sondes) between the observed and modelled wind speeds for each height, with dark and light grey shading indicating one and two standard deviations respectively.~~

3 Model wind validation using co-located radiosonde measurements

3.1 ~~Validation of modelled winds using co-located radiosonde measurements~~

Before we compare our simulated ~~gravity-wave~~ GW fields to satellite observations, ~~it is first prudent~~ we first use our co-located radiosonde observations to validate the model wind fields~~against the co-located radiosonde observations. Since~~ . Surface wind flow over orography is ~~likely to be the dominant driver of gravity-wave activity directly~~ the key driver of mountain wave activity over the island (e.g Alexander and Grimsdell, 2013; Vosper, 2015; Moffat-Griffin et al., 2017; Jackson et al., 2018) and upper tropospheric and stratospheric winds determine the upward propagation of these orographically forced waves. Thus, model winds should first be tested to ensure they are a fair representation of reality before any GW investigations are undertaken.

~~Although the~~ The boundary conditions of the local-area model are initialised daily by Met Office operational analyses, but these winds are poorly constrained by conventional observations over the Southern Ocean, relying largely on temperatures nudged by assimilated satellite radiances. Wright and Hindley (2018) showed that a lack of observations can result in significant stratospheric biases in this region in global models. ~~Thus,~~ The radiosonde measurements described here are not assimilated into the operational analysis. Thus, to our knowledge, these radiosonde observations ~~represent~~ are the only coincident ~~measurements available with which to accurately assess the low-altitude~~ and independent wind measurements available to assess the tropospheric wind fields in the model over the island during our period of study.

Figure ~~??~~ 2 shows hourly zonal and meridional wind against height for the two model runs during July 2013 and June-July 2015. These values are horizontally averaged over the whole model domain, so are representative of the large-scale background flow. As would be expected for a wintertime study at these latitudes, wind speeds in the zonal direction are eastward and generally increase strongly with height, with values reaching 120 ms^{-1} above 50 km altitude. ~~Variability in zonal wind speed in the stratosphere is closely related to the changing latitude of the centre of the stratospheric polar vortex during winter.~~ In the meridional direction, frequent changes between northward and southward flow are observed, with speeds reaching values near $\pm 40\ \mathrm{ms}^{-1}$ above 40 km altitude. ~~In one case near 10th July 2013, a wind reversal with height is seen at around 30 km.~~ Gravity wave activity in the model for this time period is shown in panels (e,f) discussed later in Sect. 4.

To compare ~~these wind fields~~ the model winds to radiosonde observations, each radiosonde trajectory is traced through the hourly model winds fields. ~~The modelled wind speed along this trajectory is compared to the observed radiosonde winds~~Because of the large horizontal distances travelled by the radiosondes (up to 600 km), and the length of the flight times (up to around 2.5 hours) it is necessary to evaluate the hourly model data along a path that varies in horizontal space, height and time. To do this, all model timesteps are loaded for the duration of each radiosonde flight, including one timestep before and after, and 4-dimensional linear interpolants (~~x,y,z,t~~x, y, z, t) of zonal u and meridional v wind fields are constructed. These interpolants are then evaluated for each point along the radiosonde's trajectory using the measured time, height and location ~~information~~of the balloon. This approach allows us to compensate for any time-varying effects in the model wind speeds during the radiosonde flights.

~~It is important to mention at this point that caution should be taken when measuring gravity wave momentum fluxes from slanted vertical profiles through mountain wave fields (such as radiosonde measurements here). As discussed by Vosper and Ross (2020)~~

~~, the usual assumptions required for the measurement of vertically-integrated momentum fluxes of planar monochromatic waves do not hold true for mountain waves sampled with a slanted vertical profile. For this reason we do not conduct a gravity wave comparison between the model and the radiosonde measurements here and instead only use the radiosonde measurements to validate the model winds.~~

The model winds along the radiosonde trajectories are then compared to the radiosonde wind observations themselves. Figure 3a shows the results of our wind comparison. Radiosonde launch times (UTC) and maximum recorded altitudes during the winter campaign are shown by the black lines and circles in Fig. 3a. ~~Also shown in this panel is~~ For illustration, the mean zonal wind speed over the modelling domain against altitude in also shown on panel (a), which gives us an indication of the ~~local~~ background wind conditions through which the balloons travelled.

As can be seen in Fig. 3a, several of the radiosonde balloons did not reach their desired altitudes near 30 km, instead bursting soon after launch. This was usually due to the extreme weather conditions at low altitudes during the fieldwork campaign, as reported by the radiosonde launch team. In some cases, surface winds were so strong that radiosonde balloons did not ascend fast enough to exit the bay around the launch site, colliding instead with the slopes of nearby mountains.

Panels (b-g) in Fig. 3 show the ~~results of this analysis, where "Sondes" indicates the~~ measured radiosonde wind speed ~~and "Model-as-Sondes"indicates the modelled~~ ("Sondes") and the model wind speed evaluated along the radiosondes path ("Model-as-Sondes") in the zonal and meridional directions. The two ~~data sets~~ datasets are in good general agreement, with measured and simulated zonal winds in Figs. 3(b-c) increasing from a few metres per second near the surface to around 60 ~~ms^{-1} around~~ ms^{-1} near 30 km altitude. In the meridional direction, both ~~data sets~~ datasets show wind speeds between around ± 15 ms^{-1} with little variation ~~in~~ with altitude in Figs. 3(e-f). ~~In both directions, the~~ The radiosonde measurements are found to exhibit more small-scale variability than the model fields, likely due to small-scale wave or turbulence features and measurement errors which are not present in the model. ~~Measurement artefacts from balloon bursts are also visible, in addition to instances~~ Some instances are also found where sonde measurements are ~~shown~~ present but no model-as-sonde data is available~~due the balloons~~ , which is due to the balloons horizontally exiting the model domain ~~.~~ (see Fig. 1c).

To further compare the simulated and measured wind speeds, the difference between the sonde and the model-as-sonde winds (sondes~~minus model-as-sondes~~ – model) against altitude is shown in Figs. 3(d) and (g) for the zonal and meridional directions respectively. Dark and light grey shaded regions show one and two standard deviations of all differences respectively, while the thick black line shows the mean difference for the June-July 2015 run.

In the zonal direction, the ~~campaign-mean~~ time-average difference in wind speed is ~~reasonably close to zero~~ less than 5 ms^{-1} for most altitudes ~~, and lowest~~ above 10 km, and close to zero in the low to mid-stratosphere ~~. Largest~~ between 15 and 25 km altitude. The largest differences between the sonde and model-as-sonde winds are seen between altitudes ~~of 3 to~~ below 10 km in Fig. 3d. This is near the tropopause~~region~~, and could suggest ~~significant~~ that short-timescale variability of the tropospheric jet observed over the island ~~which is not~~ is not so well represented in the model. This ~~may~~ could influence the upward propagation of mountain waves. Near the surface, below altitudes of around 3 km, a slight bias towards stronger zonal winds in the model ~~in observed, which could be due to local topographic effects around~~ is observed. We suspect that this is due to slight under-representation of the "roughness" of the complex local topographic features around the launch site in the model.

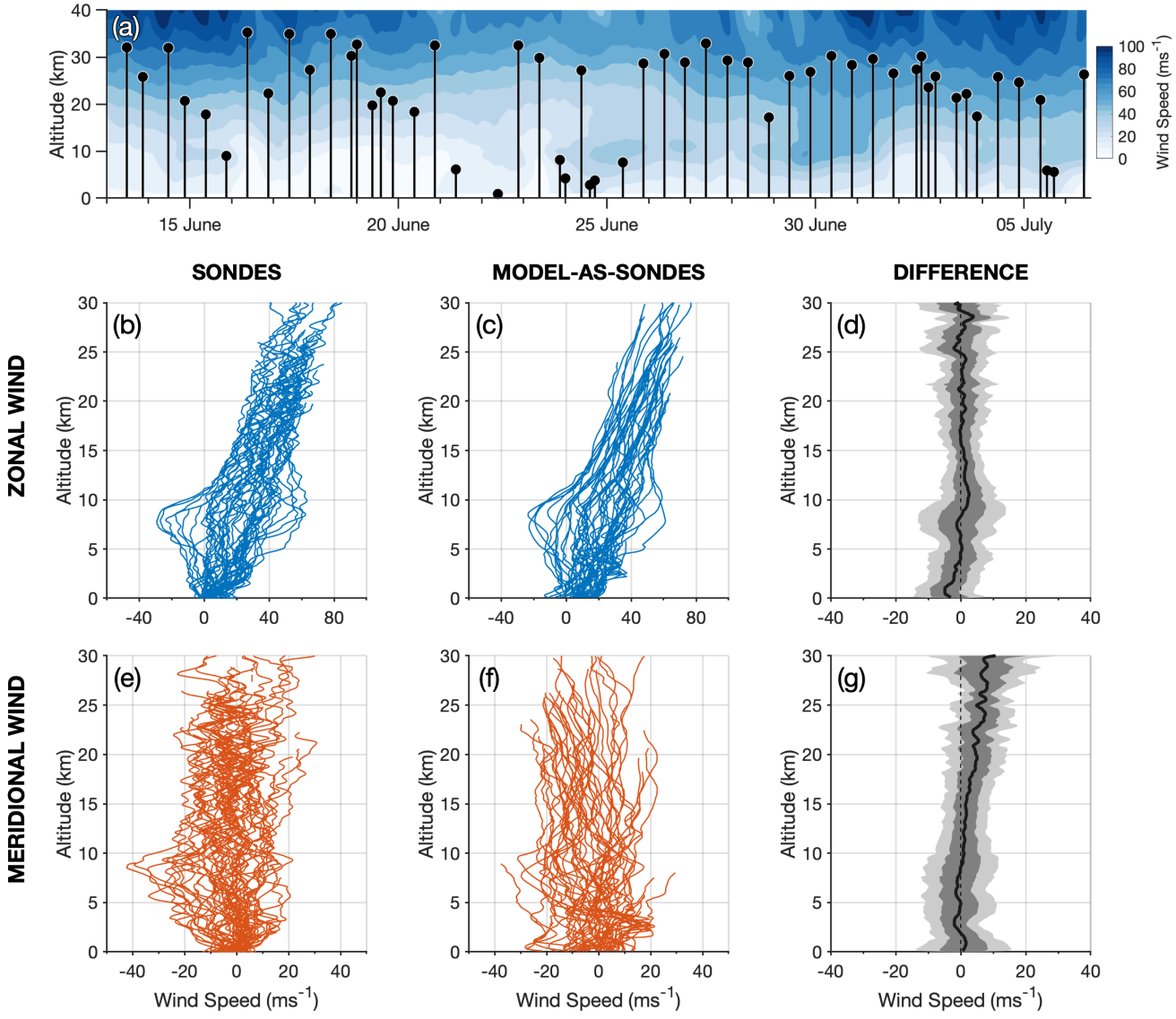

Figure 3. Comparison of wind speeds from the local-area model to coincident radiosonde observations launched from South Georgia during June-July 2015. Panel (a) shows launch times and maximum altitudes of the radiosonde observations (black lines), while coloured contours show the magnitude of the model wind speed for illustration. Panels (b,e) and (c,f) show profiles of zonal (blue) and meridional (orange) wind against height for the radiosonde measurements and the model wind, where the model wind has been evaluated along each radiosonde's trajectory. Panels (d,g) show the mean difference (thick black line) between the radiosonde and model wind speeds (Sondes – Model-as-Sondes) for each height, while dark grey and light grey shading indicates one and two standard deviations respectively.

King Edward Point ~~.~~ is located in a sheltered bay 2 km east of the main mountain ridge of the Thatcher Peninsula, which peaks at nearly 2 km high. At the 1.5 km model horizontal resolution used in this study, this mountain ridge will be at most one model grid cell away from the launch site. Thus, accurately simulating surface winds at this site will be quite challenging. Further, the model winds are not well constrained by surface observations in the area, so small surface biases are to be expected.

In the meridional direction, the time-average wind speed differences ~~between the sondes and model-as-sondes~~ are generally less than 10 ~~ms^{-1} .~~ ms^{-1} in Fig. 3g. However, a clear positive difference is observed above around 15 km altitude which increases to near 10 ms^{-1} ~~around~~ at 30 km altitude. This ~~positive difference corresponds to slight southward directional bias in the simulated model winds. That is, the meridional components of the sonde-measured winds are greater (more northward) than the meridional components of the model-as-sonde winds . This could be indicative of~~ indicates that the model slightly overestimates (underestimates) 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 ~~in the global forecast that provides the boundary conditions for the local area model over South Georgia. At these high southern latitudes,~~ but also in a small positive bias in the net horizontal wind speed. Given that global models are very poorly constrained by ~~ground-based conventional measurements, so a small bias is perhaps could be expected~~conventional observations at high southern latitudes, this directional bias is actually quite reasonable. While we do not expect this ~~bias~~ to affect our results significantly, we acknowledge that a difference in the rotation of the simulated wind vector ~~with increasing altitude~~ compared to reality could have an effect on ~~the propagation and measured structure of any mountain wave field that forms~~ wave propagation and thus the measured orientations of simulated mountain waves over the island.

It should be mentioned that some of the differences between the model and model-as-sonde winds could be due to timing or lag issues in the model, such as in the arrival of synoptic systems. Anecdotal reports from the radiosonde launch team on South Georgia suggested that the arrival of synoptic systems such as fronts and weather systems could differ from the Unified Model forecast by several hours. Although these ~~phenomena are generally located in the troposphere~~are tropospheric phenomena, they may have a stratospheric response ~~which may be~~ that is earlier or later than predicted. These would manifest as pseudo-random errors in our analysis, which could explain some of the spread in the wind speed differences. ~~Further, positional errors in the radiosonde measurements could lead to further spread, but these factors are unlikely to lead to the systematic biases reported here.~~

Aside from these ~~minor differences~~ differences however, we conclude that ~~, on a climatological level,~~ overall the model wind speed and direction over the island is ~~reasonably well simulated~~ simulated reasonably well during the June-July 2015 campaign. ~~Large differences in surface wind speed, for example, may also be indicative of difficulty in simulating the local topographic environment of the launch site rather than simulating the island as a whole. The British Antarctic Survey base at King Edward Point, from which the balloons were launched, is located in a sheltered bay 2 km east of the main mountain ridge of the Thatcher Peninsula, which is nearly 2 km high. At the 1.5 km model horizontal resolution used in this study, this mountain ridge will be at most one model grid cell away from the launch site. Thus, accurately simulating surface winds at this site will be quite challenging.~~

Caution should be taken when measuring gravity wave momentum fluxes from slanted vertical profiles through mountain wave fields (such as radiosonde measurements here). The usual assumptions required for the measurement of vertically-integrated momentum fluxes of planar monochromatic waves do not hold true for mountain waves sampled with a slanted vertical profile (e.g. de la Torre and Alexander, 1995; de la Torre et al., 2018; Vosper and Ross, 2020). For this reason, we do not conduct a GW comparison between the model and the radiosonde measurements here and only use the radiosonde measurements to validate the model winds.

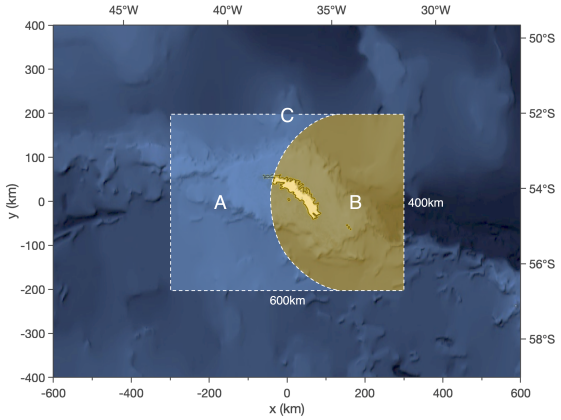

Figure 4. Illustration of the two regions to the east and west of South Georgia used to produce the values in Table 1. Region A is upwind of the island and Region B is over and downwind of the island. The two regions have equal area.

4 Gravity waves over South Georgia in the full-resolution model

After validating of the simulated winds in our local area model, we now consider simulated GW activity in the model. A key quantity in GW research is the vertical flux of horizontal pseudo-momentum, generally referred to as momentum flux. This property helps to quantify the vertical transfer of horizontal momentum by GWs. When a GW breaks, horizontal momentum will be deposited in the mean flow, resulting in a drag or driving effect on the background wind. Measuring and quantifying the momentum fluxes of mountain waves from small, isolated islands is an important area of current research (McLandress et al., 2012; Alexander and Grimsdell, 2013; Garfinkel and Oman, 2018; Jackson et al., 2018).

5 ~~Comparing gravity wave characteristics in the model to 3-D AIRS measurements~~

~~In order to make a fair comparison between gravity waves in the model and in the 3-D satellite observations, we must first ensure that the spectral range of gravity waves in each data sets is comparable. This involves re-gridding the model and observations onto a common regular grid and applying the AIRS observational filter to the model output.~~

~~The high resolution model outputs hourly temperature fields on the model grid for the July 2013~~ Figures 2(e,f) show zonal and meridional gravity wave momentum flux (GWMF) averaged between 25 km and ~~June-July 2015 modelling campaigns. During these campaigns, 39 and 48 3-D AIRS measurements~~ 45 km altitude and over a horizontal area 600×400 km centred on the island, denoted by region C in Fig. 4. Here, zonal GWMF F_x and meridional GWMF F_y are calculated as

$$(F_x, F_y) = \overline{\rho}\,(\overline{u'w'}, \overline{v'w'}) \tag{1}$$

where ρ is the background atmospheric density, u', v', w' are wind perturbations in the zonal, meridional and vertical directions and the overbar denotes an area average over GW scales Fritts and Alexander (2003); Ern et al. (2004). This relation is valid for mid-frequency GWs, where the intrinsic frequency $\hat{\omega}^2 \gg f^2$, where f is the inertial frequency. Wind perturbations u', v', w' are extracted from the background flow by subtracting a 4$^{\text{th}}$-order polynomial fit in the zonal direction. This ensures for reasonable consistency with the method used for the AIRS satellite observations described in Sect. 2.2, but the two methods are not identical and therefore should be considered separately.

Zonal and meridional GWMF timeseries in Figs 2(e,f) indicate that stratospheric GW activity over the island ~~region for~~ in the full-resolution model is intermittent, with bursts of GWMF up to around 60 mPa occurring during 7-11th July 2013~~and 2015 respectively were found to meet our selection criteria in Sect. 2.2, giving 87 co-incident AIRS measurements in total.~~, 24-30th July 2013 and 4-6th July 2015. These bursts of GWMF generally coincide with periods of increased winds speeds from the surface through to the mid-stratosphere, as shown in Figs. 2(a-d). This is indicative of strong mountain wave forcing by the surface winds and strong upper tropospheric and stratospheric winds combining to provide good conditions for mountain wave propagation to greater heights. Indeed, during periods where surface zonal winds in Figs. 2(a,b), stratospheric GWMF in Figs 2(e,f) is low.

~~In this section, we first describe how the AIRS measurements and the model output are re-gridded onto acommon regular grid. The AIRS observational filter is then applied to the model to compare the simulated wave field as if it were observed by AIRS. We then apply a 3-D Stockwell transform (3DST) to the AIRS measurements,~~ The average zonal direction of GWMF is generally westward, which is consistent with what we would expect for a mountain wave propagating against the background zonal wind in Fig. 2a. Interestingly, the area-average meridional GWMF is generally southward, regardless of the direction of the background meridional wind. The distribution of GWMF around the island, shown later in this study, indicates that this is because to the southward component of the characteristic GW pattern over the island (e.g. Alexander et al., 2009; Alexander and Grimsdell, is considerably larger than the northern component, due to the ~~model output and~~ orientation of the island.

Dotted grey lines (right hand axes) in Figs. 2(e,f) show the percentage of the total absolute GWMF $\left(F_x^2 + F_y^2\right)^{\frac{1}{2}}$ in region C contained within the region B, located downwind of the island as shown in Fig. 4. Regions A and B have areas equal to half of Region C, so a value of 50% indicates a uniform distribution of GWMF between the upwind and downwind regions to the west and east of the island. A fraction larger than 50% indicates more GWMF in the downwind region, which is a strong indication of mountain wave activity. It can be seen that during nearly all of the periods of increased GWMF in the model, this fraction is close to around 75 to 100%, which suggests that mountain waves are the dominant source of GW activity in the local-area model. This fraction rarely falls below 50%, and when it does it is during periods of low GWMF. This suggests that, relatively,

non-orographic GW activity makes only a small contribution to the GWMF in the ~~model at the AIRS resolution to measure gravity wave amplitudes, wavelengths, directions and momentum fluxes in each dataset~~local-area model at full-resolution.

4.1 ~~Applying the observational filter of AIRS to the model output~~

5 Applying the AIRS observational filter to the model

~~A key concept in the study of GWs is the~~ The GWMF results in the previous section indicate significant GW activity in the full-resolution model. But these results cannot be directly compared to AIRS satellite measurements, because GW measurements in AIRS are subject to the AIRS observational filter. ~~That is , no single instrument~~ The observational filter (Preusse et al., 2002; Alexander and Barnet, 2007) is a key concept in GW observations. No single instrument or technique can measure the full ~~gravity wave spectrum(e.g. Preusse et al., 2002; Alexander and Barnet, 2007)~~GW spectrum. For example, ~~a~~ the standard retrievals of nadir-sounding instrument such as AIRS will ~~have, in general,~~ generally have relatively low vertical resolution ($\Delta Z \sim 15$–20 km) for GWs in the stratosphere but relatively high horizontal resolution ($\Delta L \sim 50$–100 km). In contrast, limb-sounding instruments and techniques such as HIRDLS (e.g. Gille et al., 2003) or GPS radio occultation (e.g. Kursinski et al., 1997) will have relatively high vertical resolution ($\Delta Z \sim 1$ km) but relative low horizontal resolution ~~.~~($\Delta L \sim 150$–270 km). To make a fair comparison between GWs in our local-area model and coincident AIRS satellite observations, we must ensure that both datasets have the same observational filter.

~~Thus, in the present study, a simple direct comparison of gravity wave properties between the full resolution model and~~ For satellite observations, the observational filter is primarily dependent upon two things: sampling and resolution (Wright and Hindley, 2018). Below, we describe how we apply the sampling pattern and resolution of the AIRS ~~measurements is not particularly meaningful due to their different observational filters. To obtain a fair and meaningful comparison, we can apply the observational filter of AIRS to the model to make~~ observations to the local-area model to create a "model-as-AIRS" dataset that is ~~more spectrally comparable to AIRS measurements. Although the AIRS observational filter is sensitive to waves with larger horizontal and vertical scales than the model is capable of simulating, this approach is still useful. This is because, especially for the case of mountain waves over South Georgia, the accurate simulation of the large scale waves visible to AIRS is critically dependent on the accurate simulation of wave generation mechanisms over the island, which can occur on the very smallest scales.~~ comparable to the satellite observations.

5.1 Horizontal Sampling

To ~~produce~~ create the model-as-AIRS dataset ~~, we first regrid the model and AIRS data onto a common spatial grid, and then convolve the model data with a height-varying 3D Gaussian filter approximating the true observational filter of AIRS~~for our comparison to AIRS observations, we use hourly temperature output fields from the local area model. As described above, model temperature fields are on a 1.5 km horizontal grid, with 118 vertical levels from the surface to near 70 km altitude.

