Comparing gravity waves in a kilometre-scale run of the IFS to AIRS satellite observations and ERA5

Emily J. Lear[1], Corwin J. Wright[1], Neil P. Hindley[1], Inna Polichtchouk[2], Lars Hoffmann [3]

[1]Centre for Climate Adaptation and Environment Research, University of Bath, Bath, UK
[2]European Centre for Medium-Range Weather Forecasts, Reading, UK
[3]Jülich Supercomputing Center, Forschungszentrum Jülich, Jülich, Germany

Key Points:

- A kilometre-scale IFS run was resampled as AIRS to compare gravity wave properties using two different methods
- Gravity waves can be seen in the resampled IFS run and AIRS at similar times and locations
- Mean amplitudes in the resampled IFS run were found to be significantly lower than in the observations by a factor of ∼2.4

Corresponding author: Emily Lear, `ejl45@bath.ac.uk`

Abstract

Atmospheric gravity waves (GWs) impact the circulation and variability of the atmosphere. Sub-grid scale GWs, which are too small to be resolved, are parameterized in weather and climate models. However, some models are now available at resolutions at which these waves must be resolved and it is important to test whether these models do this correctly. In this study, a GW resolving run of the ECMWF (European Centre for Medium-Range Weather Forecasts) IFS (Integrated Forecasting System), run with a 1.4 km average grid spacing (TCo7999 resolution), was compared to observations from the Atmospheric Infrared Sounder (AIRS) instrument, on NASA's Aqua satellite, to test how well the model resolves these waves. In this analysis, nighttime data were used from the first 10 days of November 2018 over Asia and surrounding regions. The IFS run is resampled with AIRS's observational filter using two different methods for comparison. The ECMWF ERA5 reanalysis is also resampled as AIRS, to allow for comparison of how the high resolution IFS run resolves GWs compared to a lower resolution model that uses GW drag parametrizations. Wave properties are found in AIRS and the resampled models using a multi-dimensional S-Transform method. Orographic GWs can be seen in similar locations at similar times in all three data sets. However, wave amplitudes and momentum fluxes in the resampled IFS run were found to be significantly lower than in the observations. This could be a result of horizontal and vertical wavelengths in the IFS run being underestimated.

Plain Language Summary

Small-scale atmospheric waves known as gravity waves (GWs) transport energy and momentum and affect the dynamics of the atmosphere. At the resolution of the ECMWF IFS (Integrated Forecasting System) simulation, with an average grid spacing of 1.4 km (TCo7999 resolution), these waves need to be resolved. GWs in this IFS run were compared to those in observations from the AIRS (Atmospheric Infrared Sounder) instrument on NASA's Aqua satellite, for nighttime data over Asia and surrounding regions, during the first 10 days of November 2018, to test how well these waves are resolved in the model. The high resolution IFS run is resampled as AIRS to remove GWs outside of the wavelength ranges that can be seen in the observations, allowing the data sets to be compared. GWs with orographic sources (waves formed by wind flowing over topography), can be seen at similar times and in similar locations in the IFS run and observations, but wave amplitudes in the resampled IFS run were found to be significantly lower.

1 Introduction

Atmospheric gravity waves (GWs) are small-scale waves which transport energy and momentum throughout the atmosphere (M. J. Alexander et al., 2010; Fritts & Alexander, 2003). These waves have both direct and indirect effects on the atmosphere: to take just a few examples, GWs act as a major cause of clear-air turbulence affecting aircraft (Lane et al., 2009), contribute to ozone depletion in the polar stratosphere (Carslaw et al., 1998), affect the formation of sudden stratospheric warmings in winter by preconditioning the polar vortex (Albers & Birner, 2014) and affect the timing of the polar vortex breakdown in spring (Polichtchouk et al., 2018). GWs in the stratosphere have also been shown to impact the Brewer-Dobson Circulation (e.g. Sato & Hirano, 2019). Sources of GWs include orographic sources (wind flowing over topography) and non-orographic sources, such as convection and wind shear (Fritts & Alexander, 2003; M. J. Alexander et al., 2010).

Despite their importance for achieving realistic atmospheric circulations, GWs and their impacts remain notoriously difficult to represent in numerical models. One reason for this is because large portions of the GW spectrum occur at scales below the grid size of the model, and are therefore unresolved. Instead, the acceleration (or deceleration) of the background flow at different altitudes due to GW propagation and breaking is represented by parameterizations, which can be tuned to correct for the unknown momentum forcing due to GWs not resolved by the model. However, these parameterizations are poorly constrained by observations and contain simplifying assumptions that can lead to major circulation biases (Butchart et al., 2011; Harvey et al., 2019). Due to computational constraints, this reliance on GW parameterizations is still widespread in the vast majority of operational models used for numerical weather prediction (NWP), atmospheric research and long-term dynamical climate simulations (M. J. Alexander et al., 2010; Plougonven et al., 2020).

In recent decades, ever increasing computational power has allowed models to be developed with sufficient spatial resolution to resolve ever larger portions of the GW spectrum. In some of these specialist offline configurations, the resolution is sufficiently high that the effects of resolved waves alone are sufficient to achieve realistic circulations in the middle atmosphere, and GW parameterizations are no longer required (Sato et al., 2012; Vosper, 2015; Watanabe & Miyahara, 2009; Lund et al., 2020; Wedi et al., 2020). While these simulations are still prohibitively expensive for operational use, it is likely that this trend will continue and models will be able to resolve an increasingly large portion of the GW spectrum. This then raises a question: how realistic are the resolved waves in these high resolution simulations compared to observations?

