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 $\lambda_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.

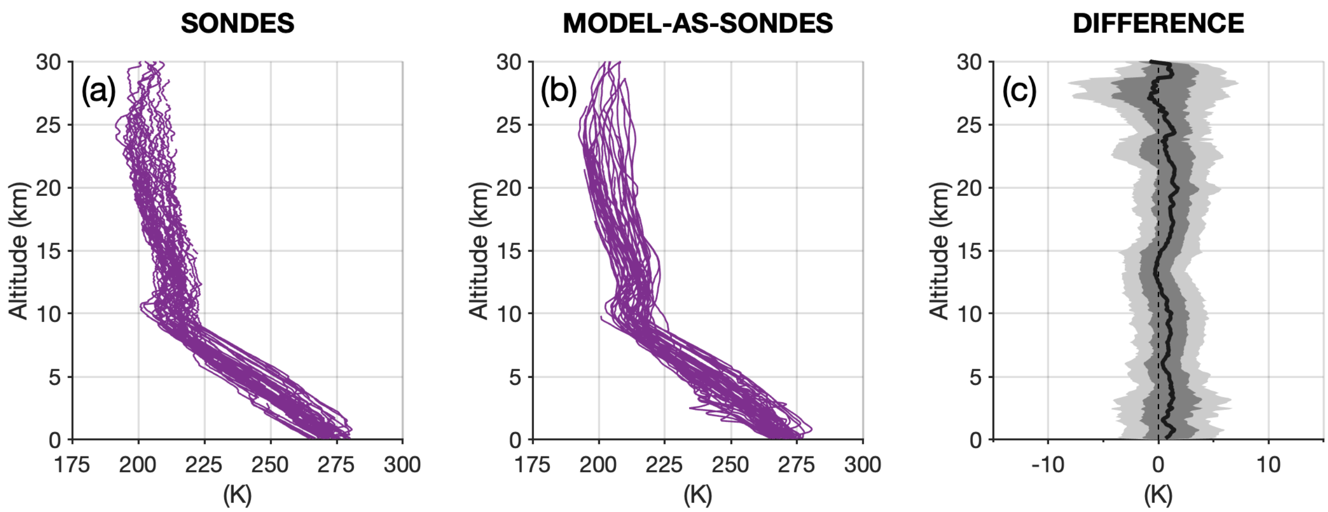

Figure R1: Temperature measurements in (a) the local-area model and (b) coincident radiosonde observations during June-July 2015. Temperature in the model are evaluated along the radiosonde flight path in 3 spatial dimensions and time. Panel (c) shows the difference (Sondes − Model) between them.

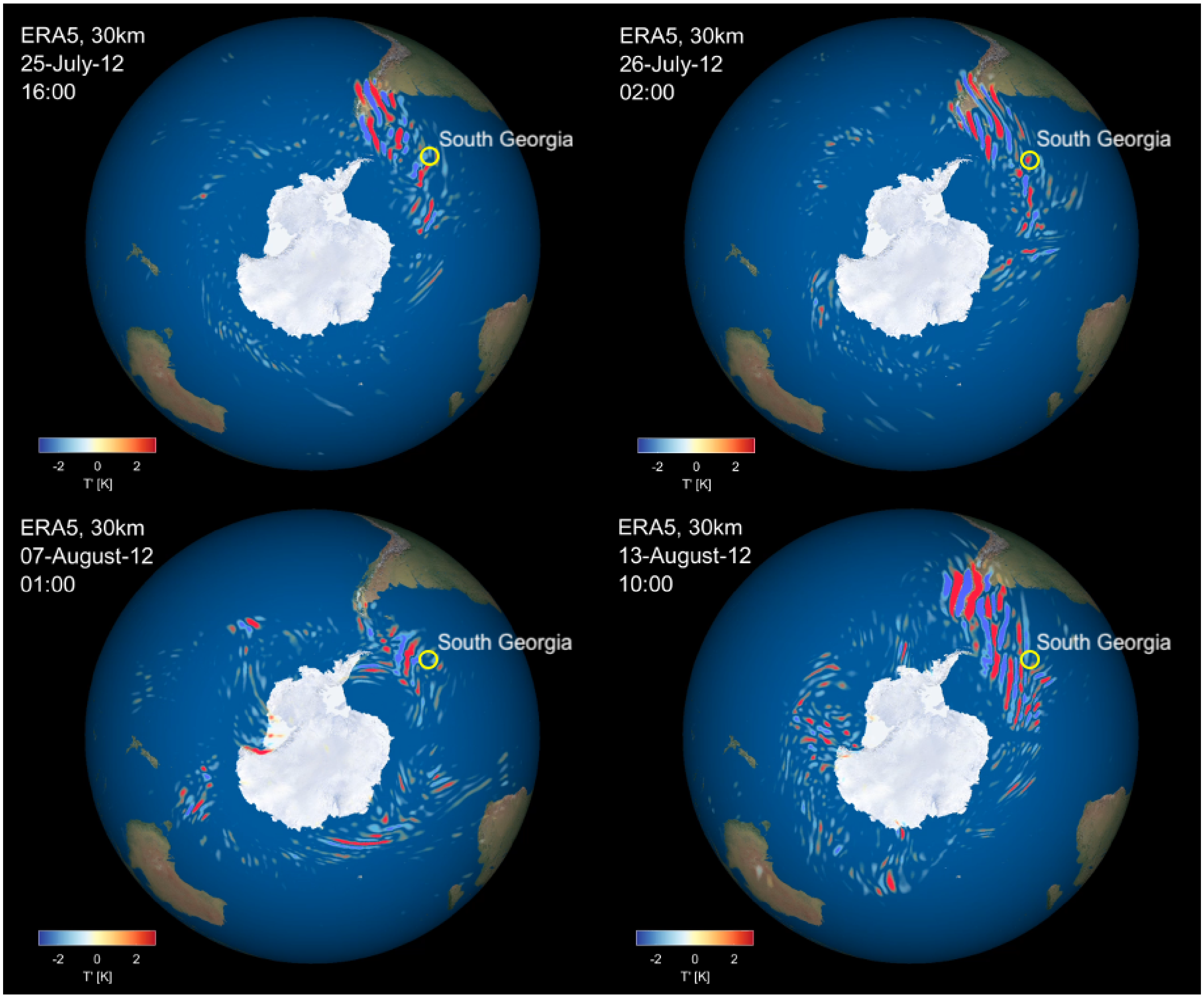

Figure R2: Selected snapshots from an animation of ERA5 reanalysis temperature perturbations over the southern hemisphere at 30 km altitude during July-August 2012.