~~For ease of explanation, we refer to these three datasets as~~

- ~~**AIRS**: the AIRS measurements described in Sect. 2.2,~~

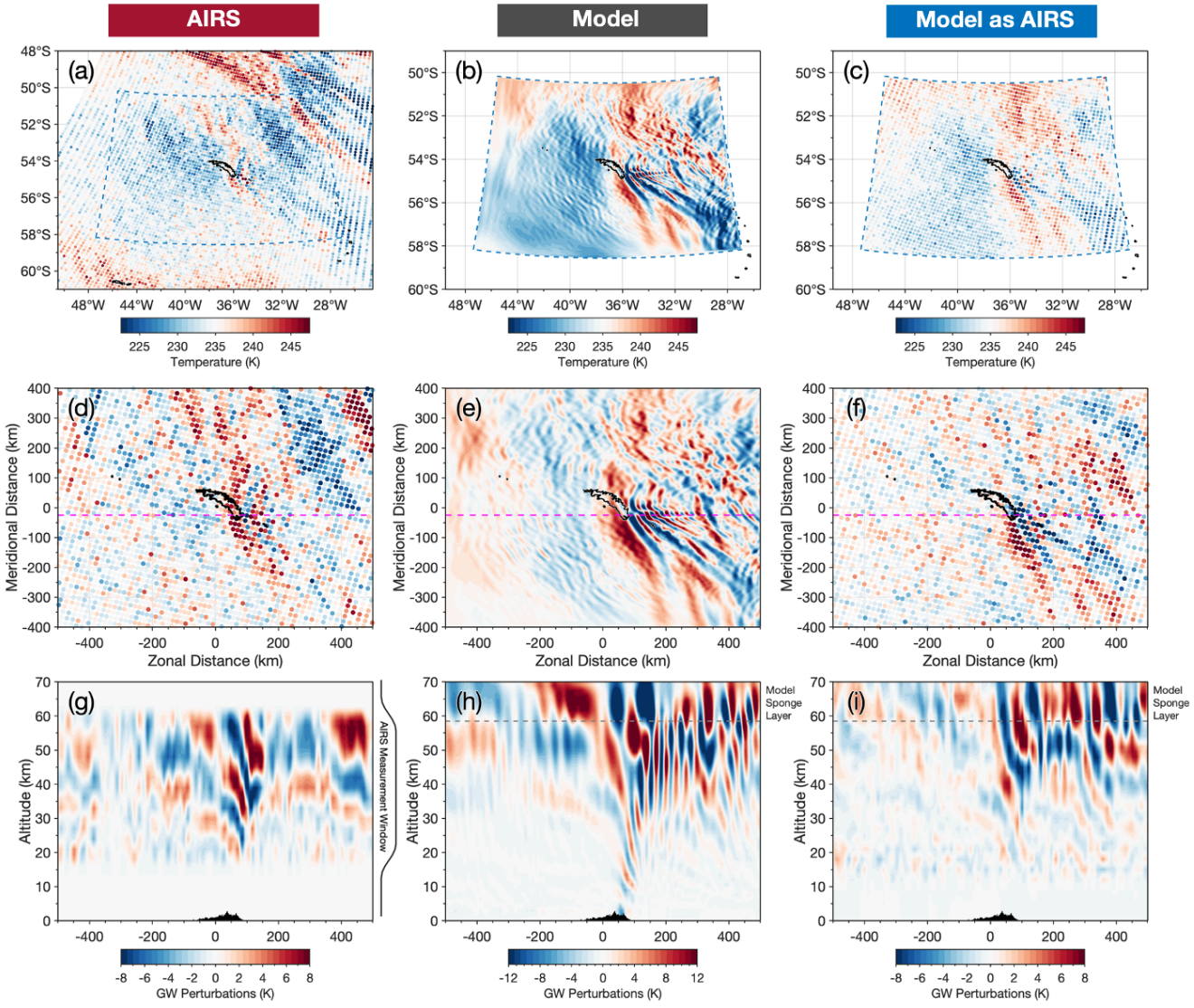

Figure 5. ~~Vertical~~ Observed and modelled temperatures (top row) and ~~horizontal~~ temperature perturbations (~~bottom~~ middle row) ~~cross-sections of observed and modelled temperature perturbations~~ at 45 km altitude over South Georgia at 0300 UTC on the 5th of July 2015 ~~for the regridded~~ in AIRS measurements, the full-resolution local-area model and the model-as-AIRS. ~~Vertical cross-sections are taken at~~ Coloured circles in (a~~meridional distance of $y = 0$ km~~ ,c) and ~~horizontal cross-sections are shown at an altitude of 45 km. Here~~(d,~~model-as-AIRS refers to~~ f) indicate the ~~model temperature perturbations with~~ size and locations of the AIRS ~~observational filter applied~~measurement footprints. ~~Different colour scales are used for~~ The bottom row shows vertical cuts through the ~~model, AIRS and model-as-AIRS so that wave structure can be seen clearly~~ temperature perturbations along the pink dashed line in ~~each data set~~panels (d-f). See ~~text in~~ Sect. ~~??~~ 5 for details on the model-as-AIRS data.

- the **model**: the high resolution model output described in Sect. 2.1, and

- the **model-as-AIRS**: the model output fields as if they were observed by AIRS , described in this section.

The data processing steps used to produce these three comparable datasets are described below. First, a common horizontal and vertical grid is specified for all three datasets. This grid must be regularly spaced and Cartesian in order for spectral analysis techniques to produce meaningful results. The chosen grid is centred at 54.5°S, 37.1°W (like the model grid) and is 1200 The first step is to simulate the AIRS horizontal footprint and sampling pattern. The AIRS sampling pattern is well illustrated in Hoffmann et al. (2014, their Fig. 2). AIRS measurements are made on a 90-element wide horizontal across-track swath, where each measurement footprint is approximately 13.5×13.5 900km wide (Hoffmann et al., 2014, their Table 1). The horizontal sampling distance between the centres of these footprints increases with increasing distance from the nadir from around 13.5 ×75 km in the zonal, meridional and vertical directions, with a grid spacing of 15km to 42 km , 15near the track edge, so it is important to consider this for GWs with relatively short horizontal scales, such as those expected directly over South Georgia.

To simulate the AIRS measurement footprints in the model, each vertical level of each model temperature timestep is convolved with a horizontal Gaussian function with full-width-at-half-maximum (FWHM) equal to 13.5×13.5 km and 1.5 km respectively. This is generally finer than the resolution km. We then interpolate the smoothed model temperatures onto the horizontal sampling grid of the AIRS measurements but much coarser than the model overpass that is closest in time to each model timestep. The Gaussian smoothing step above ensures that this is a reasonable approximation to the horizontal sampling of an AIRS measurement footprint wherever the model is sampled. This gives us model temperatures at the horizontal sampling and resolution of the nearest coincident AIRS overpass to each hourly model timestep. Since we wish to apply the AIRS horizontal resolution to the model , it is not necessary to have much finer resolution than AIRS .

Each granule or time step

5.2 Vertical Resolution

Next, we consider the vertical resolution of the AIRS measurementsand model output is linearly interpolated onto this regular grid. This produces the AIRS and model datasets used throughout this study. All analyses and figures use these regridded data. In order to achieve a consistent horizontal resolution across the AIRS measurements, a horizontal Gaussian smoothing filter with a FWHM equal to 40. To apply this vertical resolution to the model, we first need to interpolate the model on to a regular vertical grid. The chosen grid is from the surface to 75 km in both the x and y directions is applied to each vertical layer. This also helps to reduce the impact of unwanted pixel-scale noise in AIRS, as discussed in previous studies (Hindley et al., 2016, 2019), and is consistent with our model-as-AIRS processing steps belowaltitude in 1.5 km steps. This grid spacing is finer than the model vertical grid in the stratosphere, but coarser in the troposphere. Because our comparison to AIRS measurements takes place in the stratosphere, this choice will not significantly affect our results.

Since the AIRS temperature retrieval has reduced vertical resolution and accuracy outside the height range 20 to 60 km altitude (Hoffmann and Alexander, 2009), we set AIRS data outside this range to zero and apply a half-bell tapering window

to the upper and lower boundaries (see Fig. **??**, discussed below). This minimises any impact of edge effects in subsequent spectral analysis.

The average vertical resolution The vertical resolution of the 3-D AIRS temperature measurements variesbetween around 3-D AIRS retrieval for different atmospheric conditions is shown in Fig. 2 of Hindley et al. (2019), where resolution values are derived using the approach of Hoffmann and Alexander (2009). The vertical resolution varies, on average, between 7 to 14 km between altitudes of 20 to and 60 km (Hoffmann and Alexander, 2009). The full variation of vertical wavelength with altitudecan be seen inFig. 5 of Hoffmann and Alexander (2009) and Fig. 2 of Hindley et al. (2019). The horizontal resolution of the AIRS measurements is taken to be 15 km, which is derived from the average horizontal spacing of the measurement grid.

To altitude. Using the values shown by Hindley et al. (2019), we apply the AIRS vertical resolution to the model output, we consider each vertical layer separately. For each vertical layer, the re-gridded model temperature perturbations are temperature fields. This is a step-by-step process which involves the convolution of the model temperatures with vertical Gaussian functions with different FHWMs for each altitude. For example, the vertical resolution at 30 km altitude is approximately 7.5 km (Hindley et al., 2019, their Fig. 2b) so the full 3-D temperature volume is convolved with a 3-D Gaussian smoothing kernel with a vertical Gaussian function with FWHM equal to 407.5 km in both horizontal directions. This is approximately the Nyquist limit of the AIRS measurements, that is, twice the average AIRS horizontal spacing or around 20and the horizontal level at 30 km . The FWHM in the vertical is taken to be the vertical resolution for that vertical layer (see Fig. 2b of Hindley et al. (2019)). This vertical layer is extracted from the smoothed model to produce one layer of the model-as-AIRS dataset. The unfiltered model is then convolved again in 3-D with the next vertical resolution to produce the next vertical layerof the model-as-AIRS dataset. The rest of the vertical layers are then built up one by one using this method, which is then applied to each hourly model timestep for both the July 2013 and the June-July 2015 modelling campaigns. altitude is then extracted and stored separately. This Gaussian-filtering approach is less accurate than using the the full retrieval algorithm of (Hoffmann and Alexander, 2009) or the fine-resampling algorithm of (Wright and Hindley, 2018), however it allows us to produce the observationally-filtered data for this study at much lower computational cost over applying the full 3-D AIRS retrieval scheme on the high-resolution model data. process is performed for each altitude level allowing us to build up a smoothed temperature field, layer by layer, for each model timestep. The result of this procedure is a 3-D volume of model temperatures sampled on the AIRS horizontal scan track and smoothed to the AIRS vertical resolution.

5.3 Retrieval noise

Because the horizontal andvertical grid spacing of the local-area model is much finer than retrieval grid of the AIRS measurements, the sensitivity to gravity waves of the AIRS and model-as-AIRS should be dominated by the AIRS observational filter and regridding approaches that we have applied here. Since these approaches are the same for both data sets, we expect the resulting portions of the gravity wave spectrum visible in the AIRS Finally, we consider the effect of AIRS retrieval noise. Noise in AIRS measurements can arise due to thermal noise in the AIRS instrument and/or deviations of the atmospheric state from local thermodynamic equilibrium, which is assumed in the retrieval (Hoffmann and Alexander, 2009). These factors vary for

different spectral channels in the AIRS instrument, and as a result the estimated retrieval noise varies between 1.2 K and 1.5 K between 25 and 45 km altitude, as shown in Fig. 2a of Hindley et al. (2019) and Fig. 5 of Hoffmann and Alexander (2009) . However, because the retrieval noise is pseudo-random and incoherent in the horizontal, coherent wave features at large horizontal scales with amplitudes slightly below these noise values can be detected under reasonable conditions (Hindley et al., 2019) . In the general case however, we cannot routinely separate retrieval noise from GW perturbations in AIRS measurements, and so to rule out the possibility of retrieval noise affecting our comparison we add specified AIRS retrieval noise to our model-as-AIRS~~data to be comparable~~.

~~Gravity wave temperature perturbations over South Georgia at 1700 UTC on 5th July 2015 for (a) AIRS, (b) model and (c) model-as-AIRS. Here, values are shown as red and blue isosurfaces after a factor of $\kappa(z) = \exp\left(\frac{z-z_{\text{ref}}}{2H}\right)$, where $z_{\text{ref}} = 40$ km is a reference altitude and $H = 7$ km is the scale height of the atmosphere, has been applied in order to see the vertical structure of the wave field clearly. Blue and red dashed lines in each panel indicate the boundaries of the model domain and the AIRS measurements respectively.~~

5.4 ~~Results of applying the AIRS observational filter to the model~~

~~The effect of applying the AIRS observational filter~~ To apply the AIRS retrieval noise to the model~~is illustrated in Figure **??**.~~ ~~For~~ , we select an AIRS overpass at ~~03:11~~1700 UTC ~~5th July~~ on 20th June 2015 ~~, the regridded temperature perturbations of the AIRS, model and model-as-AIRS data for are plotted as vertical (top row) and horizontal (bottom) cross-sections at $y = 0$~~ (granule numbers 174 and ~~45~~175) containing no discernible wave features at any altitude level. Once the background temperature is removed using the method below, the residual perturbations exhibit an approximate standard deviation of around 0.5 K at 39 km altitude~~in our regular grid. The model timestep shown is 03:00 UTC on the same day, so the time separation is very small and~~ . For each altitude level, the ~~local wind conditions can be expected to be very close to reality.~~

~~The effect of restricting the vertical extent of the AIRS data (as mentioned above) is illustrated in Fig. **??**a , where low-quality measurements outside of altitudes between around 20 and 60 km are removed. The black line to the side of the panel shows the extent of the usable vertical measurement window.~~ residual noise perturbations from this overpass are randomised and then added to the model temperature fields for each model timestep to simulate AIRS retrieval noise. The use of synthetic random Gaussian noise was considered for this purpose, but since AIRS noise characteristics vary with altitude we found that using genuine AIRS noise provided more realistic results.

~~Clear features consistent with gravity wave temperature perturbations are visible in each dataset in Fig. **??**. In AIRS measurements and the full resolution model in Figs. **??**a and **??**b, westward-sloping phase fronts characteristic of a mountain wave field are observed over the island. Longer vertical wavelengths are observed directly over the island itself. The topography of South Georgia is shown in black at the base of the figure, and is to scale.~~

~~As might be expected, the modelled gravity wave field exhibits much more fine horizontal structure than the AIRS measurements due to the higher resolution. These fine horizontal scale structures in the model also exhibit much higher gravity wave amplitudes than those observed by AIRS. The spectral regimes observed by the two datasets also are different.~~

~~Once the AIRS observational filter is applied to the model in Fig. ??c, much of the fine horizontal structure in the modelled gravity wave field is removed and the wave amplitudes are more comparable to those observed in AIRS.~~

~~The horizontal cross-sections at an altitude of 45~~

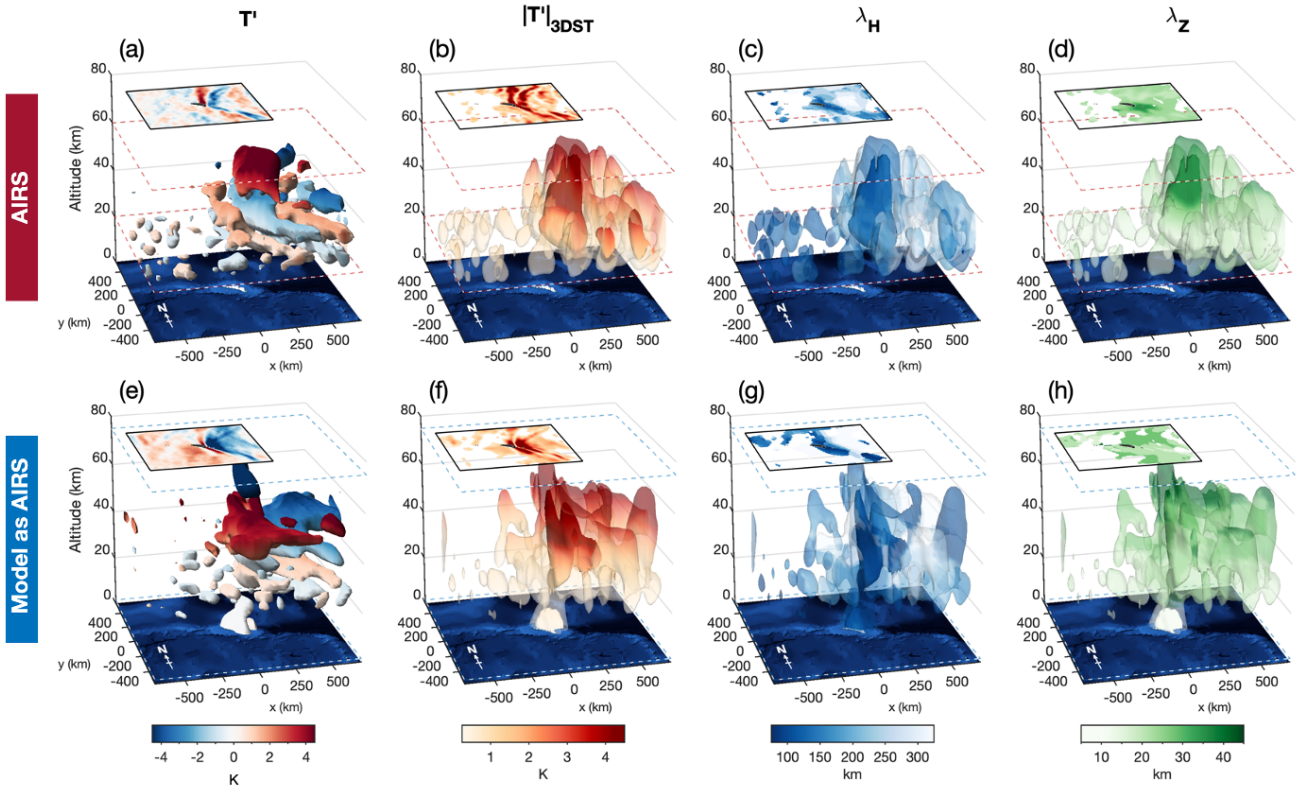

Figure 6. 3-D S-transform (3DST) analysis of temperature perturbations from AIRS satellite observations (top row) and the model-as-AIRS (bottom row) over South Georgia for 1700 UTC on 5th July 2015. Coloured isosurfaces in panels (a,e) show the AIRS and model-as-AIRS temperature perturbations T', while panels (b,f), (c,g) and (d,h) show 3DST-measured absolute wave amplitude $|T'|_{3DST}$, horizontal wavelength λ_H and vertical wavelength λ_Z respectively. Blue and red dashed lines denote the upper and lower boundaries of the model domain and the AIRS measurements respectively. Horizontal cuts through the data at 40 km altitude are shown in the top left hand corners of each panel, which share a colour scale with the isosurfaces.

6 Measuring 3-D gravity wave properties

To investigate the properties of the GWs over South Georgia in our AIRS and model-as-AIRS datasets, we first extract GW temperature perturbations from the background, then we measure GW properties using 3-D S-transform spectral analysis technique.

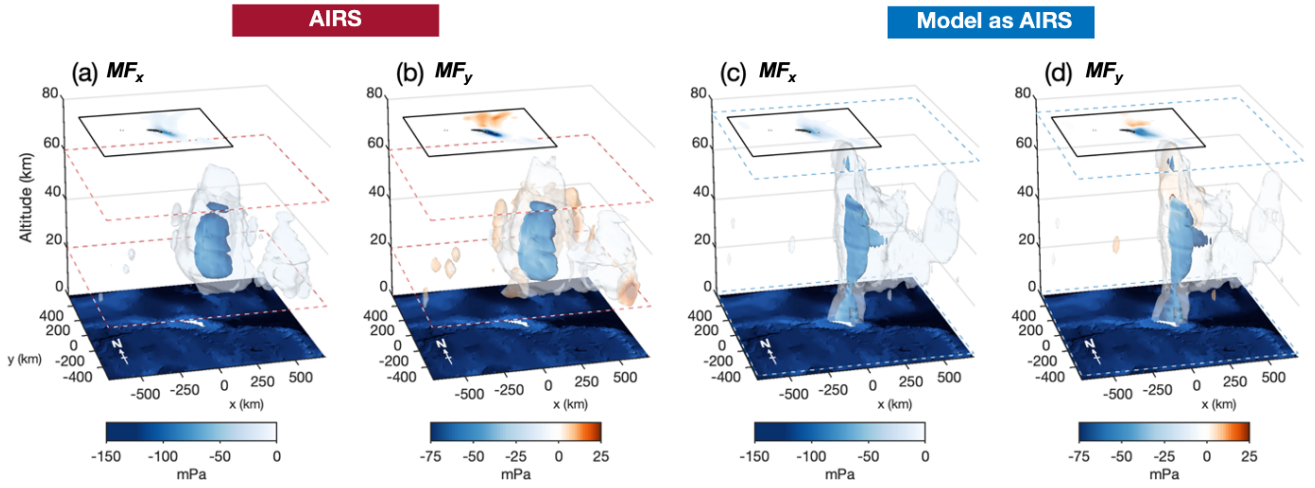

Figure 7. As Fig. 6, but for the zonal and meridional components of gravity wave momentum flux MF_x and MF_y for the AIRS and model-as-AIRS data at 1700 UTC on 5th June 2015. In this example, westward propagation has been assumed in order to break the directional ambiguity in the 3-D measurements.

6.1 Extracting gravity waves temperature perturbations

As a result of the steps in the previous section, the temperature data for each model-as-AIRS timestep lies on the same grid as the nearest AIRS overpass. This means that we can use the same background removal method to extract GW temperature perturbations from both datasets. This is important because it ensures that our analysis method does not introduce differences in the spectral range of GWs visible to each dataset that would invalidate our comparison.

To extract GW temperature perturbations at each altitude level, a horizontal 4th-order polynomial fit is performed in the across-track direction for each cross-track row (e.g. Wu, 2004; Alexander and Barnet, 2007; Hoffmann et al., 2014; Wright et al., 2017; Hi . Slowly-varying background signals due to large-scale temperature gradients or planetary wave activity are contained in this fit. This is then subtracted from each cross-track row to reveal residual GW perturbations.

As a result of the steps above, our AIRS and model-as-AIRS temperature perturbations are sensitive to GWs with vertical wavelengths between $8 \lesssim \lambda_Z \lesssim 40$ km~~in Figs. ??(d-f) also show the same effect. A characteristic "bow wave" pattern centred over the south-eastern tip of the island is observed. This horizontalstructure is typical of orographic "mountain waves" generated by small isolated islands (e.g. Alexander and Grimsdell, 2013). Once again, the model exhibits much more fine horizontal structure than is present in the AIRS measurements. Once the AIRS observational filter is applied in Fig. ??f, this structure is removed and~~ , as defined by the AIRS vertical resolution. In the horizontal, the ~~wave field more closely resembles the AIRS measurements in Fig. ??d.~~ sensitivity cutoff for short horizontal wavelengths is determined by the AIRS footprint spacing (2×13.5 km at nadir and 2×40 km at the scan edges). For longer horizontal wavelengths, sensitivity falls below 90% for $\lambda_H \gtrsim 700$ km and below 10% at $\lambda_H \gtrsim 1400$ km as a result of the 4th order polynomial background fit (Hoffmann et al., 2014).

Sensitivity functions for the 3-D AIRS retrieval to stratospheric GWs can be found in Hindley et al. (2019), Hoffmann et al. (2014) and Ern et al. (2017).

~~Interestingly, it can be seen in both cross-sections that wave amplitudes in the model-as-AIRS data are found to be around~~ Because the AIRS temperature retrieval has reduced vertical resolution and accuracy outside the height range 20 ~~% lower than the corresponding amplitudes observed by AIRS. This could occur if our application of the AIRS observational filter is too harsh, which suppresses these wavelengths more than in the real AIRS measurements, or it could be that the modelled waves at these wavelengths have amplitudes that are too low compared to reality.~~

~~An interesting additional feature is seen at in the AIRS measurements in Figs **??**(a,d). Between altitudes of around 30~~ to 60 km ~~at a location of around $x, y = 500, 300$km relative to the island, a large wave with phase fronts aligned at $+30°$ to the north is observed. By examining this structure in 3-D (not shown) we determined that it is very likely to be a gravity wave. However this wave is not seen in the either the model or~~ altitude (Hoffmann and Alexander, 2009), we set AIRS and model-as-AIRS ~~data in Figs **??**(b-c,e-f) . The absence of this wave in the model , in addition to its size and orientation, suggests that it is not a mountain wave generated by wind flow over South Georgia and may instead be a non-orographic gravity wave propagating into the region from outside, or it may be generated in situ by processes not present in the model. More examples of wave structures like this one can be seen in Fig. **??**, with further discussion in Sect. **??**.~~ GW perturbations outside this range to zero and apply a half-bell tapering window to the upper and lower boundaries. This minimises any impact of edge effects in our subsequent spectral analysis.

~~A clearer impression of the effect of the AIRS observational filter can be gained by visualising the gravity wave structure in 3-D. Figure **??** shows temperature perturbations for AIRS, the model and the~~ Figure 5 shows temperature measurements near 45 km altitude from AIRS, the full-resolution model and the model-as-AIRS ~~at 17:00~~during an AIRS overpass at 0300 UTC on 5th July ~~2015. The AIRS overpass occurred at 16:41 UTC and the model timestep shown is at 17:00 UTC. This is 14 hours later than the example shown in Fig. **??**.~~ 2015. Coloured circles in (a,c) and (d,f) show the locations and horizontal sampling of the AIRS measurements footprints for this overpass. The dashed blue line denotes the horizontal boundary of the model domain.