Here this question is investigated for one such model: a high-resolution "kilometre-scale" configuration of the Integrated Forecasting System (IFS) model developed by the European Centre for Medium Range Weather Forecasts (ECMWF), as described by Wedi et al. (2020). This configuration was run at TCo7999 resolution (Wedi et al., 2020), which is equivalent to an average horizontal grid spacing of around 1.4 km globally, and no GW parameterizations were used. In this study, the amplitudes, wavelengths and momentum fluxes of resolved GWs in the model stratosphere are compared to 3-D satellite observations from the Atmospheric Infrared Sounder (AIRS) instrument using the retrieval of Hoffmann and Alexander (2009). For further comparison, we also investigate resolved GWs in the ERA5 reanalysis, also produced by ECMWF, which is generated using a 9 km version of the IFS, to understand the impact of the increased resolution of the km-scale IFS on resolved GW properties compared to operational configurations.

In Kruse et al. (2022), four numerical weather prediction models, including the IFS run with an average grid spacing of $\sim$9 km, were compared to AIRS data, which showed that the models reproduced mountain waves in the observations well, near to the Drake Passage, but wave amplitudes were lower than those observed. GW momentum fluxes in the ECMWF operational analyses, produced using the IFS and 4D variational data assimilation, at a resolution of 0.125° in longitude and latitude (approximately 16 km) with 91 model levels, were found to be a factor of 5 lower than in Concordiasi balloon observations (Jewtoukoff et al., 2015). The ECMWF operational analyses was also found to have lower wave amplitudes compared to AIRS observations by a factor of 2–3 (Hoffmann et al., 2017), using data from 2003 to 2012. In this time period, the effective horizontal resolution and number of model levels was increased from 39 to 16 km and 60 to 91, respectively. Okui et al. (2023), found generally good agreement between the amplitudes and momentum fluxes of GWs in the Japanese Atmospheric GCM for Upper Atmosphere Research (JAGUAR) (Watanabe & Miyahara, 2009), a high vertical resolution model which uses no GW drag parameterizations, and AIRS observations.

However, making comparisons between observed and simulated GWs is not straightforward. This is because no instrument (or model) can observe the full GW spectrum. The sampling and resolution characteristics of a particular observing instrument (such as AIRS as used here) limit the range of observable GW horizontal and vertical wavelengths, a phenomenon known as the "observational filter" of the instrument (Preusse et al., 2002; M. J. Alexander & Barnet, 2007). Likewise, the spatial resolution of a model limits its ability to simulate all GW wavelengths. Therefore, to make a fair comparison between observations of GWs and resolved GWs in a model, we must first sample the model as if it were observed by the instrument by applying the instrument's sampling pattern and horizontal and vertical resolutions to the model output fields (Wright & Hindley, 2018; Hindley et al., 2021). This model-sampled-as-observations dataset can then be analysed in exactly the same way as the observations and a fair comparison between the measured GW properties can be made. The approach taken to perform this sampling method however can vary between studies, so here we investigate two different sampling methods to create this dataset: one using a simplified approach described by Hindley et al. (2021) and the second using the more rigorous, but more computationally-expensive, approach of Wright and Hindley (2018).

The selected km-scale configuration of the IFS was run globally for the period of November 2018. During this time, significant stratospheric GW activity was observed in the model, AIRS observations and the ERA5 reanalysis over continental Asia and the surrounding regions, so we select this region over which to perform our comparison (see Fig. 1). The region is likely to contain numerous sources of orographic GW activity generated by surface flow over mountain ranges, such as the Abakanski Khrebet Mountain range, the Ural mountains, the Pamir mountains and other hotspots as observed by Hoffmann et al. (2013) and Hindley et al. (2020). GWs in this region have previously been shown to be strongly visible in AIRS (Hindley et al., 2020) and aircraft (Wright & Banyard, 2020) observations, but not in limb sounder observations (Geller et al., 2013; Ern et al., 2018), suggesting a strong role for long-vertical-short-horizontal-wavelength GWs of the type this model should be well-configured to accurately resolve. The region is also likely to contain non-orographic GW activity from jets, fronts and spontaneous geostrophic adjustment processes around the edge of the wintertime stratospheric polar vortex. This region and time period therefore presents an ideal opportunity to investigate the realism of resolved GWs in the high resolution IFS simulation compared to observations and to the lower resolution reanalyses.

The data sets used in this study are described in Section 2. In Section 3, the methods for resampling the models as AIRS and calculating the GW properties are described. The results of the comparison between the resampled models and AIRS observations are presented in Section 4. These results are discussed in Section 5, and the summary and conclusions are presented in Section 6.

2 Data

2.1 AIRS

Stratospheric temperature data were used from the Atmospheric Infrared Sounder (AIRS) instrument on NASA's Aqua satellite (Hoffmann & Alexander, 2009). The Aqua satellite's orbit is sun-synchronous and near-polar, with a period of 98.8 minutes. This allows AIRS to obtain data with near global coverage over a day. AIRS has 2378 channels which measure infrared radiation in the wavelength range of 3.7–15.4 µm and 4 channels that measure near-infrared and visible radiation with a range of 0.4–0.94 µm (Parkinson, 2003). AIRS scans from +49.5° to -49.5° across track, with 90 elements and a swath width of $\sim$ 1780 km and has a horizontal resolution of $\sim$ 13.5 km $\times$ 13.5 km at nadir which reduces to 41 km $\times$ 21.4 km at the track edge (Chahine et al., 2006). The data are stored in granules containing 6 minutes of data, with 240 granules for each day (Aumann et al., 2003).

The 3D temperature data used in this study is calculated from AIRS radiance measurements using the retrieval scheme described by Hoffmann and Alexander (2009). This retrieval has an improved horizontal resolution by a factor of 3, in comparison with AIRS operational data, in both the along- and across-track directions, allowing more GW features to be seen in the data. The retrieval uses 4 µm and 15 µm AIRS CO_2 emission channels for nighttime, but only the 15 µm channels are used for the daytime retrievals; specifically 12 15 µm channels are used in the retrieval for daytime and nighttime and 23 4 µm channels for nighttime. In daytime, the radiance measurements for the 4 µm channels are affected by non-LTE (local thermodynamic equilibrium) effects due to solar excitation, so these channels are not used. In the middle and upper stratosphere, few of the 15 µm channels are sensitive to temperature perturbations and therefore, GWs, compared to the 4 µm channels. The estimated total retrieval error of the temperature measurements is 1.6–3.0 K for altitudes from 20 to 60 km. The retrieved temperatures have a vertical resolution of $\sim$ 7–15 km (Hoffmann & Alexander, 2009). Figure 2a–c of Hindley et al. (2019) show estimated AIRS temperature retrieval errors due to noise and vertical resolution with altitude. An altitude range of 27–54 km, was chosen for the AIRS retrieval data used in this study, because nighttime data have relatively low noise and high resolution in this range.