~~In order to visually inspect the full vertical structure, the temperature perturbations~~ Characteristic bow-wave patterns are visible over South Georgia in ~~Fig. **??** are multiplied by a factor of $\kappa(z) = \exp\left(\frac{z-z_{\mathrm{ref}}}{2H}\right)$, where $z_{\mathrm{ref}} = 40$ km is a reference altitudeand $H = 7$ km is~~ all three datasets in Figs. 5(a-c). These are typical of orographic "mountain waves" from a small isolated island source. These features are apparent as GW perturbations in Figs. 5(d-f). Significant fine horizontal-scale wave structure is also visible in the ~~scale height of the atmosphere, following the approach of (Sato et al., 2012; Wright et al., 2017). This effectively removes the exponential increase (decrease) in wave amplitude above (below) the reference altitude $z_{\mathrm{ref}} = 40$~~full-resolution model, where temperature perturbations exceed ± 12 ~~km~~K directly over the island. The horizontal scales and amplitudes of GW perturbations in the AIRS and model-as-AIRS datasets however show good qualitative similarity, with GW amplitudes around 6-8 K over the island in both datasets. The addition of the AIRS retrieval noise in the model-as-AIRS is also apparent in Figs. 5(c,f). ~~At the reference altitude $z = z_{\mathrm{ref}}$, which is near the centre of the usable vertical range of AIRS measurements, the factor~~

~~$\kappa(z) = 1$ and the isosurfaces plotted represent the true measured temperature perturbation values. This approach allows us to visually inspect the full vertical extent of the gravity wave structure clearly.~~

~~With the scale factor $\kappa(z)$ applied, red and blue isosurfaces are plotted for temperature perturbations at ± 2~~Figures 5(g-i) show a vertical cut through the AIRS, model and model-as-AIRS temperature perturbations along the dashed pink line shown in panels (d-f). Both AIRS and model-as-AIRS measurements are limited to between 20 to 60 ~~K, ± 3~~km altitude, where the retrieval is most reliable (Hoffmann and Alexander, 2009), but for this example we show the full height range of data in the model-as-AIRS for completeness. Westward-sloping GW phase fronts with increasing altitude are found over the island in each of the datasets. These are characteristic of upwardly propagating mountain waves subject to eastward prevailing winds (e.g. Vosper, 2015). Again, the full resolution model in Fig. 5h exhibits large-amplitude wave structure at short horizontal scales (λ_H around 30–40 ~~K and ± 2~~km) over the island and up to around 300 ~~K for the~~ km to the east. Once the AIRS ~~observations, the model and the~~ vertical resolution and horizontal sampling is applied in the model-as-AIRS ~~respectively. Blue and red dashed lines~~ however (Fig. 5i), these short horizontal scale structures are diminished, and the remaining large scale wave structures ($\lambda_H \approx$ 50-150 km) are qualitatively similar to the wave features found in AIRS in Fig. ~~?? show the spatial extent of the model and AIRS measurements respectively. A characteristic~~ 5g. While it is not expected that the phase structure of the mountain wave field ~~is observed~~ in the model and observations should match exactly, the agreement is reasonable. This example indicates that the horizontal and vertical scales of GWs in the model-as-AIRS show good qualitative agreement with GWs observed in AIRS.

To the north east of the island in ~~all three datasets. In the AIRS measurements in~~ Fig. ~~??a, between altitudes of 20 to 40 km, an extended leeward wake~~ 5a, a large horizontal scale GW structure is observed ~~extending to the east beyond the eastern boundary of the model domain. The wave field directly over the island forms a series of near-vertical phase fronts , while the downstream wake pattern becomes increasingly poorly-defined with increasing altitude. Small-scale measurement noise is observed in a speckled pattern at lower altitudes near 20 to 25~~in the AIRS measurements. Close inspection of this example suggests that the phase fronts shown in the AIRS vertical cut in Fig. 5h between 300 km ~~. The model shows a more clearly defined wake pattern relative to AIRS, which extends downstream from the island at all altitudes without any significant loss of definition. Once the AIRS observational filter is applied in Figure ??c, only the large-scale wave structureremains and the observed and modelled wave fields become visually similar. Once again, wave amplitudes are somewhat lower in the model-as-AIRS data than are observed in the AIRS measurements, but the horizontal structure within the AIRS vertical measurement window of 20 to 40~~and 500 km east of the island are part of this same wave structure. We find that wave structures of this kind are commonly observed in AIRS measurements in the region during winter (e.g. Hindley et al., 2019, their Fig. 1), but their origin is unclear (Hendricks et al., 2014). Due to their physical scale and orientation, waves like this example are unlikely to have originated from South Georgia.

No clear evidence of this wave is found in the model or model-as-AIRS, but this is not unexpected. The global forecast that supplies the lateral boundary conditions for our local area model has a coarse vertical grid, with only 70 vertical levels from the surface to near 80 km~~altitude is quite similar.~~ , so GWs such as this one are unlikely to be accurately simulated. Further, even if they are accurately simulated, it is not clear how realistically these GWs would be transferred through the model boundary

conditions into the local area model. As a result, we expect our model and model-as-AIRS temperature fields to under-represent GWs of this kind. This is discussed further in Sect. 9.

It is interesting to note that the additional wave feature seen north east of the South Georgia in Figure **??**(a ,d) is now, 14 hours later in Fig. **??**, no longer observed. This further suggests that this wave structure may be a transitory wave packet from a source outside the model domain.

6.2 Measuring gravity wave properties with a 3-D *S*-transform

3-D S-transform (3DST) analysis of temperature perturbations from AIRS satellite observations (top row) and the model-as-AIRS over South Georgia (bottom row) for 1700 UTC on 5th July 2015. Coloured isosurfaces in panels (a,e) show the AIRS and model-as-AIRS temperature perturbations T', while panels (b,f), (c,g) and (d,h) show 3DST-measured absolute wave amplitude $|T'|_{3DST}$, horizontal wavelength λ_H and vertical wavelength λ_Z respectively. Blue dashed and red dashed lines denote the upper and lower boundaries of the model domain and AIRS measurements. Horizontal cross-sections through the data at 40 km altitude are shown in the top left hand corners of each panel, sharing a colour scale with the isosurfaces. As in Fig. **??**, temperature perturbations in panels (a) and (e) are scaled by the factor $\kappa(z)$ in order to see the vertical wave structure clearly.

As Fig. 6, but for the zonal and meridional components of gravity wave momentum flux MF_x and MF_y for the AIRS and model-as-AIRS data at 1700 UTC on 5th June 2015.

6.3 Measuring 3-D gravity wave parameters with a 3-D *S*-transform

We identified 87 co-located 3-D AIRS overpasses that coincided with the time periods our local-area modelling simulations over South Georgia during winter July 2013 and June-July 2015. By applying the AIRS observational filter to In Sect. 4 we used directional wind perturbations u', v' and w' to estimate GW momentum flux in the full-resolution model via Eqn. 1. However, AIRS can only measure GW temperature perturbations, so we must use these to make our comparison between AIRS and the model output, we are able to make a fair comparison between the observed and simulated gravity wave fields in 3-Dmodel-as-AIRS.

In order to investigate the properties of the gravity waves in our AIRS, model and We can use spatially-localised measurements of GW temperature amplitudes T' horizontal wavenumbers k and l and vertical wavenumber m to estimate directional GWMF in AIRS and model-as-AIRS datasets, spectral analysis is needed. In particular, we need to be able to measure wave amplitudes, wavelengths measurements via the relation

$$(\mathrm{MF}_x, \mathrm{MF}_y) = \frac{\rho}{2}\left(\frac{g}{N}\right)^2\left(\frac{|T'|}{\bar{T}}\right)^2\left(\frac{k}{m}, \frac{l}{m}\right) \tag{2}$$

where MF_x and directions but also localise these quantities spatially. MF_y are the zonal and meridional components of GWMF, ρ is atmospheric density, g is the acceleration due to gravity, N is the buoyancy frequency and $\bar{T}$ is the background atmospheric temperature (Ern et al., 2004). Zonal, meridional and vertical wavenumbers k, l and m are related to spatial wavelengths as

$k = 2\pi/\lambda_x$, $l = 2\pi/\lambda_y$ and $m = 2\pi/\lambda_z$ respectively. Ern et al. (2017) showed that this relation, based upon the mid-frequency approximation (Fritts and Alexander, 2003), is valid for GWs within the spectral range visible to AIRS.

To ~~do this~~obtain spatially-localised measurements of GW amplitudes and wavelengths, we use a 3-D adaptation of the S-transform (also known as the Stockwell transform). Developed by Stockwell et al. (1996), the S-transform is a widely-used spectral analysis technique that can localise and measure the amplitudes of individual frequencies (or wavenumbers) in a time-series or distance profile. The S-transform has been applied for ~~gravity-wave~~ GW analysis in a variety of geophysical datasets ~~(e.g. Fritts et al., 1998; Stockwell and Lowe, 2001; Alexander and Barnet, 2007; Alexander et al., 2008; Stockwell et al., 2011; McDonald~~ (e.g. Fritts et al., 1998; Stockwell and Lowe, 2001; Alexander and Barnet, 2007; Alexander et al., 2008; Stockwell et al., 2011; McDonald and has also been applied in a variety of other fields, such as the planetary (Wright, 2012), engineering (Kuyuk, 2015) and biomedical sciences (e.g. Goodyear et al., 2004; Brown et al., 2010; Yan et al., 2015).

Here we ~~apply the 3-D S-transform (3DST) to our AIRS, model and model-as-AIRS temperature perturbations. We~~ use the N-dimensional S-transform (NDST) software package ~~developed~~ as described by Hindley et al. (2019). This version builds on the work of previous multi-dimensional S-transform analysis by Hindley et al. (2016) and Wright et al. (2017), but applies a superior wave amplitude measurement technique and features a much faster computational methodology which reduces computation time by around a factor of 10 over previous 3DST versions for AIRS analysis. A step-by-step guide describing how the 3DST method is applied to 3-D ~~Airs~~ AIRS measurements is described in Hindley et al. (2019, their Sect. 3). ~~The same process is followed here for all three datasets~~Validation of the 3DST analysis method using synthetic wave fields can be found in Hindley et al. (2016) and Hindley et al. (2019).

~~The~~ To make meaningful 3DST ~~is applied to the regridded temperature perturbations of the~~ measurements of wavelengths, a regular orthogonal grid is required. The AIRS and model-as-AIRS datasets have irregular across-track spacing (Fig. 5), so we interpolate the GW temperature perturbations for each AIRS overpass and each model-as-AIRS timestep onto a 10×10 km horizontal grid centred on South Georgia. This is finer than the horizontal sampling of the AIRS grid, so aliasing effects are unlikely to be significant. If any aliasing effects do occur, their effects will be equal for the AIRS and the model-as-AIRS, so this will not affect our comparison. In the vertical, we interpolate on to a 1.5 km vertical grid which is finer than the stratospheric vertical grids (and vertical resolutions) of both the AIRS retrieval and the model. This regridding is therefore unlikely to affect our results.

We apply the 3DST to regularly gridded GW temperature perturbations for 87 3-D AIRS measurements and ~~every hourly time step for the model and~~ 1320 hourly model-as-AIRS ~~output~~ timesteps during July 2013 and June-July 2015. ~~The results are spatially localised amplitudes, wavelengths and directions for all three datasets. Because all three datasets are on the same grid, the same frequencies can be analysed for all three datasets.~~ Following the approach of Hindley et al. (2019), we set the 3DST scaling parameter $c_x = c_y = c_z = 0.25$ and analyse for the 1000 largest-amplitude wave signals with wavelengths greater than ~~30~~27 km, ~~30~~27 km and ~~3~~6 km in the x, y and z directions respectively. These are Nyquist sampling limits of twice the smallest separation of original AIRS sampling pattern (2×13.5 km) in the horizontal, and twice the spacing of original vertical grid of the AIRS retrieval (2×3 km). Because both datasets are analysed on the same regular grid, the exact same frequencies are

be analysed for both. These steps provide spatially-localised measurements of GW temperature amplitudes, wavelengths and directions for the AIRS and model-as-AIRS datasets.

6.3 Case study comparison of 3-D gravity wave properties in AIRS and the model-as-AIRS

We inspect 3DST measurements of GW properties in AIRS and model-as-AIRS for an AIRS overpass at 1700 UTC on the 5th July 2015 in Figs. 6 and 7. This overpass occurs 14 hours after the example shown in Fig. 5, and is one of the most intense examples of mountain wave activity observed during the time periods of the model runs. The purpose of this case study comparison is not only to compare the model-as-AIRS to the AIRS observations, but also to confirm that we can measure the 3-D properties of the dominant wave structure with the 3DST.

Figure 6 shows the ~~results of our~~ 3DST analysis ~~for an example of~~ results for AIRS measurements (top row) and model-as-AIRS (bottom row) at 1700 UTC on 5th July 2015. ~~These are the examples shown in Fig. ??(a,e). Input and absolute 3DST-measured~~ Input temperature perturbations and measured wave amplitudes are shown in ~~Figs. 6~~panels (a,e) and ~~Figs. 6~~(b,f) respectively. ~~As in Fig. ??, the factor $\kappa(z)$ has been applied the the temperature perturbations in panels (a) and (e) in order to show the vertical structure clearly. Panels (b) and (f) show the true measured values.~~ Horizontal and vertical wavelengths λ_H and λ_Z are shown in ~~Figs. 6~~panels (c,g) and ~~Figs. 6~~(d,h) respectively. In each panel, a horizontal cross-section through the data at an altitude of 40 km is overlaid in the top left hand corner, ~~and the~~ which shares a colour scale with the isosurfaces. The extent of the AIRS and model data are shown by red and blue dashed line respectively. In this figure, a 3×3 element horizontal boxcar filter has been applied to make the isosurfaces smoother for visual clarity.

~~A clear bow-wave pattern is observed in~~ In both the AIRS measurements and the model-as-AIRS in Figs. 6(a-b) and 6(e-f)~~.~~ ~~The 3-D structure of the~~ , temperature perturbations exhibit a bow-wave pattern, which is characteristic of a mountain wave field ~~can be clearly seen in both datasets, and there is reasonable apparent similarity between the two~~over a small isolated island such as South Georgia (e.g. Vosper, 2015). The largest wave amplitudes are localised over the island in both datasets~~and are observed to increase with altitude, exceeding~~ , where values exceed 5 K at ~~an altitude of~~ 40 km ~~directly over~~ altitude directly over and immediately downwind of the island. The leeward "wings" of the mountain wave field that extend to the north and south are more prominent in AIRS measurements than in the ~~island in both AIRS and the~~ model-as-AIRS~~. In general however,~~ , but measured wave amplitudes closer to the island are comparable. As in Fig. 5, real and specified retrieval noise is apparent in the AIRS and model-as-AIRS ~~exhibits slightly lower wave amplitudes further away from the island , and the large leeward wave structures east of the island are have significantly lower amplitudes than their counterparts in the AIRS measurements. This effect, which is seen throughout this study, suggests that the model has a tendency to generate tightly localised waves in a focused column directly over the island, while observations show that broader bow-wave shaped wake regions are more enhanced that in the simulated model-as-AIRS, particularly at lower altitudes~~temperature perturbations respectively, as we intended.

~~Measured horizontal wavelength $\lambda_h = \left(\lambda_x^{-2} + \lambda_y^{-2}\right)^{-1/2}$ and vertical wavelength λ_z for this example are shown in Figs.~~ Figures 6(c,g) ~~and 6(d,h)~~ show measured horizontal wavelengths $\lambda_h = \left(\lambda_x^{-2} + \lambda_y^{-2}\right)^{-1/2}$ for the AIRS and model-as-AIRS respectively. In both datasets, ~~the mountain wave pattern is revealed as a tight column of~~ short horizontal wavelengths ~~with values~~

~~100~~$\lambda_H <$ 50–100 km ~~and shorter that is located~~ are located in a vertical column directly over the island. ~~Longer horizontal wavelengths are observed with increasing horizontal distance from the island. Again, the short horizontal wavelengths measured~~ The bow-wave patterns to the north and south exhibit longer measured horizontal wavelengths of around 200 km in AIRS, but shorter wavelengths at around 150 km in the model-as-AIRS~~are more tightly localised over the island. In regions with very little wave activity, very~~ .

Away from the island, long horizontal wavelengths are measured~~, although the colour map of Figs. 6(c,g) saturates at 350 km. This is to be expected, as a relatively featureless horizontal region will be measured in the~~ . This is due to a design choice in our 3DST analysis~~as having infinitely long horizontal wavelength. In the AIRS measurements westward of this island, small isolated regions of short horizontal wavelengths are observed. These correspond to~~ . For regions with no clear wave activity, only retrieval noise is present. The wavelength limits and scaling parameter settings in our 3DST analysis are designed so that the dominant measured horizontal wavelength in these regions is long ($\lambda_H \gtrsim$ 600–1200 km), analagous to a horizontal "flat field", following the approach of Hindley et al. (2019). In practice, we find that this choice is advantageous, because measurements of incoherent small-scale ~~uncorrelated noise features in AIRS measurements. Since their measured amplitude is low as seen in Fig. 6b adjacent, we do not expect these remaining features to significantly affect our results.~~ retrieval noise could otherwise be confused with measurements of short horizontal wavelength GWs (Alexander et al., 2009; Hindley et al., 2016, 2019) . Other studies, such as Ern et al. (2017), choose to measure these regions as having short horizontal wavelengths using the S3D method of Lehmann et al. (2012).

6.3.1 ~~Regions of downward-propagating wave structure~~

~~Figures 6(d,h) show an increase in absolute vertical wavelengths with altitude from about 10 km at altitudes near 20 km to around 30 to 40 km above altitudes near 45 km. This is consistent with the refraction of a stationary mountain wave field to longer vertical wavelengths with increasing zonal wind speeds with height. The longest vertical wavelengths are between 30 to 40 km and are observed directly over the island, with vertical wavelength decreasing slightly with increasing horizontal distance from the island. This is consistent with the phase fronts seen in Fig. **??**, 14 hours earlier.~~

~~Since the AIRS observations are, on this scale, a pseudo-instantaneous snapshot in time, an ambiguity arises between whether an observed wave is travelling "upwards and forwards"or "downwards and backwards". This is a common ambiguity in gravity wave observations. Various approaches have been used to constrain it in recent studies, such as the use of supplementary wind fields or the assumption of upward propagation (e.g. Alexander et al., 2009; Alexander and Grimsdell, 2013; Wright et al., 2016a; Hindley . To break the ambiguity for the example in Fig. 6, we assume that these measured waves are westward propagating mountain waves, whose horizontal phase speeds and wave vector directions are equal and opposite to the wind speed and direction. This results in wave field that appears stationary with respect to the ground. Simulated zonal winds in Fig. **??** during this period are almost entirely eastward, so the assumption of a mountain wave field with a westward orientated wavevector is reasonable.~~

~~When this assumption is applied, we are able to constrain the sign, and therefore the direction, of the vertical wavenumber m from the orientation of the measured wave. In Figs. 6(d) and 6h, positive (negative) vertical wavelength λ_z values imply~~

~~downward (upward) wave propagation. This sign convention follows the derivation of Fritts and Alexander (2003), where $m < 0$ for an upwardly propagating wave.~~

~~Interestingly, several large pink regions~~ Measured vertical wavelengths for AIRS and the model-as-AIRS are shown in Figs. 6(d,~~f)indicate possible regions of downward wave propagation in AIRS and the model-as-AIRS. This seems counter-intuitive given the nature of the probable gravity wave source, and so to investigate this further, we animated a vertical cross-section through the model-as-AIRS wave field as shown in Fig. **??**c for each hourly model time step between 0000 UTC on 5th July to 1200 UTC on 6th July 2015. We found that eastward sloping phase fronts leeward of the island persisted strongly from 0200 UTC on the 5th to 2300 UTC on the 6th, dispersing thereafter. Indeed, the feature can be seen~~ h). Vertical wavelengths are found to increase with altitude in both datasets. This is consistent with the expected refraction of mountain waves that are subject to increasing background wind speed with altitude, as indicated by the model winds in Fig. ~~**??**c and was particularly strong at 1700~~2(a,b). It is also consistent with the reduced vertical resolution of the AIRS retrieval with increasing height above around 40 ~~UTC on the 5th July,when the AIRS overpass occurred. If these wave structures are quasi-stationary with respect to the ground, which would be true for a mountain wave pattern, then this implies westward wave propagation into the prevailing wind. This would in turn imply that these regions do indeed contain downward propagating waves.~~

km altitude (Hoffmann and Alexander, 2009, their Fig. 5). In the AIRS measurements~~in Fig. 6d, large negative vertical wavelengths around -30 to -40~~, longer vertical wavelengths $\lambda_Z \gtrsim 35$–40 km ~~(implying upward wave propagation) are observed over the island and throughout most of the leeward wake pattern up to altitudes around~~ are found directly over and immediately to the east of the island near 40 km ~~. Above this point however, large positive vertical wavelengths around 30 to 40~~altitude. In the model-as-AIRS, vertical wavelengths are slightly shorter, with $\lambda_Z \gtrsim 25$–35 ~~km are also observed. This abrupt change in the sign of the vertical wavenumber, without any apparent change in the horizontal wavenumber in this area, implies that waves directly over the island have phase fronts aligned near-vertically, such that only a small change in inclination is required to change the sign of the vertical wavenumber. The example in Fig. **??** helps us to interpret this. In the AIRS measurements in Fig. **??**a, phase fronts above altitudes of~~ km near 40 ~~km directly over the island are aligned near vertically. For this reason, we suspect that this change in the sign of the vertical wavenumber is, in this example in AIRS , more likely to be due to measurement error as the phase fronts become near-vertical. However, given the apparent downward propagating wave structure in the~~ km altitude. This could help to explain why the measured AIRS GW temperature amplitudes in Figs. 6b exhibit slightly larger values than in the model-as-AIRS~~, we cannot rule out the possibility of a downward propagating structure in the AIRS measurements too, given the striking similarity between the two. Indeed, between around 150 and 300 km east of the island in Fig. **??**a, a very small region containing apparently eastward sloping phase fronts is observed.~~

~~Hence it is possible that we are indeed observing downwardly propagating waves in the lee of South Georgia. This result would be consistent with the gravity wave analysis of radiosonde observations in Moffat-Griffin et al. (2017), who used data from the same radiosonde campaign as shown in Sect. **??**. They found that, downwind of~~ . If the real GW structure exhibited a slightly longer vertical wavelength compared to the ~~island, 66% of observed gravity wave activity was downward propagating. However, gravity waves in radiosonde measurementscorrespond to much shorter vertical and horizontal wavelengths than considered here in AIRS measurements.~~

~~One possibility is that these downward regions in~~ simulated GW, this would increase the sensitivity of AIRS to this wave, resulting in larger measured temperature amplitudes. This could arise due to slightly stronger wind speeds than simulated in the model. Unfortunately, the ~~model-as-AIRS results are evidence of secondary gravity waves generated by mountain wave breaking or other wave-wave processes. Previous work (e.g. Vadas and Fritts, 2002; Bossert et al., 2017; Becker and Vadas, 2018; Vadas an has suggested that such secondary waves may be generated by large orographic sources, acting as a mechanism to "convert" geographically stationary mountain waves into waves with overall-conserved but non-zero propagation characteristics, including the possibility of generating downward-propagating waves. While such work to date has focused on major gravity wave hot spots such as the southern Andes, South Georgia may also be a sufficiently intense mountain wave source to exhibit such effects. It could also be the case that such downward-propagating regions of the field are generally a normal feature of the gravity wave field over small-island sources, even in the absence of secondary wave generation, but their effects are usually dominated in high-resolution models by the much more intense fine model structure (see Fig. **??**b). Given that a model damping layer above altitudes of 58.5 km was applied in the simulations, reflections from critical layers near the model top are unlikely. More investigation of wave features such as these in future studies is needed.~~ radiosondes launched from South Georgia on the afternoon of the 5th July 2015 did not reach their intended altitudes (Fig. 3a) due to extreme weather conditions reported at the launch site, so we cannot investigate this further for this example.

6.3.1 Zonal and meridional momentum fluxes

~~A key quantity in gravity wave research is the vertical flux of horizontal pseudo-momentum, generally referred to as momentum flux. This property helps to quantify the transfer of momentum carried by gravity waves, and the constrains the drag or driving effect on the mean flow that will arise when the wave eventually breaks or is absorbed. Quantifying the momentum flux budget of mountain wave sources from isolated small islands is a key current area of research (McLandress et al., 2012; Alexander and Grimsdell, 2~~

~~The zonal and meridional components of gravity-wave momentum flux MF_x and MF_y can be estimated from our 3DST measurements of wave amplitude, horizontal wavelength and vertical wavelength via the relation in Ern et al. (2004):~~

$$(F_x, F_y) = \frac{\rho}{2}\left(\frac{g}{N}\right)^2\left(\frac{|T'|}{\bar{T}}\right)^2\left(\frac{\lambda_z}{\lambda_x}, \frac{\lambda_z}{\lambda_y}\right)$$

~~where ρ is atmospheric density, g is the acceleration due to gravity, N is the buoyancy frequency, $|T'|$ is absolute wave amplitude, $\bar{T}$ is the background temperature, and λ_x, λ_y and λ_z are zonal, meridional and vertical wavelengths respectively. Wright et al. (2016a) and Ern et al. (2017) showed that this relation, based upon the mid-frequency approximation, is valid for the waves to which the 3-D AIRS retrieval is sensitive.~~

~~Fig. 7 shows zonal and meridional momentum fluxes F_x and F_y calculated via Eqn. 2 for the AIRS and~~ Figure 7 shows zonal and meridional momentum fluxes MF_x and MF_y calculated using Eqn. 2 for measured GW properties in AIRS and the model-as-AIRS ~~measured gravity wave quantities shown in Fig. 6. For each panel, isosurfaces are drawn where the absolute momentum flux $|F_{x,y}| = \sqrt{F_x^2 + F_y^2} =$ 10 and 100~~at 1700 ~~mPa. These isosurfaces are then coloured using the values of zonal~~

and meridional flux at each location. As in Fig. 6 horizontal cross section at an altitude of UTC on 5th July 2015. As in Fig. 6, horizontal cross sections through the data at an altitude of 40 km is overlaid in the top left hand corner of each panelkm are overlaid in the top left hand corner of each panel.