2.2 High resolution IFS simulation (TCo7999)

The high resolution run of the ECMWF IFS used in this study is is a global, hydrostatic simulation, based on version CY45Rl of the IFS atmospheric model (ECMWF, 2023), and run at a TCo7999 resolution (Wedi et al., 2020; Polichtchouk et al., 2022). This resolution has a horizontal grid spacing of 1.25 km at the equator, with an average of 1.4 km globally. In this paper, the simulation is referred to as the 1 km IFS run. ECMWF's operational 10 day forecasts, at the time of writing, use the IFS at a resolution of 9 km with deep convection parameterization.

The CY45R1 version of the IFS has 137 model levels, at heights from 0.01 hPa down to the surface and the spacing between the levels increases with altitude (Wedi et al., 2020). The smallest GWs are likely to be strongly damped by numerical diffusion in the IFS. To prevent wave reflection at the top of the model, the IFS has a weak sponge layer from 10 hPa to the model top, which only has a small effect on resolved waves, and a very strong sponge layer above 1 hPa (Polichtchouk et al., 2023). The contribution of the GW drag parameterizations is designed to reduce as the horizontal resolution of the model is increased, and is zero at an average grid spacing of 1.4 km. At this resolution, the simulation did not use deep convection parameterizations. The 1 km IFS simulation was initialised on 1$^{\text{st}}$ November 2018 00:00 UTC, integrated for 4 months, and ran with a time step of 60 s and a model output frequency of 3 hours. The temperature structure and background flow of the 1 km IFS remain similar to IFS simulations run for the same time period at 3.9 km and 7.8 km horizontal resolutions during the the first 15 days of the simulation (Polichtchouk et al., 2022). Polichtchouk et al. (2022, 2023) investigated the effect of the increase in horizontal resolution from $\sim$9 to $\sim$1 km and the deep convection parameterization and found GWs are still under-resolved at a grid spacing of $\sim$9 km, compared to GWs at the $\sim$1 km resolution.

In this study, 3 hourly 1 km IFS temperature data were interpolated onto a regular longitude-latitude grid, with a resolution of $0.1^\circ \times 0.1^\circ$. This reduced resolution was chosen to make the data easier to use and should not affect the results, as this is still a significantly higher horizontal resolution than the AIRS retrieval.

2.3 ERA5

The ECMWF ERA5 is a 5$^{\text{th}}$ generation global reanalysis, run from 1940 to the present (Copernicus Climate Change Service, 2023). ERA5 uses 4D-Var (4D variational) data assimilation which combines observations, including AIRS data, and hindcasts (past weather forecasts). The observations and hindcasts are combined in space and time within 12 hour assimilation windows (ECMWF, 2021). The hindcasts used in the data assimilation are from the ECMWF IFS CY41R2 (ECMWF, 2023), implemented in 2016, at TCo1279 resolution (9 km average horizontal grid spacing globally). ERA5 has the same model levels and sponge layers as in the CY45R1 version of the IFS (ECMWF, 2021, 2020). The ERA5 temperature data used were downloaded from the Copernicus Climate Data Store, for every 3 hours during the time period investigated, on a regular latitude-longitude grid with a resolution of 0.25° (Copernicus Climate Change Service, 2023).

3 Methods

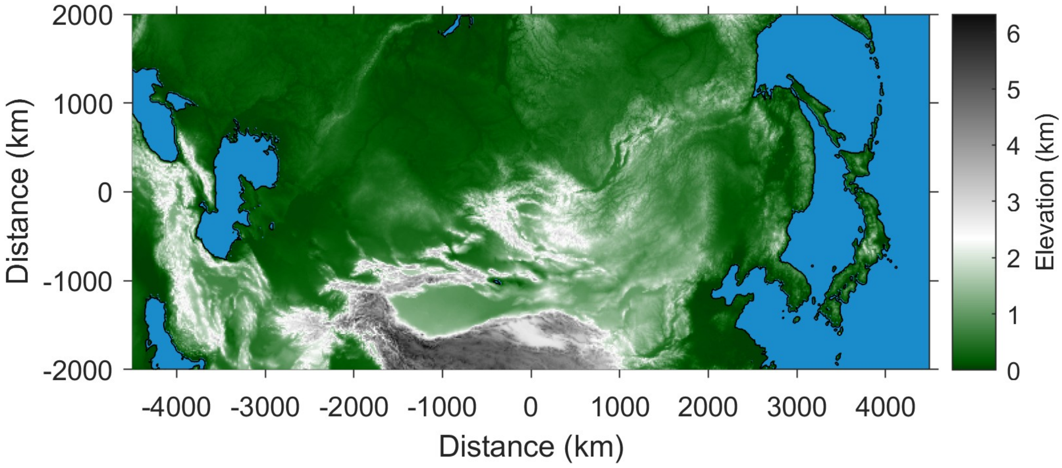

Figure 1. Map of topography in the region investigated on a regular distance grid centred at 52° latitude, 94° longitude. Coastlines are shown in black.

Data from the first 10 days of November 2018 for AIRS, the 1 km IFS run and the ERA5 reanalysis were used, as the 1 km IFS run was initialized on the 1st of this month at 00:00 UTC. This time period was chosen, because the background temperature and wind structure, which affect the generation and propagation of GWs, of the 1 km IFS (Polichtchouk et al., 2022), is expected to remain similar to observations in this period. This assumption is investigated in section 3.1.3.

Where higher magnitude temperature perturbations, indicating GWs, are present, the data are expected to have a greater variance in general. During this time period, the AIRS granules with the highest variances were located in Asia and surrounding areas, suggesting stronger GW activity. Hence, this study focuses on data from this region (shown in Figure 1). Variances of AIRS temperature perturbations were also used in Hoffmann et al. (2013) to identify individual GW events.

In this study all results use data at 39 km altitude in AIRS and the resampled models as this is at the centre of the AIRS usable height range (see Figure 2 of Hindley et al. (2019)). This is also in the altitude range where the AIRS retrieval vertical resolution is greater and the noise is lower for nighttime data in polar winter, mid latitudes and the tropics. The results are presented only for nighttime data, due to the the lower vertical resolution and higher retrieval error of the daytime AIRS retrieval.

3.1 Resampling Methods

In this study, the observational filter of the AIRS retrieval was applied to the 1 km IFS run and ERA5, to remove GWs outside of the horizontal and vertical wavelength ranges in which these waves can be seen in the observations. Data were found from the

1 km IFS run and ERA5 at the closest time of the 3 hourly 1 km IFS data used to the measurement time of each AIRS granule and were resampled as that granule. The models were not interpolated to the AIRS measurement times as this would smooth out small scale structures such as GWs (Wright & Hindley, 2018). Two different methods were used to resample the 1 km IFS run as AIRS. The first resampling method was run on a desktop computer, whereas the second is more computationally expensive (Wright & Hindley, 2018) and required the use of HPC.

3.1.1 Method 1

The first method, referred to as method 1 in this paper and applied to both the 1 km IFS run and ERA5, is described by Hindley et al. (2021). The 1 km IFS data, which was previously interpolated onto a regular longitude-latitude grid with a spacing of 0.1°, were selected at the closest 3 hourly time to each AIRS granule. This was first interpolated onto a regular distance grid in the horizontal with a point spacing of 2.7 km, which is a higher resolution than any part of the original 0.1° longitude-latitude grid spacing in the region investigated. The data were then smoothed to the approximate horizontal resolution of AIRS at track-centre, using a Gaussian with a FWHM (full width at half maximum) of 13.5 km × 13.5 km. Following this, the data were interpolated onto the location of the AIRS granule. The data was then interpolated to a regular distance spacing in the vertical of 0.1 km from 26 to 55 km altitude, so it could be smoothed to the vertical resolution of the AIRS retrieval. As the vertical resolution of the retrieval varies with altitude, the whole volume of data were smoothed in the vertical using a Gaussian function with a different FHWM for each AIRS altitude, from 27 to 54 km, with a 3 km point spacing. Different arrays of values for the FWHM at each altitude were used (shown in Figure 2, of Hindley et al. (2019)) for each of the granules containing mostly nighttime data, depending on whether they are located mostly in the tropics, midlatitudes or polar region. The nearest horizontal levels to each altitude were then found and stored in a separate array, which is the model data resampled as AIRS.

ERA5 is also resampled using this method, but since the data has a lower horizontal resolution than AIRS, it is not interpolated to a regular distance grid and smoothed to the horizontal resolution of AIRS before it is interpolated to the AIRS granule location. The 1 km IFS run and ERA5 resampled using this method are referred to as IFS 1 and ERA5 1 in this paper.

3.1.2 Method 2

The second method used to resample the 1 km IFS run as AIRS is described by Wright and Hindley (2018) and referred to as method 2 in this paper. This involves oversampling the model data, onto a grid with a spacing of 1 km in the along and across track directions and 1/20 of a decade of pressure in the vertical. These values were selected based on sensitivity testing discussed in Appendix B of Wright and Hindley (2018). Each oversampled point was then weighted by the estimated instrument sensitivity at each point and summed to produce a sample corresponding to each AIRS measurement. This aims to improve the accuracy in comparison with interpolating the model to the centre of the satellite measurement volume. Compared to interpolating to a single point, Wright and Hindley (2018) showed that this method lead to improvements in brightness temperature measurements in AIRS Level 1 data which are significant for small-scale temperature perturbations caused by GWs. The 1 km IFS data resampled as AIRS using this method is referred to as IFS 2.

3.1.3 Temperature divergence of the resampled models and AIRS

Figure 2 shows the point-wise correlation (Figure 2a) and RMSD (Figure 2b) between the temperature at 39 km altitude in the AIRS observations and resampled mod-

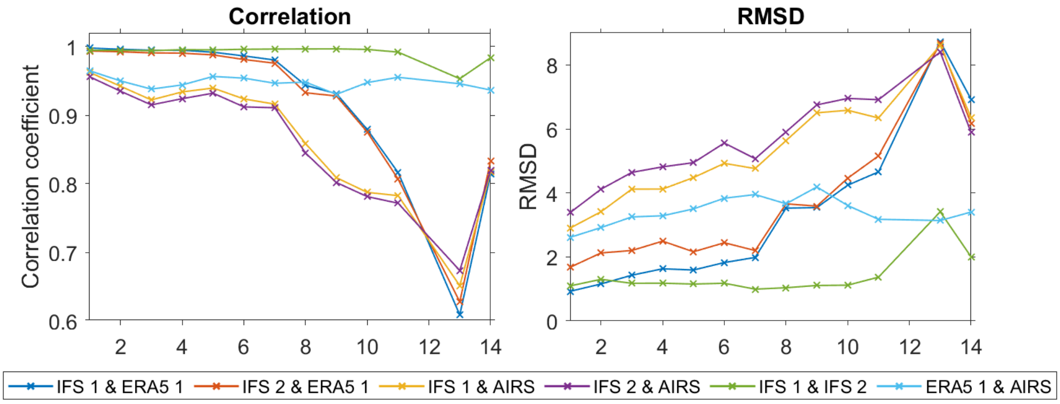

Figure 2. Point-wise correlation coefficients and RMSD between the temperature at 39 km altitude in AIRS, and the 1 km IFS run and ERA5 resampled as AIRS, for each night during the first 14 days of November 2018, within the region shown in Figure 1.

els. These are plotted for each night during the first 14 days of November 2018. The data for the 12th night is missing, because AIRS data were not recorded for most of the region studied during this night.