To show directional GWMF, we must also break a directional ambiguity in our 3-D measurements. Because each AIRS overpass only provides observations for a single moment in time, we cannot distinguish between GWs that propagate "upwards and forwards" or "downwards and backwards" (Wright et al., 2016a). For the example in Fig. 6, we inspected the time-varying wave structure in the model-as-AIRS temperature fields to determine that the simulated wave is a quasi-stationary westward-propagating mountain wave subject to eastward wind conditions. This means we can confidently break the directional ambiguity for this example and assume westward propagation, since the agreement in the mountain wave structure between the AIRS and the model-as-AIRS is good. But this is not possible for all AIRS and model-as-AIRS measurements in our study, because not all measured waves are expected to be clear mountain waves. In the general case therefore, we assume upward propagation ($m < 0$) for observed waves in all subsequent results. This follows the approach of several previous studies involving AIRS measurements (Ern et al., 2017; Wright et al., 2017; Hindley et al., 2019, 2020). Ern et al. (2017) and Hindley et al. (2020) found that a realistic horizontal directionality of global stratospheric GWMF can be obtained by making this upward assumption.

Highest momentum fluxes

The largest GWMF values in Fig. 7 are observed in a vertical column directly over the island in both the AIRS and model-as-AIRS wave fields. These regions coincide with the largest wave amplitudes, shortest horizontal wavelengths and longest vertical wavelengths in Fig. 6, so this is to be expected. Zonal momentum fluxes are almost entirely directed westward(negative)directed westward, with values that are significantly higher than the meridional components, exceeding -150between 50–150 mPa in both the AIRS and model-as-AIRS data directly over the island. In the meridional direction, regions of northward (positive) and southward (negative) flux are observed Meridional GWMF is predominantly directed southward over the island, with values between 50–75 mPa in both datasets, indicating a south-westward direction of the net GWMF. A northward component of MF_u is also found to the north and south of the islandrespectively in both datasets in a characteristic bow-wave pattern. This is an encouraging result that suggests that our 3DST technique analysis is correctly localising the opposing diverging meridional components of the characteristic bow-wave pattern . Southward fluxes are around 3 times larger than their northward counterparts, exceeding 75 mPa in both the AIRS and model-as-AIRS data.

The AIRS measurements for this example in Fig. 7 exhibit higher momentum fluxes than the coincident model-as-AIRS time step at 1700 UTC on 5th July 2015. Inspection of Fig. 6 suggests this is primarily due to the generally increased wave amplitudes measured by AIRS as discussed above. Once again, it is apparent that the AIRS measurements exhibit a much more enhanced leeward bow-wave pattern than is observed in the model-as-AIRS, whose flux is more concentrated into a tight column directly over the to the north and south of the island.

Time series of median gravity-wave amplitudes (coloured lines) from AIRS observations (a,b), the full-resolution model (c,d) and the model-as-AIRS (e,f) over a horizontal region of 250 km radius centred 100 km east of the island and a vertical region between altitudes of 25 and 45 km for July 2013 (left) and June-July 2015 (right). Shaded regions in panels (a-f) show the 25th and 75th, 15th and 85th, and 5th and 95th percentiles of measured wave amplitudes over the same region. Red circles

in (a,b) show the overpass times of the AIRS measurements. Panels (g) and (h) show the magnitude of the wind speed from the model for the 2013 and 2015 modelling campaigns respectively.

Time series of zonal gravity-wave momentum flux MF_x for the AIRS observations (a,b), the full-resolution model (c,d) and the model-as-AIRS (e,f) over South Georgia during June-July 2013 and 2015 for the same region used in Fig. **??**. Here, coloured lines show the average net zonal momentum flux, while the grey shaded areas show the average eastward and westward fluxes over the region. Positive (negative) values indicate an eastward (westward) direction. Panels (g) and (h) show horizontally-averaged zonal wind speed from the model for the 2013 and 2015 modelling campaigns respectively.

7 Results

6.1 Time series of wave amplitudes and directional momentum fluxes in the AIRS, model and model-as-AIRS

7 Gravity wave properties in AIRS and the model-as-AIRS over South Georgia

The examples shown in Sect. **??** demonstrate the effect of applying the AIRS observational filter 6 demonstrate that the AIRS sampling and resolution can be applied to the model output and to make a comparable model-as-AIRS dataset. We then showed that wave amplitudes, wavelengths and directional momentum fluxes can be measured in each dataset using the using a 3DST method . In this section we investigate how these properties vary temporally during the two modelling campaigns in a case study example. Here, we apply this method to all available AIRS observations and model-as-AIRS timesteps during the model runs in July 2013 and June-July 2015.

Figures **??**, **??** and **??** show time series of wave amplitude, zonal momentum flux MF_x

7.1 Timeseries of wave amplitude and directional GWMF

Figure 8 shows measured wave amplitudes and zonal and meridional momentum flux MF_y respectively. For each AIRS overpass and time step of the model and fluxes against time for AIRS and model-as-AIRS , the median wave amplitude and mean directional momentum fluxes are found within measurements. Values are averaged over a horizontal region of radius $r = 250$600×400 km centred 100 km east of the island between altitudes of on the island (region C in Fig. 4) between 25 km and 45 km . This horizontal region was chosen so as to capture the most intense region of the gravity wave fields just to the east of the island (this is best illustrated in Figs. **??** and **??**, which are discussed later). The height region was chosen in order to take advantage of the best AIRS vertical resolution at lower altitudes and to avoid any effects from the model damping layer at high altitudes. Red markers in panels (a) and (b) of each figure indicate the overpass times of the AIRS instrument. Model wind speeds, averaged horizontally over the model domain, are shown in panels (g) and (h) of each figure.

In Fig. **??**, coloured lines show the median wave amplitude in the region. The area within the shaded regions show the 5th altitude. The grey shaded areas in Fig. 8 show the extent of the 10th and 95th, 15th and 85th and 25th and 75th percentiles of all measurements within the region. The 5th and 95th percentile shading is the lightest, while the 25th and 75th percentile shading is the darkest. This approach is useful for two reasons. First, gravity wave amplitudes are likely to exhibit a log-normal

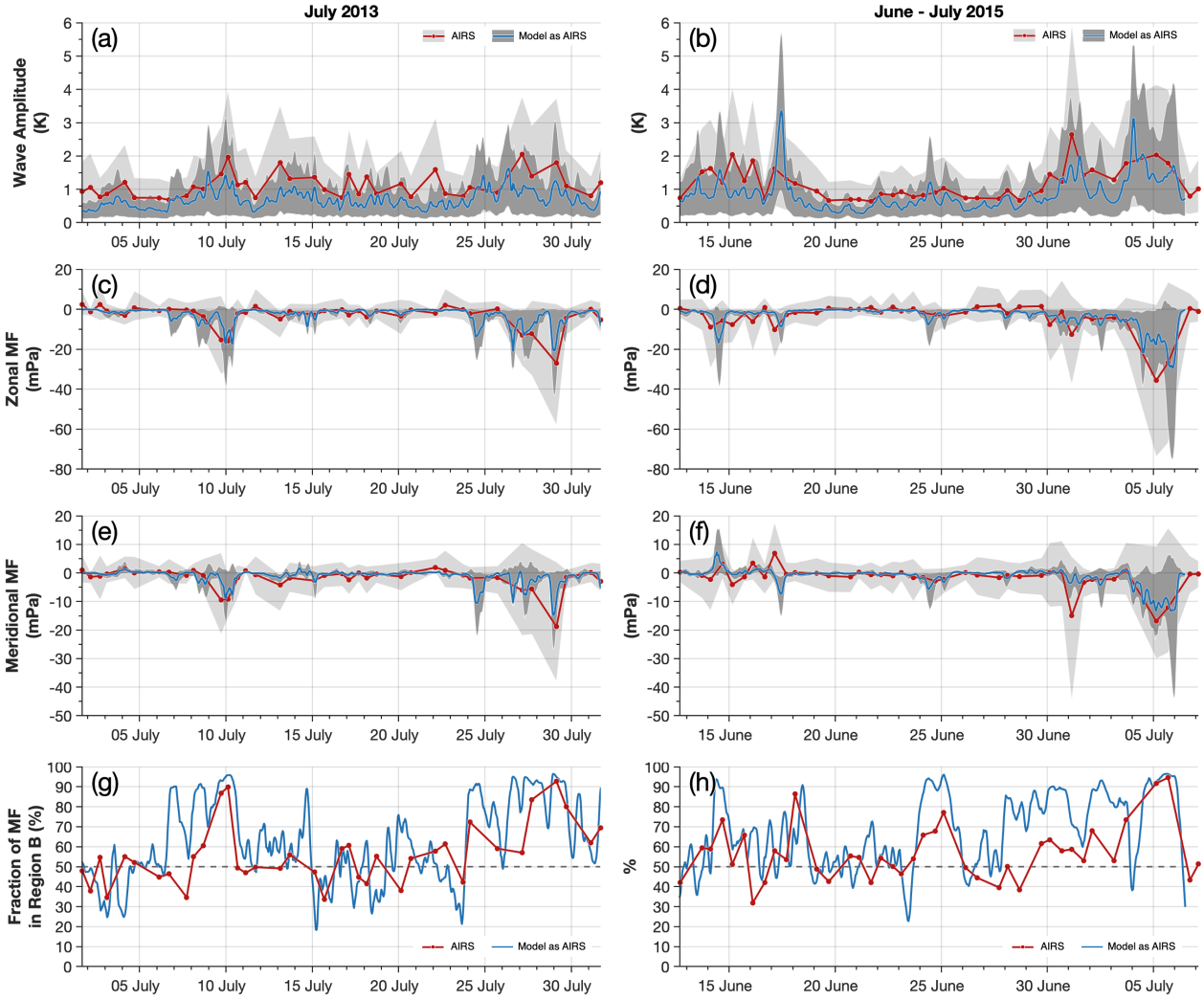

Figure 8. ~~As Fig. ??, but for the meridional component~~ Time series of ~~gravity-wave~~ median GW amplitudes and net zonal and meridional momentum ~~flux MF_y in panels~~ fluxes derived from AIRS measurements (~~a~~red) ~~to~~ and the model-as-AIRS (~~f~~blue) for July 2013 and ~~meridional wind in panels~~ June-July 2015. Values are averaged between altitudes of 25 and 45 km over a horizontal region 600×400 km centred on the island (~~g~~region C in Fig. 4)~~an (h)~~. ~~Coloured lines~~ Red circles show the ~~average net meridional flux, while~~ overpass times of the AIRS measurements. Light and dark grey shaded areas show the ~~average northward~~ 5th and ~~southward~~ 95th percentiles of measured wave amplitudes and momentum fluxes over the same region for AIRS and the model-as-AIRS respectively. ~~Here~~As in Fig. 2, ~~positive~~ panels (~~negative~~g) ~~values indicate an northward~~ and (~~southward~~h) ~~direction~~show the percentage of the total GWMF measured downwind of South Georgia (region B in Fig. 4). Percentage values larger 50% are a good indication of mountain wave activity.

distribution, so the median is more appropriate than the mean. Second, this approach can give us information about the spatial distribution of wave amplitudes within the region. 90th percentiles of measured wave amplitude and GWMF over this region for AIRS (light grey) and the model-as-AIRS (dark grey) respectively.

Figure ?? reveals that gravity wave activity over the island The timeseries in Fig. 8 indicate that GW activity over South Georgia is highly intermittent during our period of study. Several time periods of enhanced increased gravity activity are seen in all three datasetsobserved in both the AIRS and model-as-AIRS, such as the periods 7th - 11th July and 24th - 31th July during 7th–11th July and 24th–31st July 2013, and the periods 14th - 16th June, 24th - 26th during 14th–16th June, 24th–26th June and 29th June - 6th July during June–6th July 2015. These periods of increased stratospheric wave activity generally coincide with high wind speeds that extend down into the troposphere. This gives rise to avertical conduit of high wind speeds that allows mountain waves from South Georgia to propagate vertically without encountering critical levels.

During these periods, median gravity wave amplitudes between 1 and 2Figures 8(a,b) indicate that during these events, area-averaged GW amplitudes increase to around 1–2 K are observed in AIRS measurements, with higher values of around 2 to 6 K observed in the model. Somewhat lower median values typically less than around 1 Kare observed in the model-as-AIRS. The shaded percentile regions in Fig. ?? yield information within this distribution and reveal high wave amplitudes exceeding 16however reveal that some locations in the region can exhibit much larger amplitudes during these periods, where the 90th percentile of measured amplitudes can exceed 5 K at the 95th percentile in the model on 17th June and K. This is consistent with the large wave amplitudes measured over the island in the examples in Figs. 5 and 6 for the overpasses on 5th July 2015. The latter example is the one shown in Fig. 6. The fact that the median values are quite low for the model suggests that these high amplitudes are confined to a small region immediately downwind of the island, as we will show later in Fig. ??.

Wave amplitudes do not reach such high values in the AIRS and model-as-AIRS, whose 95th percentile values typically range between 5 and 6 K. Interestingly, the AIRS measurements exhibit a somewhat narrower distribution of wave amplitudes during periods of high wave activity, since the percentile boundaries are quite close together. This suggests that wave amplitudes are more broadly distributed over the horizontal region rather than tightly localised, which could be indicative of large-scale non-orographic wave activity or simply longer horizontal wavelengths with large amplitudes in AIRS measurements. The impact of measurement noise in AIRS is also apparent in the measured wave amplitudes , with median values almost never falling below 0.5 K, even during periods of low wave activity. This suggests a usable noise threshold of around 0.5 K for AIRS amplitudes. It is very difficult to determine whether this noise is due to instrument measurement noise, 3-D retrieval noise, artefacts arising from unresolved waves or true anisotropy in the real atmosphere, but it is likely that all of these sources of error play a part. In the model and model-as-AIRS, median amplitudes fall to near zero due to the lack of this measurement noise.

The median model-as-AIRS amplitudes over this region rarely exceed those in the AIRS measurements, however there are some notable exceptions. One such example is on 17th June 2015, where a brief spike of increased wave amplitude is observed with median wave amplitudes nearly reaching 3 K in the model-as-AIRS and the 95th percentile exceeding 5 K. Inspection of the model wave field during this period revealed a series of large and likely non-orographic wave fronts that formed over most of the model domain. These phase fronts did not appear to move with time, but also did not appear to be orographic (that is, no

clear bow-wave structure), suggesting that they may have originated in the global forecast that supplied the model boundary conditions.

Time series of the The timeseries of net zonal and meridional components of gravity-wave momentum flux momentum fluxes in Figs. ?? and ?? show a similar pattern. Here, coloured lines indicate the average (net) zonal and meridional flux over the same horizontal region as used in Fig. ?? between altitudes of 25 and 458(c-f) also reveal high intermittency. During periods of increased GW activity, area-average GWMF values are found to increase to around 20–40 km. Shaded regions indicate the average positive and negative flux values within the region, with positive (negative) values indicating eastward and northward (westward and southward) directions in Figs. ?? and ?? respectively. These shaded values are found by taking the average of all the positive (negative) fluxes with all the negative (positive) fluxes set to zero. This information is important since, for a typical mountain wave field such as might be observed over South Georgia, significant northward and southward flux could be generated but the net meridional flux could be very low since the two directions may cancel out, suggesting incorrectly that there is little gravity wave activity. Zonal and meridional wind speeds against altitude from the model, averaged horizontally over the model domain, are shown in panels (g,h) in both figures.

Net zonal fluxes in Fig. ?? are largely westward in all three data sets. As discussed above, an assumption of upward wave propagation is used to break the directional ambiguity in the data. The fact that the zonal flux is overwhelmingly westward, which is what we would expect given the eastward wind conditions, after this assumption has been made is a good indication that the assumption is reasonable for the majority of wave measurements here.

A similar pattern of gravity wave activity with time is observed compared to Fig. ??, with wave activity occurring in bursts that last several days at a time. In terms of the timing of these bursts of momentum flux, the model and model-as-AIRS reproduce the AIRS observations remarkably well. Average mPa in the zonal direction and 10–20 mPa in the meridional. As with the wave amplitudes, the 10th and 90th percentile shading regions indicate that peak GWMF values in the model over the region peak around 200 to 300region reached much higher values, exceeding 70 mPa during these wave bursts, falling to near zero during periods of low wave activity. Zonal flux valuesare more comparable in AIRS and the model-as-AIRS, peaking at around 10 to 20 mPa and 5 to 15in the zonal direction and 40 mPa respectively. As with the wave amplitudes, average zonal fluxes in the model-as-AIRS can be around 25% lower than the AIRS fluxes during the same periods. This suggests more wave activity in the model may be occurring outside of the AIRS observational window, most likely at very short horizontal scales below around 50 km.

Although the direction of the zonal fluxes is overwhelmingly westward, particularly in the model, small periods of eastward flux are observed, for example during 9th-10th in the meridional during the largest wave events in July 2013 in the model-as-AIRS and 4th-7th July 2015 in 2015.

The directionality of net zonal and meridional GWMF in Figs. 8(c-f) is generally negative for both AIRS and the model-as-AIRS. Although much of this eastward flux could be real, if we recall that , indicating a predominantly south-westward net direction. This is consistent with the results for the case study in Fig. 6(d, h) we identified regions where the gravity wave phase fronts were aligned so close to the vertical that they were measured as being eastward sloping with height, which could suggest downward wave propagation . If we always assume upward wave propagation, as we have done here, then the direction of flux

from downward propagating waves would be reversed. Here, such downward flux would be incorrectly measured as eastward, so we may be underestimating the westward zonal flux in Fig. **??** during these time periods.

Meridional fluxes in Fig. **??** are largely southward, particularly in the model and model-as-AIRS data. As before, 7, but we should recall here that for this timeseries we assumed upward propagation for all measured waves. The fact that the horizontal directionality agrees well with the case study example, where westward propagation was assumed, gives us additional confidence in the directionality of our measured GWMF values. Further, we can see from Fig. 8f that during the mountain wave event on 5th July 2015, the variability of meridional flux shows a similar pattern with time in all three data sets, where intermittent bursts of gravity wave activity appear to last a few days at a time. Interestingly, it can be seen in panels shaded percentile regions reveal increased northward and southward meridional momentum fluxes, although the southward component is dominant. This is consistent with the northward and southward components of a characteristic mountain wave field from an island source (e.g. Vosper, 2015).

Panels (g) and (h) that these periods of enhanced flux coincide with periods of northward winds below altitudes of around 20 km down to the surface. This makes sense, since the topography of South Georgia presents a larger cross-section to the prevailing surface wind when the wind vector is directed northward, resulting in a larger southward orographic wave forcing. This feature has been observed in previous studies (e.g. Alexander et al., 2009; Alexander and Grimsdell, 2013; Hindley et al., 2016), where peak fluxes are often found over the south eastern tip of the island. Larger wave amplitudes are generally observed in the southern section of the mountain wave pattern, usually corresponding to a net southward momentum flux. This highlights the importance of resolution in models, since this behaviour would not be replicated if South Georgia was modelled simply as a single grid cell point source in a low-resolution simulation.

The AIRS measurements exhibit a slightly lower tendency towards southward fluxes, with significant northward flux observed in Figs. **??**(a,b). A small fraction of this distribution is likely to be measurement error, but the results may still be significant. The increased northward flux in the AIRS observations is likely to be due to either the presence of more non-orographic waves in the observations, which could be expected to have a more random distribution of wavevector directions, or it may be indicative of a southward bias in the modelled fluxes. This bias could be caused by a northward wind bias in the model winds. Our comparison of model wind speeds to radiosondes observations in Sect. **??** revealed a small northward bias in the model between altitudes of 5 and 10 km, but a more significant southward bias was observed above altitudes around 15 km, so it is not clear if this could be a significant factorin Fig. 8 show the percentage of the total GWMF in region C that was contained in region B, as illustrated in Fig. 4. Since region C is made up of the two regions A and B, both of which have equal area, this percentage provides us with a useful metric for determining how much of the total GWMF was distributed upwind or downwind of the island. This metric is useful because it is consistent for both the AIRS and model-as-AIRS GWMF measurements.

Only one significant period of northward flux is observed in all three data sets at the same time, during 14th June 2015. This coincided with a period of enhanced northward flux in the AIRS observations and a period of southward winds of 10 to 30 ms^{-1} at altitudes from the surface to 60 km in the model.

Selected examples of temperature perturbations from AIRS measurements and the model-as-AIRS during the July 2013 (top) and June-July 2015 (bottom) modelling campaigns. Temperature perturbations are shown as coloured isosurfaces at ± 2 K for AIRS data and ± 1 K for model-as-AIRS, with red indicating positive values and blue negative. As in Fig. **??**, the factor $\kappa(z)$ has been applied to the temperature perturbations in all examples in order to see the vertical wave structure clearly. Red and blue dashed lines illustrate the spatial extent of the AIRS measurements and model domain.

7.2 Examples of simulated and observed gravity wave structures in 3-D

To make a more in-depth comparison of individual cases, we inspect the wave temperature perturbations in 3-D. Fig. **??** shows AIRS and model-as-AIRS gravity wave temperature perturbations for ten selected AIRS overpasses and their corresponding model time steps. Isosurfaces are drawn at ± 2 K for AIRS data and ± 1 K for the model-as-AIRS. As in Fig. **??**, During periods of increased wave activity, a larger percentage of the factor $\kappa(z)$ has been applied to the temperature perturbations in order to see the vertical wave structure clearly. The selected events are arranged in panels in chronological order.

For these select examples, the observed and simulated wave fields are quite similar. Mountain wave structures are clearly visible in both the AIRS and model-as-AIRS data on the 9th, 10th, 27th and total GWMF is usually measured downwind of the island in region B in both datasets. This is a strong indication of mountain wave activity, since we would normally expect non-orographic wave activity would be distributed more evenly over regions A and B, although we acknowledge this may not always be the case. During periods around 29th July 2013 and 14th June 2015 in Fig. **??**, not including the examples on 5th July 2015 shown in Figs. **??** and **??**. The directionality of the mountain wave field can also be inspected. In Figs. **??**(a,b,c) the mountain wave field in AIRS exhibits a slightly southward orientation on 9th, 10th and 29th however, where large GWMF values are measured, over 90% of the total GWMF was contained downwind of the island in region B in both AIRS and the model-as-AIRS. Inspection of the temperature perturbations during these events revealed characteristic bow-wave mountain wave patterns downwind of South Georgia. During periods of relatively low wave activity, such as during 15th–23rd July 2013 , while in Figs. **??**(d, f) more northward orientations are observed on 27th July 2013 and 14th or 19th–24th June 2015. These examples correspond to periods of net southward and northward momentum fluxes respectively in the time series 2015, this percentage is close of 50%, indicating a relatively uniform distribution of GWMF over regions A and B.

The agreement between the AIRS measurements and the model-as-AIRS in Fig. **??**(a,b). This provides confidence in the direction-finding of the 3DST analysis. In these examples in early July 2013 however, very little meridional flux is observed 8 is generally reasonable. The timing and magnitude of increased GWMF found during GW events is similar between the two datasets. However, although GWMF results indicate similar magnitudes, GW temperature amplitudes in the model-as-AIRS . Since these are clear mountain wave structures, it suggests that are consistently around 20–30% smaller than found in AIRS. One reason for this could be due to errors in the speed and direction of the background wind in the model.

One other interesting event occurred at 0300 UTC on 1st July 2015, shown in Fig. **??**. Here, AIRS shows the largest measured wave amplitudes for the whole 2013 and 2015 periods, with 95th percentile values exceeding 8 K. This event corresponded to large westward and southward momentum fluxes of around 10 and 12 mPa respectively in Figs. **??** and **??**. This event on the 1st July 2015 is shown in Fig. **??**j. Here, a large wave structure is observed in AIRS measurements occupying almost the

entirely domain. Dominant wave vectors for this wave structure are orientated in a southward direction. In the model-as-AIRS, a similar structure is observed but wave amplitudes are significantly lower and the wave is orientated further southwards. Significant wave amplitudes are observed both upwind and downwind of the island, particularly in AIRS, where phase fronts aligned in the opposite direction to what we would expect for a mountain wave pattern. Our interpretation is that this is a clear example of a large-scale the use of the area-average. If AIRS measurements exhibit more GW activity at large distances from the island, which could be indicative of non-orographic wave in the region. Other examples of potential non-orographic waves are shown in panels (c),(g), (h) and (i).