The point-wise correlation (Figure 2a) between IFS 1 and 2 and AIRS decreases over time from when the model was initialised up to the 11th night, which is expected since the free-running 1 km IFS run diverges from the 'truth'. On night 13, the correlation is lower due to a single large anomalous wave covering a large fraction of the region investigated in the raw model data. The correlation then increases on the 14th night. As ERA5 assimilates data from observations, including AIRS, the resampled ERA5 data does not have a decreasing correlation with the AIRS retrieval. Since IFS 1 and 2 are the same data resampled as AIRS using different methods, the correlation coefficient between these data sets remains very high, but is lowest on the 13th night.

The RMSD (Figure 2b) is greatest between IFS 1 and 2 and AIRS and increases up to night 11. The RMSD also increases between IFS 1 and 2 and ERA5 1 up to night 11. On the 13th night there is a peak in the RMSD between the 1 km IFS run resampled using both methods and the other data sets, and between IFS 1 and 2 as a result of the large anomalous wave in the 1 km IFS run data for this night. Due to the correlations and RMSD's shown in Figure 2, data is only used from the 1st – 10th November 2018 for the results presented in Section 4.

3.2 Regridding the data to a regular distance grid and finding temperature perturbations

The AIRS granules and model data resampled as each AIRS granule, with data in the region investigated were regridded onto regular 3D distance grids, as this is required for 3D spectral analysis. The grids have a horizontal point spacing of ∼20 km in the across track direction and ∼18 km in the along track direction, so that the swath width and number of across-track and along-track points remain the same after the data has been regridded, and a vertical spacing of 3 km. Following this, the background is removed from AIRS and the resampled models using a 4th order polynomial fit in the cross track direction (Wu, 2004; M. J. Alexander & Barnet, 2007).

Separately, temperature perturbations were also found for the 1 km IFS run and ERA5 before resampling as AIRS to allow for comparison to the resampled models and observations (see Section 4.3). The 1 km IFS run and ERA5 were interpolated to a regular distance grids with a point spacing of 1 km in the vertical and a horizontal point spacing of 15 km for the 1 km IFS run, and 30 km for ERA5. The background was found by smoothing both data sets using a Gaussian filter with a convolution kernel size of 11 $\times$ 11 points and a standard deviation of 7.15 points. This was then subtracted from the temperature data to find the perturbations.

3.3 Adding AIRS retrieval noise to the resampled models

Since temperature perturbations in the AIRS retrieval data can not be separated from the noise, AIRS noise is added to the temperature perturbations found for the resampled models (ERA5 1, IFS 1 and IFS 2), so that only waves that can still be seen with AIRS noise added are compared. To find granules containing only noise, the granules were sorted from lowest to highest variance of the temperature perturbations, and checked in this order to find granules without subjectively clear visible waves. In Hindley et al. (2021), temperature perturbations were found for an AIRS overpass containing 2 granules with no waves, which were then randomised at each altitude, and added to the model resampled as AIRS. This method was also used in Okui et al. (2023) for one AIRS granule. However, the noise added in Hindley et al. (2021) is uncorrelated pixel-scale noise, meaning that any noise structures larger than around 30–50 km in the AIRS retrieval data would not be included. Okui et al. (2023) found that adding noise using this method resulted in a lower background amplitude due to noise in the resampled model than in AIRS and may not be suitable for adding AIRS retrieval noise at global scales. Therefore, in this study a different method was used.

30 nighttime AIRS granules containing only noise were selected in total, with 10 in the tropics, mid-latitudes and polar region, respectively. Granules were chosen with over 10% of the data points in the region investigated. For the tropics and mid-latitudes, the granules were selected in the first 14 days of November in 2018. As there were not enough nighttime polar AIRS granules containing only noise in the first two weeks of November 2018, granules were chosen during the first 2 weeks of November in years from 2016 to 2020 for the polar region. The temperature perturbations of the AIRS noise granules (shown in Supplementary Figures S1–S3) were found using the method described above. The data in the arrays of temperature perturbations were reversed in the along and across-track directions separately and saved so that there were 30 noise granules in total for each group, to increase the number of granules that could be selected. For each granule of model data resampled as AIRS, an array of noise temperature perturbations was chosen randomly from the corresponding group and added to the resampled model temperature perturbations, depending on if the granule contains nighttime data which is mostly from the tropics ($<30°$), mid-latitudes ($30°$–$60°$ latitude) or polar region ($>60°$ latitude).

3.4 2D+1 S-Transform

The S-Transform (ST) is commonly used for the analysis of GWs (e.g. Fritts et al., 1998; M. J. Alexander et al., 2008). The 2D+1 ST is based on the 2D S-Transform (Hindley et al., 2016) and the 3D S-Transform (Wright et al., 2017). The 2D ST and 3D ST are extensions of the 1D S-Transform (Stockwell et al., 1996).

The 2D+1 ST calculates wavelengths using phase shifts between spectral features, which allows it to measure waves more effectively for 3D data with low resolution in one dimension compared with the variations in the wave field. Nadir-sensing instruments, such as AIRS have high horizontal resolution but low vertical resolution. This means there are a low number of vertical points, for the data from these instruments, in the stratosphere in comparison with the point numbers in the horizontal, limiting estimates of the vertical wavelengths of GWs. For the 2D+1 ST, 2D S-Transforms are found for the data levels in the horizontal and the phase differences between them are calculated (Wright et al., 2021).