The origin of these large-scale non-orographic waves is not clear, but it is clear that their amplitudes appear to be significantly under-estimated in GW activity, this would lead to a larger area-average. But the shaded percentile regions in Fig. 8(a,b), also indicate that the 90th percentile of measured amplitudes in AIRS is consistently larger than in the model-as-AIRS in these examples. These waves may originate from in situ non-orographic processes such as jet adjustment around the edge of the polar vortex, or from intense storms and fronts in the Drake Passage region. Hindley et al. (2019) reported that large-scale gravity waves with very similar characteristics to the example shown in Fig. **??**j were commonly observed in AIRS measurements over the Southern Ocean during winter. The geographic location of South Georgia may also be significant. 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 (e.g. Hoffmann et al., 2013, 2016). One other possibility therefore is that these waves are in fact mountain wave structures that formed over the southern Andes and Antarctic Peninsula but have since become detached due to changing wind conditions, as discussed by Sato et al. (2012); Garfinkel and Oman (2018); Hindley et al. (2019) and others. These waves could also be secondary waves generated as a result of intense primary mountain wave breaking over the southern Andes near the stratopause region (Woods and Smith, 2010; Bossert et al., 2017; Vadas et al., 2018; Becker and Vadas, 2018). These secondary waves would have non-zero phase speeds so could be observed far downwind of their source regions. Further investigation is needed to quantify the relative contributions of these different wave sources observed in the AIRS measurements so that we can determine the reasons for the under-representation of these waves in the model. by a similar amount. This suggests that large amplitude events in AIRS also exhibit larger amplitudes than their counterparts in the model-as-AIRS. These results are discussed further in Sect. 9.

7.2 Horizontal distributions and intermittency of gravity waves around the islandwave amplitude, λ_H and directional GWMF

In Figs. **??**-**??** we found gravity wave activity over the model domain during June-July 2013 and 2015 often occurred in intermittent bursts. As discussed above, orographic waves from South Georgia may only make up one part of this wave activity, so in order to assess this here we investigate of the The horizontal distribution of wave activity around the island in AIRS, the model and the model-as-AIRS.

Figure **??** shows the distributions of measured wave amplitudesas a function of horizontal distance east and west from the model centre. A region is selected that is $y = \pm 250$ km in the meridional direction and between GW properties around South Georgia is shown in Fig. 9. For this analysis, measured GW amplitudes, horizontal wavelengths λ_H and zonal and meridional

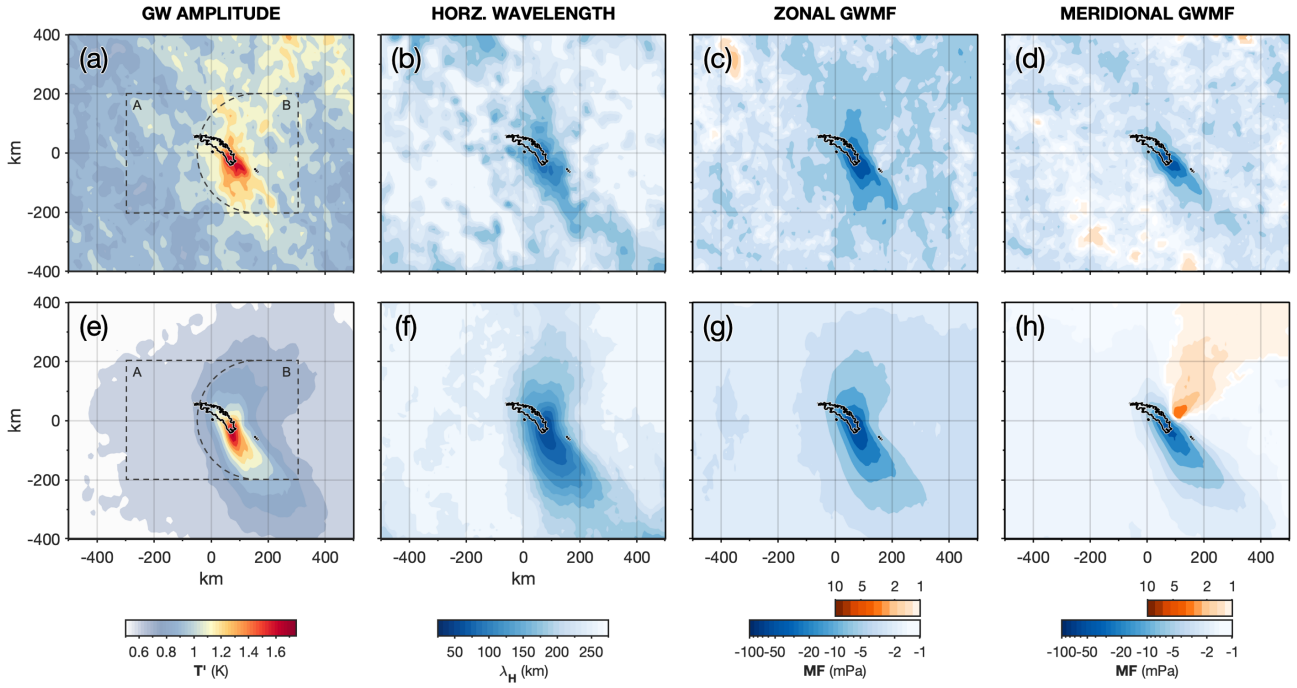

Figure 9. ~~Measured gravity wave~~ Average GW temperature amplitudes ~~as a function of distance east (positive x)~~ T', horizontal wavelengths λ_H and ~~west~~ zonal and meridional momentum flux (~~negative x~~GWMF) MF_x and MF_y over South Georgia from ~~the model centre for (a)~~ AIRS ~~,~~ measurements (~~b~~top row) ~~the model~~ and ~~(c)~~ the model-as-AIRS (bottom row) during ~~June-July~~ both modelling campaigns in July 2013 and June-July 2015. ~~Median wave amplitudes~~ Data are ~~shown by red, black and blue coloured lines in (a-c) for~~ averaged over a ~~meridional region ±250 km and a~~ vertical region between ~~altitudes of~~ 25 km and 45 km ~~, while the shaded regions show the 5th and 95th, 15th and 85th, and 25th and 75th percentiles~~altitude. ~~Panel (d) shows the Gini coefficient of the average absolute gravity wave momentum flux (GWMF) over the same region for all AIRS overpasses and model time steps~~For horizontal wavelengths, ~~which is related to~~ only λ_H measurements for GWs with amplitudes $T' > 1.5$ K are included in the ~~intermittency of GWMF over time~~average. ~~Red, black and blue~~ Black dashed lines in (~~d~~a) ~~correspond to AIRS, the model~~ and ~~the model-as-AIRS respectively. The right hand axis of panel~~ (~~d~~e) ~~also shows~~ show the ~~corresponding percentage~~ extent of ~~wave events that carried 90% of~~ the ~~total GWMF during the 2013 and 2015 campaigns. The topography of South Georgia is shown~~ regions described in ~~black at the bottom of each panel for illustration~~Fig. 4.

momentum fluxes for AIRS and the model-as-AIRS are averaged over 25 to 45 km ~~in the vertical. Median wave amplitudes are found in this region for all AIRS measurements and model time steps during June-July~~ altitude for all measurements during July 2013 and June-July 2015. ~~As in Fig. ??, coloured lines~~ For λ_H, only values for GWs with amplitudes $T' > 1.5$ K are included in the average (Hindley et al., 2019).

Average GW amplitudes in Figs. ~~??(a-c) show the median wave amplitude as a function of horizontal distance east and west of the island. The shaded areas show the extent of the 5th and 95th,15th and 85th and 25th and 75th percentiles of the wave amplitude distribution. The island topography is shown in the bottom centre of each panel to guide the eye.~~

In all three datasets , largest amplitudes are measured just to the east of the island , around 1009(a,e) exceed 1.5 km away from the model centre. Median values of around 1 K are observed at this location in AIRS, but in the K directly over the island in both AIRS and the model-as-AIRS median values peak at around to 0.6for this two-month period. Both datasets exhibit increased GW amplitudes directly over the island and in a region extending around 150 K. In the model results in Fig. **??**b, largest median amplitudes reach 1.6 K just to the east km to the south, but AIRS exhibits regions of increased GW amplitudes further to the north and south in a somewhat disorderly pattern. To the east and west of the island, but the 95th percentile reaches nearly 7GW amplitudes near 0.9 K in the model compared to nearly 3are measured in AIRS, compared to just 0.7 K for AIRS and just over 2 K for in the model-as-AIRS. This indicates that tightly localised wave amplitudes over the island in the model were significantly higher than in AIRS measurements, as we would expect. However,

Because we added specified AIRS retrieval noise to the model-as-AIRS, it is unlikely that this difference is due to noise in AIRS measurements. Instead, it may be due to non-orographic gravity wave (NGW) activity in the real atmosphere that is not well represented in this local-area model configuration. Recent satellite and modelling studies have suggested significant NGW activity can be found in this region during winter (e.g. Sato et al., 2012; Choi and Chun, 2013; Hendricks et al., 2014; Plougonven and Zhar . Even if such NGWs are poorly resolved by AIRS, their partial detection creates general variability and anisotropy in the AIRS temperature perturbations, which are then measured as GW amplitudes in our 3DST analysis. Direct inspection of the AIRS measurements suggests that this effect is quite different from the effects of pixel-scale retrieval noise, and does not appear in the model-as-AIRSvalues did not reach similar values. . This is discussed further in Sect. 9.

The distribution of wave amplitudes with zonal distance from the island in Fig. **??** is somewhat different too. In the model and model-as-AIRS,low amplitude values are found upwind (to the west) and a large spike is seen immediately downwind of the shortest average horizontal wavelengths in Figs. 9(b,f) are found directly over the island, with values decreasing sharply with eastward distance. In the AIRS results in panel (a), the largest amplitudes are also found over and immediately downwind of the island. However, a larger relative distribution of increased wave amplitudes is seen at large horizontal distances from the island than is observed in the model-as-AIRSaround 60 km and 80 km in the model-as-AIRS and AIRS respectively. But caution should be taken when considering time-averaged wavelengths. The characteristic horizontal wavelength of a generalised mountain wave field directly over the island is related to the size of the orographic obstacle in the direction of the prevailing wind. This is indicated by the amplitude percentile regions being spaced somewhat further apart than in the model-as-AIRS. Although wave amplitudes do decrease somewhat with increasing distance from the island (including slightly larger values to the east than the west, as is seen in the model) this tendency towards larger wave amplitudes away from the island is strongly suggestive of relatively large-amplitude around 30–40 km for South Georgia under westerly wind conditions. The fact that both datasets exhibit longer horizontal wavelengths over the island suggests that other (probably non-orographicwave activity measured by AIRS. Although there is significant non-orographic wave activity in the model, the results of Fig. **??** suggest that their amplitudes are much lower in the model-as-AIRS than is observed in reality. Although measurement noise in AIRS is likely to play a part in these results, the increased 85th and 95th percentile values in Fig. **??** are suggestive of large amplitude wave events that would be unlikely to be produced solely by retrieval noise.) waves with longer λ_H are included in the average.

Because AIRS exhibits around 30% longer average horizontal wavelengths over the island than in the model-as-AIRS, this could indicate that NGWs with $T' > 1.5$ K are more often found in the AIRS observations here.

Figure ??d shows the intermittency of absolute gravity wave momentum flux throughout June-July 2013 Zonal GWMF in Figs. 9(c,g) is almost entirely westward, which is consistent with expected propagation of GWs into the background wind. Over the island, westward GWMF exceeds 50 mPa in both datasets. Meridional GWMF in Figs. 9(h) exhibits a north-south divergence in the model-as-AIRS that is centred on the island. This is characteristic of a bow-wave mountain wave field. We recall here that we did not specify this horizontal directionality and only upward propagation was assumed. This further suggests that our assumption of upward propagation for GWs visible to AIRS during winter in this region is generally valid. We acknowledge however that any downwardly propagating waves ($m > 0$) will exhibit the opposite horizontal directionality ($k \rightarrow -k$ and 2015 as a function of zonal distance from the island, using the same region as used in panels (a) to (e). To quantify intermittency, we use the Gini coefficient (Gini, 1912). The Gini coefficient quantifies the unevenness of a distribution using a scalar value between zero and one, and has been used in numerous gravity wave studies in recent years (e.g. Hertzog et al., 2012; Plougonve . Here, a high value for the Gini coefficient implies that momentum fluxes are unevenly distributed into a few large events during the 2013 and 2015 campaigns (high intermittency). A low value for the Gini coefficient implies that fluxes are more evenly distributed into more frequently occurring events of comparable intensity (low intermittency). Generally, orographic wave sources have been found to exhibit higher intermittency than non-orographic sources over long timescales in previous studies (Hertzog et al., 2012; Plougonven et al., 2013; Wright et al., 2013; Hindley et al., 2019).

All three data sets exhibit the highest intermittency immediately to $l \rightarrow -l$) in our analysis due to being mislabelled as upwardly propagating. Our results here however do not suggest that this has a significant effect on the directionality of our measured GWMF over long timescales, and even if such an effect is present it would be equal for both AIRS and the east of the island, with Gini coefficient values near 0.65, 0.92 and 0.75 for the AIRS , model and model-as-AIRSrespectively. This is consistent with the results of previous studies that found orographic gravity wave sources to exhibit higher intermittency than non-orographic sources. Gini coefficient values decrease with increasing distance from the island, but values are higher east than to the west, as a result of the leeward mountain wave field from South Georgia which is more intermittent than non-orographic waves measured upwind. , so it would not affect the validity of our comparison.

Gini coefficient values in the AIRS measurements are significantly lower than the model and Both the AIRS and the model-as-AIRS at all horizontal distances from the island, implying lower intermittencyexhibit large southward GWMF of more than 50 mPa to the south of the island in Figs. 9(d,h), but only the model-as-AIRS exhibits a clear northward component in this time-average, albeit at comparatively weak values of up to 4 mPa. One reason for this is likely to be measurement noise, which would reduce the Gini coefficient value. This is because, during periods of very little wave activity, measurement noise in AIRS would still exhibit some small amplitudes rather than falling to near zero as seen could be due to the small meridional wind bias in the model and model-as-AIRS in Fig. ??. However, given the increased wave amplitudes away from the island in Fig. ??a, it could also be indicative of non-orographic activity in the AIRS measurements that is under-represented in the model .

It is useful to relate these intermittency values to a more meaningful quantity. Following the approach of Hindley et al. (2019) , we can use the Gini coefficient to find the percentage of gravity wave events that contributed to 90% of the total momentum

		Amplitude T' (K)			Zonal MF (mpa)			Eastward	Merid. MF (mpa)	Westward
		A	1.02 B	0.55	-2.28 A	0.85 B	-1.49	43% A	B	1.11
Model	1.06	-1.94 1.58	57% 2.49	A	1.56 -2.09	0.01 -14.53	-3.14	0.35 -0.91	-2.69 -8.40	15%
AIRS	0.49	-16.83 0.97	1.75 1.14	-12.90	85% -1.56	A -5.34	0.61	0.01 -0.70	-1.18 -2.67	0.14
Model as AIRS	B	0.81 0.65	0.17 0.90		-1.10	-5.57		-0.35	-2.24	0.47

Table 1. Average gravity wave Measured GW amplitudes and directional momentum fluxes over in upwind (A) and downwind (B) of South Georgia during June-July 2013 and 2015 for in the full-resolution model, AIRS observations , the model and the model-as-AIRS. Values are shown for two geographical averaged between 25 km and 45 km altitude over regions A and B as shown in (see Fig. 4) for altitudes between 25 to 55 km. The two regions have equal area, all GW measurements during July 2013 and the June-July 2015. The rightmost column shows the percentage fraction of the total absolute momentum flux GWMF in each data set region C that was contained measured in each region A and B. Note that GWMF in the full resolution model is calculated using Eqn. 1 but AIRS and model-as-AIRS GWMF is calculated via Eqn. 2.

flux during both time periods. This is shown on the right hand axis of Fig. **??**d. For a perfectly even distribution, 90% of the total flux would be carried by 90% of the wave events. For an uneven distribution, such as we observe here, this percentage is significantly lower.

We find that in the model, 90% of the total momentum flux over the island was carried by less than wind shown in Sect. 3. We found that the model exhibited a southward wind bias of up to 10% of wave events, compared to around 22% and 32% in the model-as-AIRS and AIRS measurements respectively. This shows that, despite the effects of retrieval noise, the same fraction of the total momentum flux over the island in AIRS measurements was carried by significantly more wave events than in the model or model-as-AIRS. Because the distribution of wave amplitudes with horizontal distance in Figs. **??**a implies increased wave activity away from the island, and particularly upwind, this result could further support the hypothesis that the model is under-representing the contribution of large-scale non-orographic wave events to the total momentum flux over the time period of ms^{-1} between 15 to 30 km altitude compared to coincident radiosonde observations. Although our radiosonde measurements do not extend further than 30 km, it is possible that this observed wind bias could persist to altitudes between 25 and 45 km, where our GWMF measurements in Fig. 9 are shown. This southward wind bias could lead to a stronger northward section of the campaignsimulated mountain wave field than is observed in AIRS, due to the preferential propagation of mountain waves into the background wind.

Illustration of the two regions to the east and west of South Georgia used to produce the values in Table 1. Region A is upwind of the island and Region B is over and downwind of the island. The two regions have equal area.

7.3 Gravity wave properties upwind and downwind of South Georgia

To further explore the question of orographic versus non-orographic wave activity in the region, we divide the model domain into two regions A and B, as shown in Fig. 4. These two regions have equal area of 453750 km^2, and together they form a rectangular region that is 1100 km × 825 km as shown. Region A is designed to capture non-orographic wave activity upwind (to the west) of the island, while region B is designed to capture both orographic and non-orographic wave activity over and downwind (to the east) of the island. The shape of region B consists of a rectangle and half circle of diameter 825 km. This shape was chosen to be a simple representation of the mountain wave field region (see Fig. **??**(d-f)) with a horizontal area that is straightforward to calculate. In the vertical, these regions extend between altitudes of 25 and 45 km, which is the same height range as used in Figs. **??** to **??**.

Table 1 shows The results of Fig. 9(a-f) are summarised in Table 1 over the two region A and B. Here, average wave amplitudes , directional momentum fluxes and the percentage of the total momentum flux in each region A and B during June-July and net GWMF are shown for AIRS, the model-as-AIRS and full-resolution model for all GW measurements during July 2013 and June-July 2015. As in Figs. **??** and **??**, campaign-average directional flux values are found by taking the average of all the positive flux values with the negative values set to zero and vice versa to give the true area average. The larger area of these regions mean that this approach yields values that are somewhat lower than those shown in previous figures. Because these two regions have equal area, the average directional flux over both regions is simply the average of the two values shown for regions A and B in Table 1. The net flux over both regions is then the sum of the average eastward and westward (or northward and southward) values. Note that amplitudes and GWMF in the full-resolution model are not directly comparable to values in the AIRS or model-as-AIRS, due to the different observational filter and processing methods, but they are included for context.

Average All three datasets exhibit larger wave amplitudes and fluxes are significantly larger downwind of the island net GWMF in region B than upwind (downwind) than in region A in all three data sets, but values also differ significantly between the data sets. For example, in AIRS measurements the net zonal momentum flux in region B is -2.97 mPa, but (upwind), but average wave amplitudes in region B in the model-as-AIRS it is -2.07are around 20% smaller than found in AIRS. Despite this, average GWMF values in region B are similar, where the magnitude of the net flux $\left(\mathrm{MF}_x^2 + \mathrm{MF}_y^2\right)^{\frac{1}{2}}$ in both datasets is around 6 mPa, some 31% lower.

Further contrast is found when we compare the fraction of the total momentum flux in each dataset contained within region A and region B. Interestingly, 85% of the total absolute momentum flux in the model and 72% in the model-as-AIRS was found in region B , but in the AIRS measurements only 57% of the total flux was found in region B compared to 43% in region A. If retrieval noise in AIRS is assumed to be uniformly distributed both upwind and downwind of the island, then such noise will not affect this valuemPa. This suggests that gravity wave momentum flux is significantly more evenly distributed over the whole region in AIRS measurements than was simulated in the model. When the AIRS observational filter is applied, the because average λ_H over the island is longer in AIRS than in the model-as-AIRSpercentage becomes closer to the AIRS value for region B, but is still 15% higher.

This result reinforces some of the earlier findings of this study. Namely, the model seems to underestimate the momentum flux associated with non-orographic waves, or perhaps overestimates the flux associated with orographic waves from South

Georgia, when compared to AIRS measurements. Direct inspection of the gravity wave fields in the model do reveal significant non-orographic wave features but, as shown above, wave amplitudes tend to be around 20 to 25% lower in the model-as-AIRS compared to observations. Since this is also true for mountain waves in the model-as-AIRS, we suspect therefore that it is indeed an underestimation of non-orographic wave activity in the model rather than an overestimation of orographic wave activity that leads to this discrepancy, the larger average wave amplitudes in AIRS do not lead to larger GWMF values via Eqn. 2.

Cross-sections of the mean absolute gravity-wave momentum flux (GWMF) from the AIRS observations, the full-resolution model and the model-as-AIRS for both campaigns during July 2013 and June-July 2015. The top row shows a horizontal cross-section through the domain at 30 km altitude, while the bottom row shows a vertical cross-section at $y = 0$ km, where y is the meridional direction. Panels (d) and (h) show the differences between mean GWMF from the AIRS observations and the model-as-AIRS for the horizontal and vertical cross-sections respectively. Both panels in each column share the same colour scale below. The black lines to the side of panels (e), (g) and (h) illustrate the AIRS vertical measurement window. The topography of South Georgia is shown in black at the bottom of panels (e-h).

7.3 Campaign-mean momentum fluxes

A useful quantity to constrain is the long-timescale mean of stratospheric momentum flux from South Georgia. These values are useful for simplified parameterisation schemes and model tuning for climate simulations.

Figure ?? shows horizontal and vertical cross-sections through the campaign-mean absolute momentum flux derived from the AIRS, full-resolution model and model-as-AIRS data for both June-July 2013 and 2015. The horizontal cross-sections reveal, as expected, localised maxima in momentum flux over the islandextending eastwards and to the south in all three data sets, with values in AIRS and model-as-AIRS peaking exceeding 15 mPa. The model-as-AIRS flux in Fig. ??c is however more tightly localised over the south-eastern part The rightmost column of Table 1 shows the fraction of the total absolute GWMF measured upwind and downwind of the island. Around 35% of the total GWMF in AIRS is found upwind of the island than is observed in AIRS in Fig. ??a. As a result, the area-average momentum flux values in region A, compared to only 17% in the model-as-AIRSare lower. The AIRS flux is centred on the same location, but large flux values are more spatially distributed into the beginnings of a characteristic bow-wave shape. This reduced area-average flux in the model-as-AIRS can also be seen in Fig. ?? and ?? in Sect. 7.1, where model-as-AIRS fluxes are generally lower than AIRS measurements during bursts of coincident wave activity. As we would expect, both the AIRS observations and the model-as-AIRS exhibit flux values that are more than an order of magnitude lower than in the model, whose values exceed 600 mPa in Fig. ??b, and are tightly localised over the same region as the model-as-AIRS (the colour scale is saturated to show the spatial distribution clearly).

These results suggest that the downwind bow wave pattern formed to . Further, the east of the island observed in AIRS measurements occurs is under-represented by the model simulation, since the downwind mountain wave pattern is far less apparent in the magnitude of the net GWMF in the upwind region is around 45% larger in AIRS than in the model-as-AIRS. Fig. ??d highlights this difference between the campaign-average . These results indicate that the model-as-AIRS may underestimate NGW activity upwind of South Georgia compared to observations.

There is also a small difference in the direction $\tan^{-1}(\mathrm{MF}_x/\mathrm{MF}_y)$ of the net GWMF in region B between the AIRS and model-as-AIRS~~fluxes. The AIRS observations exhibit around 10 mPa more flux away from the island to the south east than~~, which exhibit directions of 243° and 248° clockwise from north respectively. Although these directions are close, this indicates a small northward bias in the model-as-AIRS, ~~but around 5 mPa less over the island itself.~~ which could be related to a southward wind bias in the background stratospheric wind, as discussed in Sect. 7.2 above.

~~The vertical cross-sections in Figs. **??**(e,f,g) reveal a similar picture, with a localised maximum of momentum flux directly over the south-eastern part of the island. With the AIRS vertical measurement window applied, the AIRS and~~

7.3 Wave amplitude growth with height

The results in previous sections show persistent differences in measured wave amplitudes between AIRS and the model-as-AIRS~~fluxes maximise between altitudes of around 20 to 30 km before slowly decreasing with increasing altitude, while the~~. To investigate how these differences vary with altitude, Fig. 10 shows vertical profiles of measured GW amplitudes in AIRS, the full-resolution ~~model shows an intense vertical column of flux from the surface to around 40 km altitude before doing the same.~~ model and the model-as-AIRS averaged over region B during June 2013 and June-July 2015.

~~The vertical difference plot in Fig. **??**h shows that between altitudes of 20 to 30~~Average wave amplitudes in AIRS are up to 0.4 ~~km AIRS measurements exhibit slightly more flux just to the east of the island but significantly less directly over the island compared to the~~ K larger than in the model-as-AIRS ~~. The model-as-AIRS also exhibits fluxes around 2 to 10 mPa larger than AIRS in a localised vertical column directly over the island~~ at all altitudes ~~, but smaller fluxes over a horizontal area average. Again, the high resolution model fluxes are at least an order of magnitude larger than observed in AIRS and the model-as-AIRS, with average values exceeding 300~~up to around 45 ~~mPa at altitudes around 30~~ km~~and more than 1000 mPa below 10 km (once again the colour scale is saturated to show the spatial distribution clearly). This is due to large wave amplitudes at very short horizontal wavelengths in the model that are not visible to AIRS~~. As shown in Fig. 9(a,e), this is likely due to larger GW amplitudes found at large distances from the island in AIRS, which increases the area average. Interestingly however, although stratospheric GW amplitudes increase exponentially with altitude in all three datasets, but they appear to increase at different rates.

~~Vertical profiles of the campaign-mean absolute gravity-wave momentum flux (a) and measured wave amplitudes (b,c) against height for the AIRS observations (red), full-resolution model (black) and model-as-AIRS (blue) for June-July 2013 and 2015. Panel (c) is as panel (b) but with a logarithmic x-axis. Grey diagonal lines in (c) show the exponential adiabatic growth of gravity-wave amplitudes with altitude $e^{\frac{z}{2H}}$ expected from theory for an atmospheric scale height of $H = 7$ km. The dashed blue line indicates the start of the model damping layer which begins at $z = 58.5$ km and extends to the model top.~~

7.4 ~~Wave amplitude and momentum flux growth with altitude~~

~~Figure **??** shows campaign-mean measured wave amplitudes and absolute momentum fluxes against height for a vertical column with horizontal radius $r = 50$ km centred on a location 50 km east of the model centre. This region was selected so as to capture the peak of the flux distributions based on the results of the cross-sections in Fig. **??**.~~

Campaign-mean fluxes in Fig. **??**a show reasonable agreement in AIRS and model-as-AIRS derived fluxes between altitudes of Figure 10b shows the same data as 10a but on a logarithmic amplitude scale. Between 25and 45 km , although the AIRS fluxes are slightly higher, consistent with what we have seen in previous sections. Above altitudes of and 45 km however, the AIRS fluxes are lower than the model-as-AIRS values. The model exhibits flux values just over an order of magnitude higher than AIRS altitude, the model and the model-as-AIRS . This is due to the fine horizontal scale structure in the wave field, as illustrated in Fig. **??**b. All three data sets show a general decrease in gravity-wave momentum flux with increasing altitude, except for the lowest region of the AIRS fluxes where an increase is observed between altitudes of 20 and 25closely follow the expected exponential adiabatic amplitude growth with height as $e^{\frac{z}{2H}}$ (thin grey lines), where $H = 7$ km . This is due to the half-bell tapering functions that we applied to the upper and lower boundaries of usable AIRS measurements (shown on the right hand side of panel (c)), which slowly reduced the perturbations to zero below these altitudes as illustrated. This reduction is not physical but an artefact of our method, and we would expect that in reality the red line denoting AIRS flux is the approximate scale height of the atmosphere. Linear fits to the curves in Fig. **??**a is likely to continue to increase below 10b between altitudes of 25 km altitude following the blue and 45 km altitude yield gradients of 0.032 and 0.028 for the full-resolution model and the model-as-AIRS line. respectively. These values are close to the gradient of around 0.031 (thin grey lines) that denotes theoretical exponential growth with height.

Figures **??**b and **??**c show campaign-mean measured wave amplitude against height in the same vertical column over South Georgia. In the model and model-as-AIRS, wave amplitudes are found to decrease slightly with increasing heightfrom the surface to the tropopause, then increase steadily through the stratosphere up to altitudes around 40AIRS GW amplitudes however are found to increase more slowly with height. This is particularly evident even for altitudes between 25 and 35 km, as expected from exponential amplitude growth of $e^{\frac{z}{2*H}}$ under dry conditions, where $H = 7$ km is the average scale height for the atmosphere. Following the approach of Wright et al. (2016b), thin grey diagonal lines in Fig. **??**c show this theoretical growth .

Wave amplitudes in the model and model-as-AIRS despite the fact that this is where the AIRS vertical resolution is best (∼7–8 km, Hoffmann and Alexander, 2009). The amplitude growth rate in AIRS is the smallest over this height range. But this reduced growth rate is consistent with growth rates in GW potential energy during winter from limb-sounding observations as found by Wright et al. (2016b). A linear fit of the AIRS curve in Fig. **??**c closely follow lines of exponential growth between altitudes of 20 to 4010b between 25 and 45 km . Above 40 km altitude however, exponential amplitude growth altitude yields a gradient of 0.015, approximately half the growth rate found in the modeldeclines sharply, along with a smaller decline in the model-as-AIRS. This is likely to be related to the wave field in the model reaching saturation above these altitudes as the wave fields experience super-adiabatic conditions. The reduced measured amplitudes in the spectral range of the . The fact that the model-as-AIRS data as shown in Fig. **??**b would make this condition less likely, so the reduction is less abrupt. Above an altitude of 58.5 km, the model damping layer is applied as described in Vosper (2015), so results for measured wave amplitudes are does not follow the same reduced growth rate as in the AIRS observations indicates that this difference is not likely to be physical at these heights. Note also that the reduction in wave amplitude below 25 km altitude is due to our usable vertical measurement window and not likely physical as mentioned above.

An interesting observation is that between altitudes of 25 due to 45 km, average wave amplitudes in AIRS measurements increase slower with increasing altitudethan both the model and the model-as-AIRS in Fig. **??**c, and significantly slower than expected from idealised conditions using a scale height of $H = 7$ km. This could be indicative of increased changes in AIRS vertical resolution or retrieval noise with altitude. The reduced growth rate in AIRS could simply be because the scale height in the real atmosphere during this time period was greater than in the model, or it could indicate that some wave breaking, saturation or dissipation effects that occur in the real atmosphere that are not accurately simulated in the model, or it could suggest that the atmospheric scale height in this region was greater than 7 km during this period. It is also possible that this could be a result of changes in the AIRS vertical resolution or noise levels at these altitudes, but the AIRS measurement noise actually increases significantly above altitudes of 50 km (Hoffmann and Alexander, 2009; Hindley et al., 2019) so this is unlikely to be the cause of this reduction in wave amplitudes. If it were due to reduced AIRS vertical resolution with altitude, we would expect to see a similar sharp reduction in the model-as-AIRS measured amplitudes, which we do not. We therefore suspect this change may be due to a physical process. . If the vertical resolution of our local-area model is too coarse, GWs are prevented from dissipating and would continue to increase in amplitude exponentially with altitude.

Given the results of the previous sections, and in particular taking into account the variability in the radiosonde-observed winds, we propose that this sharper reduction in AIRS amplitudes above 45 km altitudeis due to a reduced likelihood of mountain wave structures being stable up to altitudes above 45 km in the real atmosphere due to variability in surface and stratospheric winds.

7.4 Distribution of GWMF with wave amplitude and horizontal wavelength

8 Large amplitude mountain waves at short horizontal scales

Our final analysis concerns the spectral properties of the observed and simulated wave fields. In the previous sections we have investigated the distribution of gravity wave amplitudes and momentum fluxes in the geographical region around South Georgia. Here we investigate how the campaign-average momentum fluxes are distributed as a function of wave amplitude and horizontal wavelength . This analysis will help us to understand what features of the wave field may give rise to the discrepancies reported earlier in the study.

For a range of wave amplitudes between 0 and 25

In this section we consider GW measurements over the island at the very shortest horizontal scales visible to AIRS. Large-amplitude mountain waves are generally expected either directly above or just downwind of an orographic obstacle. The horizontal wavelength for the central region of a mountain wave field is primarily determined by the size of the obstacle in the direction of the prevailing wind, which is around 30–40 K and a range of horizontal wavelengths between 25 and 1000 km, the average absolute momentum flux is found for each amplitude and wavelength combination during all overpasses and timesteps in June-July 2013 and 2015. Average values are taken over the same geographical region as used in Figs. **??** to **??** and between altitudes of 25 to 45 km . As before, regions that do not have a particular amplitude and wavelength combination are set to zero before the average is taken to give the true area average. This yields values thatare quite low compared to values

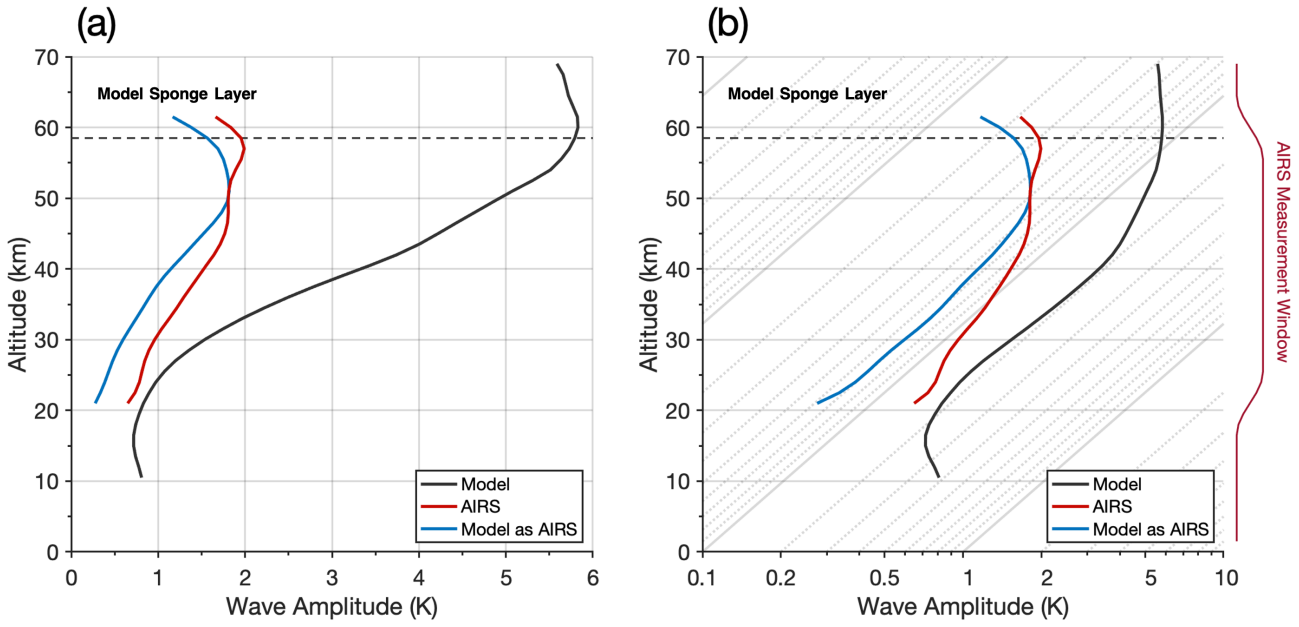

Figure 10. Vertical profiles of the average measured wave amplitudes over South Georgia in AIRS observations (red), the full-resolution model (black) and the model-as-AIRS (blue) for June-July 2013 and 2015. Values are averaged over a horizontal area 600×400 km centred on the island (region C in Fig. 4). Both panels (a) and (b) show the same data, but in (b) the data are plotted on a logarithmic x axis where thin grey diagonal lines show the exponential adiabatic growth rate of GW amplitude with altitude $e^{\frac{z}{2H}}$ expected from theory for an atmospheric scale height of $H = 7$ km. The model damping "sponge" layer (dashed black line) begins at $z = 58.5$ km and extends to the model top.

~~shown in previous figures. The results of this analysis for AIRS~~measurements, ~~the full-resolution~~ for South Georgia. These large amplitude and short horizontal wavelength waves can carry large momentum fluxes. In this section we show that, under favourable viewing conditions, AIRS can observe these waves.

Figure 11(a-c) shows AIRS, model and model-as-AIRS ~~are shown in Fig. ??.~~

~~In each panel in Fig. ??, colours indicate the area-averaged momentum flux for measured waves with a specific amplitude and wavelength. In panel (a), two distributions of waves become clear in the model: one group with large amplitudes up to nearly 25~~temperature measurements at 45 ~~K and horizontal wavelengths between 30 and 75~~km altitude over South Georgia during an overpass at 1700 ~~km and a second group with amplitudes below around 5 K and horizontal wavelengths between 100 and 300 km. Given the horizontal scale of the first group, and the results in~~ UTC on 5th July 2015. As in Fig. ~~??, this is likely to correspond to the very~~ 5, coloured circles indicate the location and extent of the AIRS measurement footprints. Alternating red-blue circles close to the island are indicative of large amplitude GWs at the Nyquist sampling limit of the AIRS instrument, with a horizontal wavelength of around two AIRS footprints (around 2×13.5 km). Normally, we would be suspicious of such wave detections in AIRS measurements due to the retrieval noise, but the orientation of these features, their large magnitudes and their proximity to the island show close agreement with the expected mountain wave field in the full resolution model.

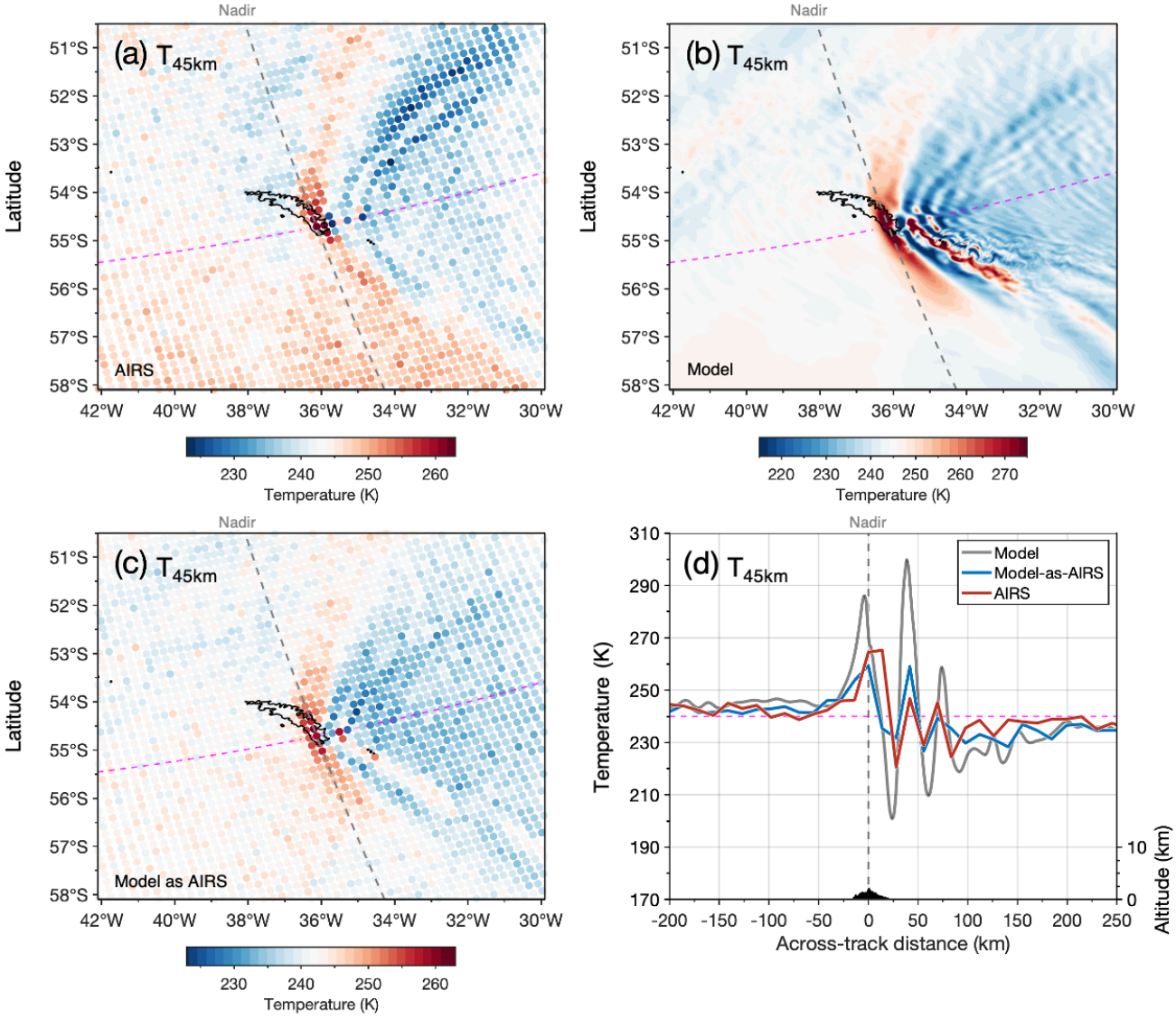

Figure 11. ~~Average absolute gravity-wave momentum flux as a function of wave amplitude and horizontal wavelength λ_H~~ Temperature measurements over ~~the whole model domain between altitudes of 25 to 55~~South Georgia at 45 km altitude at 1700 UTC on 5th July 2015 for (a) ~~the~~ AIRS ~~measurements~~satellite observations, ~~the~~ (b) ~~the~~ full-resolution model and (c) the model-as-AIRS~~during June-July 2013~~. Coloured circles in (a) and ~~2015.~~ (c) show the locations of the AIRS measurement footprints, while dashed grey lines show the satellite nadir, which passes directly over the island in this example. Panel (d) shows measured temperature at 45 km altitude against horizontal across track distance along the ~~difference between~~ pink dashed line in (a-c). Surface elevation along this path (right hand axes) is shown in black at the ~~AIRS measurements and the model-as-AIRS~~bottom of (d).

Inspection of the AIRS sampling pattern in Fig. 11a reveals three preferential conditions for the measurement of short horizontal scale ~~waves directly over the island, while the second group likely corresponds to the mountain wave wake region and non-orographic wave activity in the model.~~

~~In the AIRS and the model-as-AIRS distributions in panels (b)and (c), only the second group of waves is apparent, where largest average momentum fluxes occur for horizontal wavelengths between 150 and 200~~mountain waves over South Georgia. Firstly, this overpass occurred during intense mountain wave activity in our study, as shown in Figs. 2 and 8. GWMF values near to this overpass are some of the largest measured during the time period studied here. Secondly, the nadir of the AIRS scan track passed directly over the the island, as shown by the dashed grey line in Figs. 11(a-c). The horizontal sampling between adjacent AIRS footprints is closest at nadir (~13.5 ~~km. This peak occurs at amplitudes between around 0.5 and 2 K in AIRS measurements, but typically below 1 K in the model-as-AIRS. Panel (d) shows the difference between the AIRS and model-as-AIRS flux distributions, where up to 50% more average flux is observed in AIRS for amplitudes larger than around 1 K and horizontal wavelengths longer than around 100 km. The only part of the distribution where slightly more average flux is found in the model-as-AIRS occurs for amplitudes between around 3 and 6 K and horizontal wavelengthsnear 75 km. This increased flux towards larger amplitudes and shorter horizontal wavelengths is apparent in the model-as-AIRS distribution in panel (c)~~km), which provides the best possible horizontal resolution for GWs.

Thirdly, we can see from Fig. 11a that the across-track scan direction is aligned perpendicular to the central section of the chevron-shaped mountain wave field, where the horizontal wavelengths are shortest. This means that these across track rows, shown by the dashed pink line, bisect the mountain wave field perpendicular to the GW phase fronts in the central section of the mountain wave field, providing the most favourable viewing geometry for these short horizontal wavelengths. The orientation of the mountain wave field over the island, which is ~~of course simply a subset of the model distribution, and is similar to that of the model before the observational filter of AIRS removes flux from short horizontal scale waves to near zero~~strongly related to the direction of the prevailing wind, does not always preferentially align with the across track scan direction in such a way. For all the AIRS overpasses in our study that were inspected, only this example on 5th July 2015 showed such a clear alignment.

~~These distributions suggest an interesting conclusion. The AIRS distribution~~ The full-resolution model in Fig. ~~??b does not show the same tendency towards larger amplitudes at horizontal wavelengths shorter than 100~~11b exhibits a mountain wave field with fine-horizontal scale structure. Short horizontal wavelengths near 30–40 km ~~in the same way that the model and model-as-AIRS does. In the model and model-as-AIRS, these shorter horizontal wavelength waves with large amplitudes correspond to the mountain wave field directly~~ are found over and immediately downwind of the island~~itself, as shown in Figs. 6 and ??. The fact that the AIRS measurements do not appear to follow this distribution could suggest that such short horizontal wavelength waves over the island, which carry the largest momentum fluxes individually via Eqn. 2, are not as observed so commonly in the real atmosphere as they are in the model.~~

~~This is not likely to be an AIRS horizontal resolution issue, since the island is usually located near to the centre of the AIRS swath for the overpasses we have selected in Sect. 2.2, where the spacing between AIRS pixels is as low as 15 to 20 km, yielding aNyquist resolution of around 40 km. It is interesting therefore that in the~~ , while turbulent eddies are apparent on the southern part of the mountain wave field. When the AIRS sampling and resolution is applied to the model, the mountain wave

structure in panel (c) exhibits good qualitative agreement with the AIRS ~~distribution we do not see the beginning of distribution of the first group of short-wavelength waves as seen in the model in panel (a) or the model-as-AIRS in panel (c). This is is also unlikely to be an artefact introduced by the application of the AIRS observational filter to~~ observations in panel (a). As was shown in Fig. 6 above, the AIRS observations exhibit more mountain wave structure at larger horizontal distances to the north and south of the ~~model to generate the~~ island.

We next take an across-track cut along the dashed pink line at 45 km altitude through all three datasets in Fig. 11(a-c) to show temperature against horizontal distance from the island in Fig. 11d. Because it is not straightforward to extract GW perturbations from model temperatures in the same way as the AIRS and model-as-AIRS~~, as described in Sect. ??. This is because the same method is followed for both data sets: measurements are gridded onto our 15~~, we present raw temperature measurements here to avoid any artefacts that may arise from inconsistent background removal methods for this example.

Temperatures rise and fall with increasing horizontal distance downwind of the island to the east in all three datasets with an apparent horizontal wavelength close to ~30–40 km, and there is good agreement in GW phase. The full resolution model exhibits large temperature perturbations of around ±45 ~~×~~K above and below background temperatures of ~245 K upwind of the island. When the AIRS sampling and resolution are applied, the model-as-AIRS exhibits perturbations close to ±15 ~~km regular horizontal grid then convolved with a Gaussian with a horizontal FWHM equal to 40~~K. These are in reasonable agreement with AIRS measurements, which exhibit temperature perturbations of around 15–20 ~~km × 40 km. The model is downsampled onto this grid, whereas the AIRS measurements are slightly upsampled, so if any artefacts were introduced by this process we would expect to see an underestimation of the flux from short horizontal scale waves in the model-as-AIRS due to the downsampling. We do not see evidence of this, so it is likely that this increased in flux at short horizontal scales~~ K. Interestingly, positive perturbations in AIRS are larger than in the model-as-AIRS ~~compared to AIRS is physical.~~

~~One possible reason for this could be due to variability in the background wind. Mountain wave fields require a stable vertical column of strong tropospheric and stratospheric winds in order to propagate vertically into the stratosphere without encountering critical levels. Here, hourly boundary conditions for the local-area domain are provided by a global forecast which is initialised at midnight each day and integrated forward in time in hourly steps. These hourly boundary conditions are then linearly interpolated in time to the time step of the local-area simulation. One issue with this approach is that variability of the background winds on timescales shorter than one hour will not be present in the local-area model . Further, the reduced vertical grid of the global forecast, which has only 70 vertical levels, will mean that the boundary conditions will exhibit reduced wind variability in the vertical.~~

~~These factors could result in reduced variability of~~ in the ~~background winds in the local-area model. This could potentially provide more favourable conditions for~~ first positive peak directly over the island, but smaller in the second peak at around 40 km to the ~~generation~~ east. The third peak exhibits comparable perturbation amplitudes. This discrepancy could be caused by a slight offset in the position of the mountain wave structure in the model compared to observations, or a slightly different horizontal wavelength, which could affect the measured wave amplitude when the AIRS sampling is applied to the model.

Several conclusions can be drawn from Fig. 11. Overall, the physical scale of observed and modelled mountain wave structures over the island shows good qualitative agreement for this example. The full-resolution model (1.5 km horizontal

grid) indicates that the shortest characteristic horizontal wavelength of mountain waves ~~with short horizontal scales, since it is~~ directly over the island is around 30–40 km. When the AIRS sampling pattern is aligned preferentially, AIRS can resolve these short horizontal scale waves~~that would the be first to be disrupted by sudden changes in the wind vector. This could explain the discrepancy in Fig. ??d, where the model-as-AIRS exhibited more momentum flux at~~ , as shown in AIRS measurements and the model-as-AIRS.