Using the 2D+1 ST, vertical wavelengths can be calculated more precisely in comparison with the 3D ST, and this method is therefore more effective for measuring waves with long vertical wavelengths. A further improvement compared to the 3D ST is that the 2D+1 ST does not quantize vertical wavelengths to Fourier modes, so these wavelengths vary smoothly in the output. However, the 2D+1 ST is computationally slower.

The 2D+1 ST was used to find wave properties for the resampled 1 km IFS run and ERA5 reanalysis, both models before being resampled, and each AIRS granule. The wave amplitude is an output of the 2D+1 ST. The horizontal and vertical wavelengths were calculated using the granule-relative wave frequencies from the 2D+1 ST. The zonal M_x and meridional M_y components of the momentum flux were calculated using the following equation derived in Ern et al. (2004),

$$M_x, M_y = -\frac{\rho}{2}\left(\frac{k}{m}, \frac{l}{m}\right)\left(\frac{g}{N_B}\right)^2\left(\frac{|T'|}{\overline{T}}\right)^2 \tag{1}$$

where ρ is the atmospheric density, k, l and m are the wavenumbers in the zonal, meridional and vertical directions respectively, g is the acceleration due to gravity and N_B is the buoyancy frequency, set 0.02 s^{-1} in this study. $\overline{T}$ is the local background temperature and $|T'|$ is the amplitude. The wavenumbers are signed to preserve the sign (direction) of the zonal and meridional momentum flux components (P. Alexander et al., 2018).

The altitude range selected of the model data resampled as AIRS and of the AIRS data used for the 2D+1 ST analysis was 27–54 km. Including altitudes outside of this range with higher noise would affect the wave properties calculated using the 2D+1 ST. The 2D+1 ST was tuned to select waves with horizontal wavelengths ranging from 60 to 800 km. Areas where the vertical wavelength is below 6 km or above 45 km were removed from the data for the wave properties calculated.

4 Results

4.1 Example Case Studies

Two case studies comparing the resampled models and AIRS are presented in this section. The first case study includes data from AIRS granules with mean times of 19:38–19:50 UTC on the 5$^{\text{th}}$ November and the resampled model granules at the closest 3 hourly times to the observations. The temperature perturbations for this example are shown in Figure 3. The top row shows the AIRS swath (Figure 3a) and resampled models with AIRS noise added (Figure 3b–d). The topography of the area is shown in Figure 3e as well as the resampled model swaths before adding noise (Figure 3f–h). 8 µm AIRS brightness temperatures and ERA5 winds are shown for both case studies in Supplementary Figures S4 and 5.

The large wave on the right seen in AIRS and the resampled models (Figure 3a–d) is likely to be orographic, as it is close to a region of higher topography. A curved wave can be seen on the left in AIRS (Figure 3a), which could be convective as it is located close to a region of with a brightness temperature lower than 220 K, indicating deep convection (Hoffmann & Alexander, 2010) (see Supplementary Figure S4a). This wave cannot be clearly seen in the resampled IFS (IFS 1 & 2, Figure 3b and c), which suggests that the convective source is missing or significantly reduced in strength in the 1 km IFS run. In ERA5, the wave can be seen but at a lower amplitude, suggesting the convective source was correctly assimilated in the reanalysis.

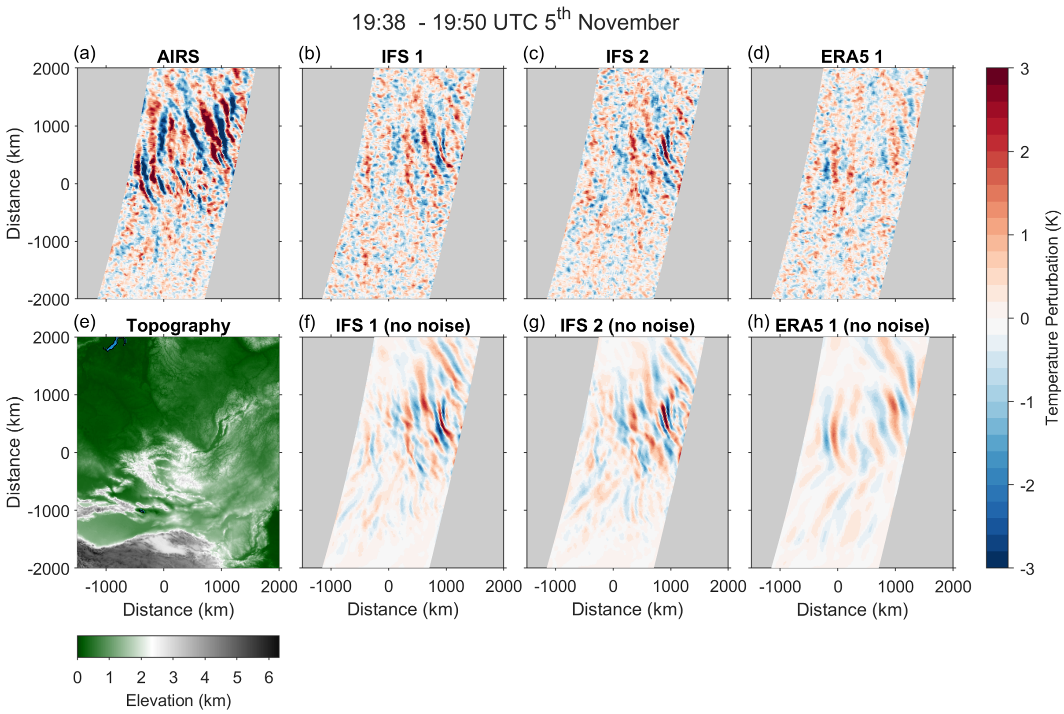

Figure 3. Temperature perturbations for granules in the region shown in Figure 1 with mean times of 19:38–19:50 UTC for AIRS (a) and at the closest time for the resampled models. Panels (b–d) show the resampled models after adding AIRS noise, with the resampled models before adding noise below them (no noise, panels f–h). The section of the topography from Figure 1, for the area is shown in panel (e), with (0,0) at 52° latitude, 94° longitude.