This is significant because the GW temperature perturbations over the island at these short horizontal wavelengths ~~and less at long horizontal wavelengths than observed in AIRS measurements.~~

9 ~~Discussion~~

~~In this study we used a range of different analysis techniques to investigate the similarity between gravity waves in a high resolution local area modelover South Georgia and coincident satellite observations. The results of these analyses, as listed in their respective sections above, are quite varied. Here we combine the key findings of this study to discuss them in a broader context.~~

~~Our interpretation of the results in~~ in this ~~study is that , when allowed to run at a high spatial resolution, a mountain wave field over the island tends to form in our local area model configuration that is somewhat over-idealised compared to observations. Specifically, this idealised wave field in the model consists of a central region of large amplitude, short horizontal wavelength waves that is tightly located over the island. Our results in Figs. ??, ?? and ?? indicate that this feature is not so commonly seen in observations.~~

~~This over-localisation of flux over the island in the model could arise due to reduced short-timescale variability in the background wind. As discussed earlier,~~ example are very large. If we had only found these waves in the model, we may ask the question of whether they could be supported in the ~~lateral boundary conditions of the local-area model are supplied hourly from an N512 global forecast with reduced vertical resolution. Variability in the background wind over timescales shorter than one hour, or over short vertical scales, might not be accurately simulated in the local-area model. This could provide more favourable conditions for the generation of short horizontal scale mountain waves than is observed in the real atmosphere. Although our radiosonde comparison in Sect. ?? showed reasonable agreement between observed and simulated winds over monthly time scales, it was found that significant discrepancies could occur for individual measurements.~~ real atmosphere. But because measured amplitudes in AIRS and the model-as-AIRS show reasonable agreement, this suggests that the magnitude of wave amplitudes close to 45 K in the full resolution model is realistic. These temperature perturbations of around $\pm$45 K in the full resolution model correspond to large horizontal and vertical wind perturbations near 45 km altitude of up to $\pm$80 ms^{-1} and $\pm$60 ms^{-1} respectively. To our knowledge, stratospheric GW temperature perturbations of this magnitude at such small horizontal scales are rarely seen in satellite observations. This is partly due to the limited observational filters of spaceborne instruments, but as we can see in Fig. 11 viewing geometry and horizontal sampling are also important.

~~We also observe a greater distribution of gravity wave activity at larger horizontal distances from~~ Recent ground-based lidar observations austral during winter over the well-known GW hot spot of the southern Andes by Kaifler et al. (2020) revealed

GW temperature perturbations near ±40 K with vertical wavelengths around 16–18 km. Comparison with operational models in their study found good agreement with between simulated and observed GWs. But the horizontal scale of the large-amplitude GWs measured in Kaifler et al. (2020) was close to $\lambda_H \approx$400 km, which is around 10 times larger than we find over South Georgia in ~~Sect. ??, particularly upwind of the island to the west~~Fig. 11. This is ~~a strong indication of an under-representation of the amplitudes and fluxes of non-orographic waves in the model. These waves, whose physical scales~~ an important result, because is suggests that the GWMF of such large-amplitude, short-λ_H waves over small mountainous islands can be very large ~~, as seen in Fig. ??j or in Fig. 1c of Hindley et al. (2019), were found to have around at least 50% larger amplitudes in the AIRS observations than in the~~ and that high resolution and favourable sampling is required in models and observations to simulate and measure this GWMF.

9 Discussion

9.1 Model-as-AIRS: Sensitivity to horizontal sampling and retrieval noise

One key process in this study is the application of the AIRS resolution and horizontal sampling to the model to create model-as-AIRS ~~. The lack of these large scale non-orographic waves could also contribute to~~ temperature perturbations that can be directly compared to AIRS observations. We found that our results were highly sensitive to the accuracy of this procedure. It is not enough to simply apply the horizontal and vertical resolutions of AIRS to the ~~overly-stable wind vector that allows fine-scale horizontal mountain wave structure to form in the model~~model; we must ensure that accurate horizontal sampling is also applied. This is because ~~a large amplitude non-orographic wave passing through the region in the real atmosphere could disrupt the background horizontal wind vector, resulting in non-linear interaction and dissipation of the fine horizontal scale mountain wave field~~short-horizontal scale ($\lambda_H \lesssim 40$ km) GWs directly over the island, which carry large momentum fluxes, are easily resolved on the model grid but are not always resolved in the AIRS observations due to limitations in sampling caused by inconsistent viewing geometry, as shown in Sect. 8. Therefore, if the AIRS horizontal sampling were not applied, the model-as-AIRS would always overestimate these short-λ_H GWs compared to observations.

~~We also showed in Fig. ?? that the short-timescale variability, or intermittency, of gravity wave momentum flux is significantly lower in AIRS observations than in either the high resolution model or the~~ It is perhaps counter-intuitive to apply unwanted retrieval noise to model output, but we found that this was also an essential step in our comparison. By applying the specified AIRS retrieval noise to the model-as-AIRS, we can cancel out any effects of noise in our regional comparisons, specifically the comparison of upwind and downwind GW properties. Because GW temperature perturbations cannot always be separated from noise perturbations, the more retrieval noise that is present in measurements, the more even the distribution of GWMF between the upwind and downwind regions in Table 1. If we did not apply retrieval noise to the model-as-AIRS, a fair comparison would not be possible.

We should note however that the specified retrieval noise that we applied is randomised to uncorrelated pixel-scale noise for each altitude level, so if there are elements of retrieval noise in AIRS measurements that have larger horizontal scales greater than around 30-50 km, these may not be correctly applied to the model-as-AIRS. ~~Although retrieval noise in the AIRS~~

measurementsmay play a part in these results, this result is consistent with a more uniform distribution of gravity wave activity over time in observations. In the model, mountain waves may be somewhat over represented, leading to discrete bursts of gravity waves activity and higher intermittency during our period of study. An under-representation of large-scale waves from non-orographic sources in

9.2 Simulation of NGWs in the local-area model

In Sect. 7.2 we found that, compared to the AIRS observations, the model-as-AIRS may under-estimate NGW wave activity at large horizontal distances from the island, particularly upwind. This is significant because de la Cámara et al. (2016) recently showed that an even balance between orographic and NGW parametrisations near 60°S had a significant impact on reducing the cold-pole biases. They found that sporadic large-amplitude NGW parameterisations from specific sources provide greater forcing on the circulations than a homogeneous distribution of NGW parameterisations.

Hindley et al. (2019, their Fig. 1) reported that sporadic large-amplitude NGWs can often be found in AIRS observations around the Southern Ocean during winter. We also find suggestions of such waves in our Fig. 5a, so their apparent underestimation in our local-area model is important. As discussed in Sect. 6, the global forecast that supplies the lateral and initial boundary conditions for the local-area model has a coarser vertical and horizontal resolution, with only 70 vertical levels from the model, which have been found to exhibit lower intermittency than orographic sources, would also be consistent with our results. One possiblecause for reduced from transitory non-orographic wave activity in the model could be that these waves are present in the hourly global forecast, but they do not survive the transmission from the forecast grid to the surface to near 80 km and a horizontal grid spacing close to 60 km at latitudes near South Georgia. Even if NGWs are realistically simulated in the global forecast, it is not clear how well these waves would be "transmitted" through the interface between the global forecast and the local-area domain through the model boundary conditions.

Thus, our answer to the question posed in Sect. 1 is two-fold. Generally speaking, the high-resolution model over South Georgia used here is found to produce a gravity wave field over South Georgia that is in good agreement with coincident observations model. Further, the time integration used between the global forecast and the local-area model timesteps may further invalidate the realism of any transferred waves. As a result, non-stationary NGWs generated outside the local-area model, such as those from storms, jets, fronts and geostrophic adjustment processes, are unlikely to be realistically simulated in the local-area model. This is a consequence of the nested model configuration used here, which is designed to produce realistic wind conditions over South Georgia for mountain wave generation. The timing and directionality of gravity wave momentum fluxes closely matches observations. Although area-averaged momentum fluxes are around 25% lower than observed, the agreement is significantly better than would be expected in a coarser resolution free-running GCM. This indicates that increasing model resolution can improve the representation of gravity waves from small islands through the accurate determination of grid cells containing mostly land or sea, as discussed by Vosper et al. (2016). This leads to further improvements such as the accurate direction of the meridional momentum flux from South Georgia, shown in Sect. 7.1, which is observed to be overwhelmingly southward in both the model and observations. This is important because, if the island orography was underrepresented in a lower resolution GCM such that it was effectively a point source, the net meridional flux for a westerly wind vector would be

close to zero. The high-resolution simulations allow the orientation of the island's topography to be properly represented, with a larger cross-section to southwesterly winds. This results in a significantly larger southward component of meridional flux, which is in good agreement with the observations without the need for modelparameterisations.

However, we can also conclude from our results that, if the background wind vector is not accurately simulated, increasing the model spatial resolution can lead to inaccurate gravity wave generationat different horizontal scales. Further, for a local region considered in this study, accurate simulation of transitory non-orographic gravity waves propagating in and out of If the horizontal extent of the region is also needed in order to produce a realistic local field of gravity wave activity. local-area simulation and the number of vertical levels in the global forecast are increased, we would expect that transitory NGWs would be better simulated in the local-area model.

It is also important to note that it is not just model resolution which is important for accurate gravity wave simulations. Model numerics can also be significant. The Unified Model, for example, Met Office Unified Model used here uses semi-implicit time integration for operational efficiency, but choosing too large a time step timestep can make the model dissipative to gravity waves (e.g. Shutts and Vosper, 2011; Vosper, 2015). This GWs (e.g. Shutts and Vosper, 2011; Vosper, 2015), which could lead to an underestimation of gravity wave amplitudesin the model, which could partly explain the reduced amplitudes and momentum fluxes in the model-as-AIRS compared to AIRS measurements. However, we find that wave amplitudes appear to decay with height at a consistent rate time-averaged GW amplitudes directly over the island in Fig. ?? in the model and 9 appear to show a reasonable agreement between AIRS and the model-as-AIRS, which could suggest suggesting that this effect may be small . As mentioned above, it is interesting to note that the AIRS amplitudes appear to decay faster with height than the is small for mountain waves, which have ground-based horizontal phase speeds close to zero.

9.3 Large-amplitude mountain waves directly over the island

In Sect. 8 we found good agreement between the AIRS and model-as-AIRS for the shortest characteristic horizontal wavelengths around 30–40 km for mountain waves directly over South Georgia. Here, these waves have large amplitudes up to 20 K in AIRS measurements and can carry large momentum fluxes.

But these GWs at short horizontal wavelengths lie at the sampling and resolution limits of AIRS measurements. They are only visible in Fig. ??, suggesting that wave amplitudes may, in general, dissipate more quickly with altitude in the real atmosphere than in the model11 due to the favourable viewing geometry of this specific AIRS overpass, where the satellite nadir passes directly over the island and the across track direction is aligned parallel (perpendicular) to the background wind vector (GW phase fronts). Because these conditions are not the same for each overpass, this means that the GWMF from these large amplitude, short-λ_H waves may be underestimated in recent AIRS GWMF climatologies (Hindley et al., 2020) . This underscores the importance of considering how instrument sampling patterns contribute to the observational filters of spaceborne GW measurements, and further highlights that future comparisons between models and observations should consider both horizontal sampling and resolution (Wright and Hindley, 2018).

10 Conclusions

In this study, we ~~posed the fundamental question: when a numerical model is allowed to run at very high spatial resolution over a~~ compare simulated stratospheric gravity waves (GWs) over the small mountainous island ~~like South Georgia , how realistic are the simulated gravity waves compared to observations? The answer to this question is key to~~ of South Georgia to coincident 3-D AIRS satellite observations. Such islands currently lie in the "grey zone" of global model resolution, where they are neither fully resolved nor fully parameterised (Vosper, 2015; Vosper et al., 2016). Thus, critically assessing simulated GW momentum fluxes generated by these islands is crucial for the development of ~~the future generations of global climate models. As spatial resolution is increased, a larger portion of the gravity wave spectrum can be resolved, thus reducing the reliance on model parameterisations which are poorly constrained by observations. The increased gravity wave fluxes from these resolved waves may in turn lead to significant reductions in model biases, such as the cold-pole problem. In the shorter term, high resolution modelling experiments are being used as reference simulations to test gravity wave parameterisations (both orographic and non-orographic) so it is important to understand their validity (Vosper, 2015; Vosper et al., 2016).~~ accurate future global models.

~~To answer this question, we compared simulated gravity waves in a high resolution local area model over South Georgia to coincident 3-D satellite observations from AIRS/Aqua. We applied the AIRS observational filter~~ We use a local-area model configuration with a high spatial resolution (1.5 km horizontal grid, 118 vertical levels) that can resolve the mountainous orography of the island and accurately simulate mountain wave generation and propagation. We apply the sampling and resolution of AIRS to the model ~~output to produce a~~ to create a "model-as-AIRS~~dataset that could be compared the observations directly. Our results show that:~~ " dataset. This allows us to make direct like-for-like comparisons of simulated and observed GW amplitudes, wavelengths and directional momentum fluxes during two periods in July 2013 and June-July 2015. We find that:

1. ~~Overall, the timing and magnitude of gravity wave~~ The timing of GW activity in the ~~model-as-AIRS is in reasonable agreement with observations, but area-averaged momentum fluxes can be up to around 25% lower than observed in AIRS measurements.~~ local-area model generally agrees well with the AIRS observations. This suggests that mountain wave forcing, propagation and background winds in the model are accurately simulated to first order, and that the 1.5 km horizontal grid is sufficient to generate realistic stratospheric mountain waves.

2. ~~In both the model and observations, meridional momentum fluxes~~ When the model is sampled as AIRS, good agreement is found in net GW momentum flux (GWMF) over the island~~are overwhelmingly southward, something that would not be accurately simulated if the island was under-resolved in a GCM.~~ . Average zonal (meridional) GWMF over this two-month period is westward (southward) at 5.3 mPa (2.7 mPa) and 5.6 mPa (-2.2 mPa) in AIRS and the model-as-AIRS respectively.

3. ~~Directly over the island, the~~ Both peak and area-averaged GW amplitudes in the model-as-AIRS ~~exhibits higher individual flux measurements but is more intermittent than the observations, with 90%~~ are ~20–30% smaller than seen in AIRS. Upwind of the island, 35% of the ~~total flux carried by just 22% of wave events, compared to 32% for AIRS.~~

4. ~~72% of the~~ total ~~flux is located downwind of the island in the model-as-AIRS~~GWMF is found in AIRS, compared to only ~~57~~17% in the ~~AIRS measurements~~model-as-AIRS. This suggests that although the model configuration used here simulates realistic orographic GWMF over the island, it under-estimates non-orographic ~~wave activity observed in AIRS is under-represented in the model , which is supported by directed inspection of the wave fields.~~ GW activity over the surrounding ocean.

5. ~~Gravity wave momentum fluxes in AIRS~~appear to dissipate more quickly with increasing height than in the Average GW amplitudes in AIRS, but not the model-as-AIRS~~, which could suggest that the model may under-represent~~ , are found to increase more slowly with height than expected from theory. This could be because simulated wave breaking or dissipation processes in the model are incomplete, either due to insufficient vertical grid spacing or underrepresented wave-wave or wave-mean flow interactions.

6. ~~Spectral analysis results, combined with spatial distributions, indicate that our high resolution model runs over-estimate the fluxes of large amplitude, short horizontal scale mountain waves directly over the island, but under-estimate the fluxes of large amplitude, large horizontal scale waves in the surrounding region. This could be due toreduced short timescale variability of the background wind in the model.~~

7. ~~Model winds exhibit a slight southward bias of around 5~~ A ~20% northward bias in meridional GWMF is found in the model-as-AIRS. This bias could be related to, or even caused by, a large southward wind bias of up to 10 ~~ms^{-1} with increasing altitude~~ ms^{-1} in the model compared to coincident radiosonde observations ~~.~~ at altitudes above ~10 km.

8. Finally, AIRS measurements reveal large-amplitude (T' ~15–20 K at 45 km altitude) mountain waves with λ_H ~30–40 km directly over the island. These waves are at the shortest horizontal scales visible to AIRS, and are only detectable due to favourable viewing geometry during one specific overpass. AIRS-measured λ_H and T' for this example show excellent agreement with the model-as-AIRS. This example provides valuable experimental evidence that large-amplitude (up to $T' \sim 45$ K at 45 km altitude) short horizontal wavelength ($\lambda_H \sim 30 - -40$ km) mountain waves, as seen in the full-resolution model here, are physical and can occur in the real atmosphere.

~~In conclusion, we find that although increasing the horizontal resolution of models can improve the representations of small islands such as South Georgia, this can also lead to an over-representation of small-scale mountain wave activity and an under-representation of large-scale non-orographic waves. Our results suggest that , in order to produce a realistic simulation of wintertime stratospheric gravity wavesover small islands in the Southern Ocean in future GCMs, a holistic approach is needed. This approach should include~~

Despite the increasing availability of global GW observations in recent years, direct comparisons of GWs in observations and models have been limited by several fundamental factors, including: (a) ~~sufficient spatial resolution to resolve the island~~

~~topography; (b) accurate wind speed, direction and short-timescale variability over the island and~~ the observational filter problem; (b) a lack of the 3-D observations needed to constrain the directionality of GW momentum fluxes; (c) insufficient model resolution to accurately resolve small-scale GWs; and (d) the need for realistic background wind conditions for specific time periods for dedicated high-resolution offline simulations.

In this study, we have overcome each of these obstacles to make accurate and detailed comparisons between observed and simulated GWs over the mountainous island of South Georgia. We find that, for a high-resolution real-time simulation that is guided by a global forecast, good agreement can be found between simulated wintertime GWs and coincident 3-D GW observations if the observational filter of the ~~wider geographical region; and (c) accurate simulation of large-scale~~ instrument is carefully applied to the model. In particular, we show that when the sampling pattern is orientated favourably with respect to the wave, agreement between GWs in AIRS observations and the model can be excellent.

Some important biases do remain between the model and the observations however. Specifically, our model configuration exhibits directional biases and underestimates non-orographic ~~waves from sources such as geostrophic adjustment processes around the vortex edge that may be found in the stratosphere over small islands during winter. As the spatial resolution of future GCMs is increased ever further in the coming years, we must ensure that each of these factors is implemented correctly in order to produce realistic gravity wave characteristics over small islands in the Southern Ocean~~GW activity in the region compared to observations. These discrepancies likely arise from the nested local-area configuration used here, and may be greatly reduced in a global model operating at this spatial resolution. As such models become available in the future, our study points to an effective way forward for future comparisons of GWs in high-resolution models and observations.

Author contributions. The Met Office Unified Model simulations were designed and carried out by SBV, ANR, and JKH. The specialised 3-D temperature data from the AIRS satellite was developed and provided by LH. The radiosonde field campaign on South Georgia was conducted by CJW and ACM with logistical support from TMG and JCK. Data analyses, scientific investigations and publication figures were developed and performed by NPH. The written manuscript was prepared by NPH, CJW and NJM with additional contributions from all authors. This investigation was carried out as part of the South Georgia Wave Experiment (SG-WEX) project, designed by NJM, AMG, DRJ, JCK, TMG and ANR.

Competing interests. The authors declare that they have no competing interests

Acknowledgements. The SG-WEX project is supported by the UK Natural Environment Research Council (NERC) under grant numbers NE/K015117/1, NE/K012584/1, NE/K012614/1, NE/R001391/1 and NE/R001235/1 and a Royal Society University Research Fellowship UF160545 supporting CJW. The authors would like to acknowledge the use of supercomputing and data archival systems at several institutions, with a special thanks to their support staff. These are: the Met Office and NERC Supercomputing Node (MONSooN), the JASMIN data analysis platform at the UK Centre for Environmental Data Archival (CEDA), the Managed Archive Storage System (MASS) at the UK Met

Office, the BALENA high-performance computing service at the University of Bath and the JURECA high-performance computing service at Forschungszentrum Jülich. We ~~would also like to acknowledge~~ acknowledge the logistical staff at British Antarctic Survey involved in the radiosonde campaign and the cooperation of the government of South Georgia and the South Sandwich Islands. ~~Finally, we would like to~~ We also thank the NASA AIRS instrument team~~and~~ , the many teams involved in developing the Met Office Unified Model, and Annelize van Niekerk and Matthew Griffith for useful discussions about the model. Finally, we would like to thank the three anonymous reviewers for their helpful suggestions and comments.

References

Alexander, M. J.: Global and seasonal variations in three-dimensional gravity wave momentum flux from satellite limb-sounding temperatures, Geophysical Research Letters, 42, 6860–6867, https://doi.org/10.1002/2015GL065234, 2015.

Alexander, M. J. and Barnet, C.: Using satellite observations to constrain parameterizations of gravity wave effects for global models., J. Atmos. Sci., 64, 1652–1665, https://doi.org/10.1175/JAS3897.1, 2007.

Alexander, M. J. and Grimsdell, A. W.: Seasonal cycle of orographic gravity wave occurrence above small islands in the Southern Hemisphere: Implications for effects on the general circulation, J. Geophys. Res., 118, 11 589–11 599, https://doi.org/10.1002/2013JD020526, 2013.

Alexander, M. J., Gille, J., Cavanaugh, C., Coffey, M., Craig, C., Eden, T., Francis, G., Halvorson, C., Hannigan, J., Khosravi, R., Kinnison, D., Lee, H., Massie, S., Nardi, B., Barnett, J., Hepplewhite, C., Lambert, A., and Dean, V.: Global estimates of gravity wave momentum flux from High Resolution Dynamics Limb Sounder observations, Journal of Geophysical Research, 113, https://doi.org/10.1029/2007jd008807, https://doi.org/10.1029/2007jd008807, 2008.

Alexander, M. J., Eckermann, S. D., Broutman, D., and Ma, J.: Momentum flux estimates for South Georgia Island mountain waves in the stratosphere observed via satellite, Geophy. Res. Lett., 36, L12 816, https://doi.org/10.1029/2009GL038587, 2009.

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., Sato, K., Eckermann, S., Ern, M., Hertzog, A., Kawatani, Y., Pulido, M., Shaw, T. A., Sigmond, M., Vincent, R., and Watanabe, S.: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models, Quart. J. Roy. Meteor. Soc., 136, 1103–1124, https://doi.org/10.1002/qj.637, 2010.

Alexander, P., Luna, D., Llamedo, P., and de la Torre, A.: A gravity waves study close to the Andes mountains in Patagonia and Antarctica with GPS radio occultation observations., Annal. Geophys., 28, 587–595, 2010.

Alexander, P., de la Torre, A., Schmidt, T., Llamedo, P., and Hierro, R.: Limb sounders tracking topographic gravity wave activity from the stratosphere to the ionosphere around midlatitude Andes, Journal of Geophysical Research: Space Physics, 120, 9014–9022, https://doi.org/10.1002/2015ja021409, 2015.

Alexander, P., de la Torre, A., Kaifler, N., Kaifler, B., Salvador, J., Llamedo, P., Hierro, R., and Hormaechea, J. L.: Temperature Profiles From Two Close Lidars and a Satellite to Infer the Structure of a Dominant Gravity Wave, Earth and Space Science, 7, e2020EA001 074, https://doi.org/10.1029/2020EA001074, 2020.

Alexander, S. P., Klekociuk, A. R., and Tsuda, T.: Gravity wave and orographic wave activity observed around the Antarctic and Arctic stratospheric vortices by the COSMIC GPS-RO satellite constellation, J. Geophys. Res., 114, https://doi.org/10.1029/2009JD011851, d17103, 2009.

Aumann, H., Chahine, M., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., Revercomb, H., Rosenkranz, P., Smith, W., Staelin, D., Strow, L., and Susskind, J.: AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems, IEEE Transactions on Geoscience and Remote Sensing, 41, 253–264, https://doi.org/10.1109/TGRS.2002.808356, 2003.

Baldwin, M. P., Birner, T., Brasseur, G., Burrows, J., Butchart, N., Garcia, R., Geller, M., Gray, L., Hamilton, K., Harnik, N., Hegglin, M. I., Langematz, U., Robock, A., Sato, K., and Scaife, A. A.: 100 Years of Progress in Understanding the Stratosphere and Mesosphere, Meteorological Monographs, 59, 27.1–27.62, https://doi.org/10.1175/amsmonographs-d-19-0003.1, 2018.

Becker, E. and Vadas, S. L.: Secondary Gravity Waves in the Winter Mesosphere: Results From a High-Resolution Global Circulation Model, Journal of Geophysical Research: Atmospheres, 123, 2605–2627, https://doi.org/10.1002/2017JD027460, 2018.