Figure 4 shows the wave properties, derived from the 2D+1 S-Transform, for the AIRS and resampled model granules shown in Figure 3a–d. Areas of higher amplitude in Figure 4a–d are seen in similar locations in each data set, but are higher in AIRS (Figure 4a) and lowest in ERA5 1 (Figure 4d). These regions can also be clearly seen for larger areas in AIRS than in the resampled models. In ERA5 1, an area of higher amplitude can be seen at the left of the granule stripe, near the centre of the y axis which is also seen in AIRS, but is not as clearly seen in IFS 1 and 2 (Figure 4b and c). Longer horizontal and vertical wavelengths can be seen in IFS 2 (Figure 4g and k), than in IFS 1 (Figure 4f and j), and in areas of noise the horizontal wavelengths are lower for all the data sets (Figure 4e–h). The magnitudes of the zonal and meridional momentum fluxes are highest in AIRS (Figure 4m and q), and lowest in ERA5 1 (Figure 4p and t).

The second case study is shown for AIRS granules with mean times of 19:14–19:26 UTC on the 9$^{\text{th}}$ November 2018, further in time from when the 1 km IFS run was initialised, and the resampled model granules at the closest 3 hourly times. Temperature perturbations for this case study are shown in Figure 5. As in Figure 3, the resampled model temperature perturbations are shown with AIRS noise (Figure 5b–d) and before adding AIRS noise (Figure 5f–h). The GWs shown in Figure 5a–d are likely to have orographic sources as they are close to regions of higher topography and not located near to a region of deep convection (shown in Supplementary Figure S4b).

In Figure 6, the wave properties are shown for the second case study (the AIRS swath, and models resampled as the swath in Figure 5a–d). The wave properties found for the resampled 1 km IFS run appear to agree less well with the AIRS observations, compared to the first case study in Figure 4. By the 9$^{\text{th}}$ November, the 1 km IFS run would have diverged further from reality compared to the first case study on the 5$^{\text{th}}$ November, which is closer to the time the simulation was initialised. The amplitudes in ERA5 1 (Figure 6d) are higher than in the example shown in Figure 4d. In the areas where the amplitude is higher in IFS 1 and 2 (Figure 6b and c), the horizontal wavelengths (Figure 6f and g) are longer than in AIRS (Figure 6e) for the same locations. In this case study, the horizontal wavelengths in the area with greater amplitudes in ERA5 1 (Figure 6h) appear to be more similar to the AIRS observations (Figure 6e) than IFS 1 and 2 (Figure 6f and g). This is also expected as ERA5 assimilates observations, unlike the 1 km IFS run.

The vertical wavelengths in IFS 1 and 2 (Figure 6j and k) are shorter than in AIRS (Figure 6i) for the areas where GWs can be seen. In the areas with higher amplitude in ERA5 1 (Figure 6d), the vertical wavelengths are generally longer (Figure 6l) than in the same locations for AIRS (Figure 6i). The magnitudes of the zonal and meridional momentum fluxes are also highest in AIRS for this case study (Figure 6m and q), due to the higher wave amplitudes, but they are lower in the resampled 1 km IFS run (IFS 1 and 2 in Figure 6n, o, r and s) than in ERA5 1 (Figure 6p and t). In both case studies, (Figures 4m–t and 6m–t) the zonal and meridional momentum flux is negative in areas where the amplitude is highest for all the data sets.

4.2 Time mean maps

Figure 7a–d shows the mean nighttime amplitudes during the first 10 days of November 2018 for the region in Figure 1. AIRS amplitudes (Figure 7a) are divided by a factor of 2 before plotting, so that areas with higher amplitudes in the resampled models can be seen more clearly using the same color bar. Areas of higher amplitude are seen in similar locations in AIRS and the resampled models. However, the amplitudes are significantly higher in AIRS than in the resampled 1 km IFS run (Figure 7c and d) and ERA5 1 (Figure 7b) and are lower in ERA5 1 than in IFS 1 and 2. The regions of higher wave amplitude are located near to mountain ranges, which can be seen in Figure 1 as areas of higher elevation, suggesting that the GWs have orographic sources. There is an area of higher amplitude over the Urals in Russia which can be seen in AIRS and the resampled IFS (1 and 2), but is not clearly seen in ERA5 1. The locations of the peaks in amplitude in AIRS and the resampled models are consistent with Hindley et al. (2020) and Wright and Banyard (2020). The maximum mean amplitudes in IFS 1 (Figure 7c) and IFS 2 (Figure 7d) are a factor of 2.5 and 2.2 lower than in AIRS respectively, so the maximum mean amplitude for the 1 km IFS run is a factor of 2.4 lower than in AIRS, averaging the results from the two resampling methods. The maximum amplitude in ERA5 (Figure 7b) is a factor of ~ 2.8 lower than in AIRS.

The mean nighttime zonal and meridional momentum fluxes are shown in Figure 7e–h and 7i–l respectively. Like the amplitude, the magnitude of the mean zonal and meridional momentum flux is significantly higher in AIRS (Figure 7e and i) than in the resampled models (Figure 7f–h and j–l). The zonal momentum flux generally has a higher magnitude than the meridional momentum flux, which is expected based on the AIRS retrieval climatology in Hindley et al. (2020) and due to background wind filtering. In areas where the amplitude is higher, the zonal momentum flux is negative (westward) and the meridional momentum flux is also negative (southward). This suggests that the highest amplitude GWs are formed by wind flowing over the northeast–southwest aligned topography shown in Figure 1.

The mean horizontal wavelengths for nighttime data in the first 10 days of November 2018 are also shown in Supplementary Figure S6.