Bossert, K., Kruse, C. G., Heale, C. J., Fritts, D. C., Williams, B. P., Snively, J. B., Pautet, P.-D., and Taylor, M. J.: Secondary gravity wave generation over New Zealand during the DEEPWAVE campaign, Journal of Geophysical Research: Atmospheres, 122, 7834–7850, https://doi.org/10.1002/2016JD026079, 2017.

Brown, R. A., Lauzon, M. L., and Frayne, R.: A General Description of Linear Time-Frequency Transforms and Formulation of a Fast, Invertible Transform That Samples the Continuous S-Transform Spectrum Nonredundantly, IEEE Transactions on Signal Processing, 58, 281–290, https://doi.org/10.1109/TSP.2009.2028972, 2010.

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., Kushner, P. J., Newman, P. A., Osprey, S. M., Perlwitz, J., Sigmond, M., Wang, L., Akiyoshi, H., Austin, J., Bekki, S., Baumgaertner, A., Braesicke, P., Brühl, C., Chipperfield, M., Dameris, M., Dhomse, S., Eyring, V., Garcia, R., Garny, H., Jöckel, P., Lamarque, J.-F., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Pawson, S., Plummer, D., Pyle, J., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., Tian, W., Waugh, D., and Yamashita, Y.: Multimodel climate and variability of the stratosphere, J. Geophys. Res., 116, n/a–n/a, https://doi.org/10.1029/2010JD014995, d05102, 2011.

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Bblaisdell, J., Chen, L., Divakarla, M., Fetzer, E. J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F. W., Kakar, R., Kalnay, E., Lambrigtsen, B. H., Lee, S.-y., Le Marshall, J., Mcmillan, W. W., Mcmillin, L., Olsen, E. T., Revercomb, H., Rosenkranz, P., Smith, W. L., Staelin, D., Strow, L. L., Susskind, J., Tobin, D., Wolf, W., and Zhou, L.: AIRS: Improving Weather Forecasting and Providing New Data on Greenhouse Gases, Bulletin of the American Meteorological Society, 87, 911–926, https://doi.org/10.1175/BAMS-87-7-911, 2006.

Choi, H.-J. and Chun, H.-Y.: Effects of Convective Gravity Wave Drag in the Southern Hemisphere Winter Stratosphere, Journal of the Atmospheric Sciences, 70, 2120–2136, https://doi.org/10.1175/JAS-D-12-0238.1, 2013.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, Quarterly Journal of the Royal Meteorological Society, 131, 1759–1782, https://doi.org/10.1256/qj.04.101, 2005.

de la Cámara, A., Lott, F., Jewtoukoff, V., Plougonven, R., and Hertzog, A.: On the Gravity Wave Forcing during the Southern Stratospheric Final Warming in LMDZ, Journal of the Atmospheric Sciences, 73, 3213–3226, https://doi.org/10.1175/jas-d-15-0377.1, 2016.

de la Torre, A. and Alexander, P.: The Interpretation of Wavelengths and Periods as Measured from Atmospheric Balloons, Journal of Applied Meteorology, 34, 2747–2754, https://doi.org/10.1175/1520-0450(1995)034<2747:TIOWAP>2.0.CO;2, 1995.

de la Torre, A. and Alexander, P.: Gravity waves above Andes detected from GPS radio occultation temperature profiles: Mountain forcing?, Geophy. Res. Lett., 32, L17 815, https://doi.org/10.1029/2005GL022959, 2005.

de la Torre, A., Alexander, P., Llamedo, P., Menéndez, C., Schmidt, T., and Wickert, J.: Gravity waves above the Andes detected from GPS radio occultation temperature profiles: Jet mechanism?, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL027343, 2006.

de la Torre, A., Alexander, P., Hierro, R., Llamedo, P., Rolla, A., Schmidt, T., and Wickert, J.: Large-amplitude gravity waves above the southern Andes, the Drake Passage, and the Antarctic Peninsula, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD016377, 2012.

de la Torre, A., Alexander, P., Schmidt, T., Llamedo, P., and Hierro, R.: On the distortions in calculated GW parameters during slanted atmospheric soundings, Atmospheric Measurement Techniques, 11, 1363–1375, https://doi.org/10.5194/amt-11-1363-2018, 2018.

Eckermann, S. D. and Preusse, P.: Global Measurements of Stratospheric Mountain Waves from Space, Science, 286, 1534–1537, https://doi.org/10.1126/science.286.5444.1534, 1999.

Ern, M., Preusse, P., Alexander, M. J., and Warner, C. D.: Absolute values of gravity wave momentum flux derived from satellite data, J. Geophys. Res., 109, D20 103, https://doi.org/10.1029/2004JD004752, 2004.

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

Fritts, D. C. and Alexander, M. J.: Gravity wave dynamics and effects in the middle atmosphere, Reviews of Geophysics, 41, 1003, https://doi.org/10.1029/2001RG000106, 2003.

Fritts, D. C., Riggin, D. M., Balsley, B. B., and Stockwell, R. G.: Recent results with an MF radar at McMurdo, Antarctica: Characteristics and variability of motions near 12-hour period in the mesosphere, Geophy. Res. Lett., 25, 297–300, https://doi.org/10.1029/97GL03702, 1998.

Fritts, D. C., Vadas, S. L., Wan, K., and Werne, J. A.: Mean and variable forcing of the middle atmosphere by gravity waves, Journal of Atmospheric and Solar-Terrestrial Physics, 68, 247 – 265, https://doi.org/doi.org/10.1016/j.jastp.2005.04.010, vertical Coupling in the Atmosphere/Ionosphere System, 2006.

Garcia, R. R., Smith, A. K., Kinnison, D. E., de la Camara, A., and Murphy, D. J.: Modification of the Gravity Wave Parameterization in the Whole Atmosphere Community Climate Model: Motivation and Results, Journal of the Atmospheric Sciences, 74, 275–291, https://doi.org/10.1175/JAS-D-16-0104.1, 2017.

Garfinkel, C. I. and Oman, L. D.: Effect of Gravity Waves From Small Islands in the Southern Ocean on the Southern Hemisphere Atmospheric Circulation, Journal of Geophysical Research: Atmospheres, 123, 1552–1561, https://doi.org/10.1002/2017JD027576, 2018.

Gille, J., Barnett, J., Whitney, J., Dials, M., Woodard, D., Rudolf, W., Lambert, A., and Mankin, W.: The high resolution dynamics limb sounder (HIRDLS) experiment on aura, in: Infrared Spaceborne Remote Sensing XI, edited by Strojnik, M., vol. 5152 of *Proceedings of SPIE-The International Society for Optical Engineering*, pp. 162–171, SPIE, https://doi.org/10.1117/12.507657, Conference on Infrared Spaceborne Remote Sensing XI, SAN DIEGO, CA, AUG 06-08, 2003, 2003.

Gini, C.: Variabilità e mutabilità (Variability and mutability, Tipografia di Paolo Cuppin, 1912.

Goodyear, B. G., Zhu, H. M., Brown, R. A., and Mitchell, J. R.: Removal of phase artifacts from fMRI data using a Stockwell transform filter improves brain activity detection, Magnetic Resonance in Medicine, 51, 16–21, https://doi.org/10.1002/mrm.10681, 2004.

Hendricks, E., Doyle, J., Eckermann, S. D., Jiang, Q., and Reinecke, P.: What Is the Source of the Stratospheric Gravity Wave Belt in Austral Winter?, J. Atmos. Sci., 71, 1583–1592, https://doi.org/10.1175/JAS-D-13-0332.1, 2014.

Hertzog, A., Boccara, G., Vincent, R. A., Vial, F., and Cocquerez, P.: Estimation of gravity wave momentum flux and phase speeds from quasi-Lagrangian stratospheric balloon flights. Part II: Results from the Vorcore campaign in Antarctica., J. Atmos. Sci., 65, 3056–3070, https://doi.org/10.1175/2008JAS2710.1, 2008.

Hertzog, A., Alexander, M. J., and Plougonven, R.: On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere, J. Atmos. Sci., 69, 3433–3448, https://doi.org/10.1175/JAS-D-12-09.1, 2012.

Hierro, R., Steiner, A. K., de la Torre, A., Alexander, P., Llamedo, P., and Cremades, P.: Orographic and convective gravity waves above the Alps and Andes Mountains during GPS radio occultation events – a case study, Atmospheric Measurement Techniques, 11, 3523–3539, https://doi.org/10.5194/amt-11-3523-2018, https://www.atmos-meas-tech.net/11/3523/2018/, 2018.

Hindley, N. P., Wright, C. J., Smith, N. D., and Mitchell, N. J.: The southern stratospheric gravity wave hot spot: individual waves and their momentum fluxes measured by COSMIC GPS-RO, Atmos. Chem. Phys., 15, 7797–7818, https://doi.org/10.5194/acp-15-7797-2015, 2015.

Hindley, N. P., Smith, N. D., Wright, C. J., Rees, D. A. S., and Mitchell, N. J.: A two-dimensional Stockwell transform for gravity wave analysis of AIRS measurements, Atmospheric Measurement Techniques, 9, 2545–2565, https://doi.org/10.5194/amt-9-2545-2016, 2016.

Hindley, N. P., Wright, C. J., Smith, N. D., Hoffmann, L., Holt, L. A., Alexander, M. J., Moffat-Griffin, T., and Mitchell, N. J.: Gravity waves in the winter stratosphere over the Southern Ocean: high-resolution satellite observations and 3-D spectral analysis, Atmospheric Chemistry and Physics, 19, 15 377–15 414, https://doi.org/10.5194/acp-19-15377-2019, 2019.

Hindley, N. P., Wright, C. J., Hoffmann, L., Moffat-Griffin, T., and Mitchell, N. J.: An 18-Year Climatology of Directional Stratospheric Gravity Wave Momentum Flux From 3-D Satellite Observations, Geophysical Research Letters, 47, https://doi.org/10.1029/2020gl089557, 2020.

Hoffmann, L. and Alexander, M. J.: Retrieval of stratospheric temperatures from Atmospheric Infrared Sounder radiance measurements for gravity wave studies, J. Geophys. Res., 114, D07 105, https://doi.org/10.1029/2008JD011241, 2009.

Hoffmann, L., Xue, X., and Alexander, M. J.: A global view of stratospheric gravity wave hotspots located with Atmospheric Infrared Sounder observations, J. Geophys. Res., 118, 416–434, https://doi.org/10.1029/2012JD018658, 2013.

Hoffmann, L., Grimsdell, A. W., and Alexander, M. J.: Stratospheric gravity waves at Southern Hemisphere orographic hotspots: 2003–2014 AIRS/Aqua observations, Atmos. Chem. Phys., 16, 9381–9397, https://doi.org/10.5194/acp-16-9381-2016, 2016.

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

Holt, L. A., Alexander, M. J., Coy, L., Liu, C., Molod, A., Putman, W., and Pawson, S.: 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, 2481–2495, https://doi.org/10.1002/qj.3101, 2017.

Höpfner, M., Larsen, N., Spang, R., Luo, B. P., Ma, J., Svendsen, S. H., Eckermann, S. D., Knudsen, B., Massoli, P., Cairo, F., Stiller, G., v. Clarmann, T., and Fischer, H.: MIPAS detects Antarctic stratospheric belt of NAT PSCs caused by mountain waves, Atmos. Chem. Phys., 6, 1221–1230, https://doi.org/10.5194/acp-6-1221-2006, 2006.

Hu, S., Ma, S., Yan, W., Hindley, N. P., Xu, K., and Jiang, J.: Measuring Gravity Wave Parameters from a Nighttime Satellite Low-Light Image Based on Two-Dimensional Stockwell Transform, Journal of Atmospheric and Oceanic Technology, 36, 41–51, https://doi.org/10.1175/JTECH-D-18-0092.1, 2019a.

Hu, S., Ma, S., Yan, W., Hindley, N. P., and Zhao, X.: Measuring internal solitary wave parameters based on VIIRS/DNB data, International Journal of Remote Sensing, 0, 1–12, https://doi.org/10.1080/01431161.2019.1608389, 2019b.

Jackson, D. R., Gadian, A., Hindley, N. P., Hoffmann, L., Hughes, J., King, J., Moffat-Griffin, T., Moss, A. C., Ross, A. N., Vosper, S. B., Wright, C. J., and Mitchell, N. J.: 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, 1027–1040, https://doi.org/10.1175/BAMS-D-16-0151.1, 2018.

Jiang, J. H., Wu, D. L., and Eckermann, S. D.: Upper Atmosphere Research Satellite (UARS) MLS observation of mountain waves over the Andes, J. Geophys. Res., 107, https://doi.org/10.1029/2002JD002091, 2002.

Kaifler, N., Kaifler, B., Dörnbrack, A., Rapp, M., Hormaechea, J. L., and de la Torre, A.: Lidar observations of large-amplitude mountain waves in the stratosphere above Tierra del Fuego, Argentina, Scientific Reports, 10, https://doi.org/10.1038/s41598-020-71443-7, 2020.

Kim, Y., Eckermann, S. D., and Chun, H.: An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather prediction models, Atmosphere-Ocean, 41, 65–98, https://doi.org/10.3137/ao.410105, 2003.

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R.: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, J. Geophys. Res., 102, 23 429–23 465, https://doi.org/10.1029/97JD01569, 1997.

Kuyuk, H. S.: On the use of Stockwell transform in structural dynamic analysis, Sadhana-Academy Proceedings in Engineering Sciences, 40, 295–306, https://doi.org/10.1007/s12046-014-0301-2, 2015.

Lehmann, C. I., Kim, Y.-H., Preusse, P., Chun, H.-Y., Ern, M., and Kim, S.-Y.: Consistency between Fourier transform and small-volume few-wave decomposition for spectral and spatial variability of gravity waves above a typhoon, Atmos. Meas. Tech., 5, 1637–1651, https://doi.org/10.5194/amt-5-1637-2012, 2012.

Lilienthal, F., Jacobi, C., Schmidt, T., Torre, A., and Alexander, P.: On the influence of zonal gravity wave distributions on the Southern Hemisphere winter circulation, Annales Geophysicae, 35, 785–798, https://doi.org/10.5194/angeo-35-785-2017, 2017.

Liu, X., Xu, J., Yue, J., Vadas, S. L., and Becker, E.: Orographic Primary and Secondary Gravity Waves in the Middle Atmosphere From 16-Year SABER Observations, Geophysical Research Letters, 46, 4512–4522, https://doi.org/10.1029/2019GL082256, 2019.

Llamedo, P., de la Torre, A., Alexander, P., Luna, D., Schmidt, T., and Wickert, J.: A gravity wave analysis near to the Andes Range from GPS radio occultation data and mesoscale numerical simulations: Two case studies, Advances in Space Research, 44, 494–500, https://doi.org/10.1016/j.asr.2009.04.023, https://doi.org/10.1016/j.asr.2009.04.023, 2009.

Llamedo, P., Salvador, J., de la Torre, A., Quiroga, J., Alexander, P., Hierro, R., Schmidt, T., Pazmiño, A., and Quel, E.: 11 Years of Rayleigh Lidar Observations of Gravity Wave Activity Above the Southern Tip of South America, Journal of Geophysical Research: Atmospheres, 124, 451–467, https://doi.org/10.1029/2018JD028673, 2019.

McDonald, A. J.: Gravity wave occurrence statistics derived from paired COSMIC/FORMOSAT3 observations, J. Geophys. Res., 117, D15 106, https://doi.org/10.1029/2011JD016715, 2012.

McLandress, C., Shepherd, T. G., Scinocca, J. F., Plummer, D. A., Sigmond, M., Jonsson, A. I., and Reader, M. C.: Separating the Dynamical Effects of Climate Change and Ozone Depletion. Part II: Southern Hemisphere Troposphere, Journal of Climate, 24, 1850–1868, https://doi.org/10.1175/2010JCLI3958.1, 2011.

McLandress, C., Shepherd, T. G., Polavarapu, S., and Beagley, S. R.: Is Missing Orographic Gravity Wave Drag near 60°S the Cause of the Stratospheric Zonal Wind Biases in Chemistry–Climate Models?, J. Atmos. Sci., 69, 802–818, https://doi.org/10.1175/JAS-D-11-0159.1, 2012.

Meyer, C. I. and Hoffmann, L.: Validation of AIRS high-resolution stratospheric temperature retrievals, Proc. SPIE, 92420, 9242–9242–10, https://doi.org/10.1117/12.2066967, 2014.

Moffat-Griffin, T., Wright, C. J., Moss, A. C., King, J. C., Colwell, S. R., Hughes, J. K., and Mitchell, N. J.: The South Georgia Wave Experiment (SG-WEX): radiosonde observations of gravity waves in the lower stratosphere. Part I: Energy density, momentum flux and wave propagation direction, Quarterly Journal of the Royal Meteorological Society, 143, 3279–3290, https://doi.org/10.1002/qj.3181, 2017.

Perlwitz, J.: Tug of war on the jet stream, Nature Climate Change, 1, 29–31, https://doi.org/10.1038/nclimate1065, 2011.

Plougonven, R. and Zhang, F.: Internal gravity waves from atmospheric jets and fronts, Reviews of Geophysics, 52, 33–76, https://doi.org/10.1002/2012RG000419, 2014.

Plougonven, R., Hertzog, A., and Guez, L.: Gravity waves over Antarctica and the Southern Ocean: consistent momentum fluxes in mesoscale simulations and stratospheric balloon observations, Quart. J. Roy. Meteor. Soc., 139, 101–118, https://doi.org/10.1002/qj.1965, 2013.

Plougonven, R., de la Cámara, A., Hertzog, A., and Lott, F.: How does knowledge of atmospheric gravity waves guide their parameterizations?, Quarterly Journal of the Royal Meteorological Society, 146, 1529–1543, https://doi.org/10.1002/qj.3732, 2020.

Polichtchouk, I. and Scott, R. K.: Spontaneous inertia-gravity wave emission from a nonlinear critical layer in the stratosphere, Quarterly Journal of the Royal Meteorological Society, 146, 1516–1528, https://doi.org/10.1002/qj.3750, 2020.

Preusse, P., Dörnbrack, A., and Eckermann, S.: Space-based measurements of stratospheric mountain waves by CRISTA 1. Sensitivity, analysis method, and a case study, J. Geophys. Res., 107, 8178, https://doi.org/10.1029/2001JD000699, 2002.

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

Sato, K., Tateno, S., Watanabe, S., and Kawatani, Y.: Gravity Wave Characteristics in the Southern Hemisphere Revealed by a High-Resolution Middle-Atmosphere General Circulation Model., J. Atmos. Sci., 69, 1378–1396, https://doi.org/10.1175/JAS-D-11-0101.1, 2012.

Sato, K., Tsuchiya, C., Alexander, M. J., and Hoffmann, L.: Climatology and ENSO-related interannual variability of gravity waves in the Southern Hemisphere subtropical stratosphere revealed by high-resolution AIRS observations, J. Geophys. Res., 121, 7622–7640, https://doi.org/10.1002/2015JD024462, 2016.

Scaife, A. A., Butchart, N., Warner, C. D., and Swinbank, R.: Impact of a Spectral Gravity Wave Parameterization on the Stratosphere in the Met Office Unified Model, Journal of the Atmospheric Sciences, 59, 1473–1489, https://doi.org/10.1175/1520-0469(2002)059<1473:ioasgw>2.0.co;2, 2002.

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

Stockwell, R. G. and Lowe, R. P.: Airglow imaging of gravity waves 1. Results from a small network of OH nightglow scanning imagers, J. Geophys. Res., 106, 17 185–17 203, https://doi.org/10.1029/2001JD900035, 2001.

Stockwell, R. G., Mansinha, L., and Lowe, R. P.: Localization of the complex spectrum: the S transform, IEEE Transactions on Signal Processing, 44, 998–1001, https://doi.org/10.1109/78.492555, 1996.

Stockwell, R. G., Taylor, M. J., Nielsen, K., and Jarvis, M. J.: The evolution of a breaking mesospheric bore wave packet, J. Geophys. Res., 116, https://doi.org/10.1029/2010JD015321, 2011.

Thompson, D. W. J., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., and Karoly, D. J.: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change, Nature Geoscience, 4, 741–749, https://doi.org/10.1038/ngeo1296, 2011.

Vadas, S. L. and Becker, E.: Numerical Modeling of the Excitation, Propagation, and Dissipation of Primary and Secondary Gravity Waves during Wintertime at McMurdo Station in the Antarctic, Journal of Geophysical Research: Atmospheres, 123, 9326–9369, https://doi.org/10.1029/2017JD027974, 2018.

Vadas, S. L. and Fritts, D. C.: The importance of spatial variability in the generation of secondary gravity waves from local body forces, Geophysical Research Letters, 29, 45–1–45–4, https://doi.org/10.1029/2002GL015574, 2002.

Vadas, S. L., Zhao, J., Chu, X., and Becker, E.: The Excitation of Secondary Gravity Waves From Local Body Forces: Theory and Observation, Journal of Geophysical Research: Atmospheres, 123, 9296–9325, https://doi.org/10.1029/2017JD027970, 2018.

Vosper, S. B.: Mountain waves and wakes generated by South Georgia: implications for drag parametrization, QJRMS, 141, 2813–2827, https://doi.org/10.1002/qj.2566, 2015.

Vosper, S. B. and Ross, A. N.: Sampling Errors in Observed Gravity Wave Momentum Fluxes from Vertical and Tilted Profiles, Atmosphere, 11, 57, https://doi.org/10.3390/atmos11010057, 2020.

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

Vosper, S. B., van Niekerk, A., Elvidge, A., Sandu, I., and Beljaars, A.: What can we learn about orographic drag parametrisation from high-resolution models? A case study over the Rocky Mountains, Quarterly Journal of the Royal Meteorological Society, 146, 979–995, https://doi.org/10.1002/qj.3720, 2020.

Warner, C. D. and McIntyre, M. E.: On the Propagation and Dissipation of Gravity Wave Spectra through a Realistic Middle Atmosphere, Journal of the Atmospheric Sciences, 53, 3213–3235, https://doi.org/10.1175/1520-0469(1996)053<3213:OTPADO>2.0.CO;2, 1996.

Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., Takahashi, M., and Sato, K.: General aspects of a T213L256 middle atmosphere general circulation model, J. Geophys. Res., 113, D12 110, https://doi.org/10.1029/2008JD010026, 2008.

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., and Thuburn, J.: 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, 1505–1520, https://doi.org/10.1002/qj.2235, 2014.

Woods, B. K. and Smith, R. B.: Energy Flux and Wavelet Diagnostics of Secondary Mountain Waves, Journal of the Atmospheric Sciences, 67, 3721–3738, https://doi.org/10.1175/2009JAS3285.1, 2010.

Wright, C. J.: A one-year seasonal analysis of martian gravity waves using MCS data, Icarus, 219, 274–282, https://doi.org/10.1016/j.icarus.2012.03.004, 2012.

Wright, C. J. and Gille, J. C.: Detecting overlapping gravity waves using the S-Transform, Geophy. Res. Lett., 40, 1850–1855, https://doi.org/10.1002/grl.50378, 2013.

Wright, C. J. and Hindley, N. P.: How well do stratospheric reanalyses reproduce high-resolution satellite temperature measurements?, Atmos. Chem. Phys., 18, 13 703–13 731, https://doi.org/10.5194/acp-18-13703-2018, 2018.

Wright, C. J., Osprey, S. M., and Gille, J. C.: Global observations of gravity wave intermittency and its impact on the observed momentum flux morphology, J. Geophys. Res., 118, 10,980–10,993, https://doi.org/10.1002/jgrd.50869, 2013.

Wright, C. J., Hindley, N. P., and Mitchell, N. J.: Combining AIRS and MLS observations for three-dimensional gravity wave measurement, Geophy. Res. Lett., 43, 884–893, https://doi.org/10.1002/2015GL067233, 2015GL067233, 2016a.

Wright, C. J., Hindley, N. P., Moss, A. C., and Mitchell, N. J.: Multi-instrument gravity-wave measurements over Tierra del Fuego and the Drake Passage - Part 1: Potential energies and vertical wavelengths from AIRS, COSMIC, HIRDLS, MLS-Aura, SAAMER, SABER and radiosondes, Atmos. Chem. Phys. Disc., 8, 6797–6876, https://doi.org/10.5194/amtd-8-6797-2015, 2016b.

Wright, C. J., Hindley, N. P., Hoffmann, L., Alexander, M. J., and Mitchell, N. J.: 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, 8553–8575, https://doi.org/10.5194/acp-17-8553-2017, 2017.

Wu, D. L.: Mesoscale gravity wave variances from AMSU-A radiances, Geophy. Res. Lett., 31, 1944–8007, https://doi.org/10.1029/2004GL019562, l12114, 2004.

Yan, A., Zhou, W., Yuan, Q., Yuan, S., Wu, Q., Zhao, X., and Wang, J.: Automatic seizure detection using Stockwell transform and boosting algorithm for long-term EEG, Epilepsy & Behavior, 45, 8–14, https://doi.org/10.1016/j.yebeh.2015.02.012, 2015.