4.3 Kernel distribution functions

Kernel distribution functions (KDFs) for the amplitudes, horizontal and vertical wavelengths and momentum flux are shown for nighttime data during the first 10 days of November 2018 in Figure 8. In the first column (Figure 8a, c, e and g), the KDFs for AIRS, and ERA5 and the 1 km IFS run (IFS) before resampling as AIRS are shown and the second column (Figure 8b, d, f and h) shows the KDFs for AIRS and the resampled models. KDFs were chosen to show the distributions of the data rather than probability density functions, due to the noise in AIRS and added noise in the resampled models. The distributions were plotted using a normal kernel smoothing function.

Areas of noise in the AIRS and resampled model data were reduced by smoothing the amplitude measurements using a 7 by 7 point boxcar filter and removing points in the original data where the amplitude of the smoothed data is below the 70$^{\text{th}}$ percentile for nighttime data in each data set ($\sim$0.90 K for IFS 1 and 2, $\sim$1.18 K for AIRS and $\sim$0.89 K for ERA5 1). For the KDFs of ERA5 and the 1 km IFS run before being resampled as AIRS, the same method was used to remove areas of the data where the amplitude is below the 70$^{\text{th}}$ percentile of AIRS. As the amplitudes are lower in the resampled models than in AIRS, the 70$^{\text{th}}$ percentiles amplitude cutoffs are lower, so less noise is removed. This means more points from areas containing only noise are included in the KDFs for the resampled models. These areas generally have lower amplitudes and therefore lower momentum fluxes, as well as lower horizontal wavelengths (shown in Figures 4 and 6), so there will be a greater proportion of the resampled model data with lower values for these wave properties.

After being resampled as AIRS, the horizontal and vertical wavelength spectra of the models are more similar to the AIRS data (Figure 8d and f), compared to the spectra before resampling (Figure 8c and e), suggesting that the resampling methods used allow a fairer comparison between the observations and models to be made. There are a higher proportion of points with shorter horizontal wavelengths in the resampled models than in AIRS (Figure 8d), but this could be due to the 70$^{\text{th}}$ percentile amplitude cutoffs used. There is some quantization of the horizontal wavelengths longer than around 200 km, leading to multiple peaks in the KDFs for all data sets in Figure 8c and d. This is a result of these waves being approximated by Fourier modes in the 2D+1 ST, since they are long relative to the data size. For the resampled models, the peaks in the vertical wavelength are offset from the observations by around 2–3 km (Figure 8f). This could be due to the values of the AIRS vertical resolution, used for both resampling methods, being overestimated, but the vertical wavelengths in the resampled models could be too short, and the different 70$^{\text{th}}$ percentile amplitude cutoffs used for each data set could also affect these results.

Before resampling there are a greater proportion of higher amplitude GWs in the 1 km IFS run, up to $\sim$17 K, than in the AIRS observations, where the KDF tails off at $\sim$15 K, and significantly higher fraction of lower amplitude waves in ERA5, where the KDF tails off at around $\sim$9 K (Figure 8a). After the AIRS observational filter is applied to the models, the wave amplitudes are generally higher in AIRS, and there is a greater fraction of data points with lower amplitudes in ERA5 1, where the KDF decreases to $\sim$4 K, than in IFS 1 and 2, which have KDFs that decrease to $\sim$ 6 K (Figure 8b). Whilst the horizontal momentum flux KDFs are similar for AIRS and the 1 km IFS run before resampling (Figure 8g), they are generally higher in AIRS than in the resampled models (Figure 8h), as the momentum flux is proportional to vertical wavelengths and the square of the amplitudes. The results shown in Figure 8b, d and f suggest that there are high amplitude GWs with longer horizontal or vertical wavelengths in the AIRS retrieval data that are not present in the resampled models.

4.4 Bivariate histograms

Bivariate histograms are shown in Figure 9 to compare the wave properties in the resampled models and AIRS. The color bars show the normalised density, i.e. the number of counts in each bin divided by the total number of counts. Areas of noise in the data used for the bivariate histograms were also reduced using the same method as for the kernel distribution functions (Figure 8), described by the previous section. The number of points used to plot the histograms for each comparison can be found in Table S1 in the appendix. These values vary as points are only included if both data sets do not have a missing value in the point location. Values will be missing if the vertical wavelength in the point location is lower than 6 km or greater than 45 km, or the amplitude at that location is below the 70th percentile amplitude cutoff for each data set. Table S1 also shows the fraction of points above (f_a) and below (f_b) the 1:1 line (grey dashed line) in Figure 9.

The amplitudes in nighttime AIRS data are significantly higher than in the resampled models (Figure m, q and u), with f_b ranging from 0.815 (AIRS & IFS 2) to 0.859 (AIRS & ERA5 1) (Table S1). Stripes with no data can be seen in the bivariate histograms of the horizontal wavelengths (Figure 9b, f, j, n, r, and v), as the horizontal wavelengths are quantized at longer wavelengths. There are also more points where the AIRS data has a longer vertical wavelength than in the resampled models (Figure 9o, s and w) with f_b ranging from 0.557 (AIRS & ERA5 1) to 0.668 (AIRS & IFS 2). The histograms also show higher momentum fluxes in AIRS (Figure 9p, t and x) as a result of the higher GW amplitudes and vertical wavelengths. For the momentum flux (Figure 9d, h, l, p, t and x), the points in the bivariate histograms are very spread out suggesting there is little point-wise correlation between the data sets except from between IFS 1 and 2 (Figure 9d). The data points are also quite spread out for the vertical wavelength plots (Figure 9c, g, k, o, s, and w) indicating a low point-wise correlation. Data points are closer to the 1:1 line for IFS 1 and 2 for higher amplitude values (Figure 9a and d) and momentum flux. The fraction of points above and below the 1:1 line for IFS 1 & IFS 2 (Figure 9a–d) are similar for all wave properties shown (see Table S1) with the greatest difference in the fractions for the horizontal wavelength where f_a is 0.557 and f_b is 0.443.

5 Discussion

The methods used in this study have allowed a fairer comparison between the models and AIRS observations compared to previous work. However some issues still remain, including how noise is selected and added, AIRS's observational filter and the amplitude cutoffs used.

The AIRS temperature perturbations containing only noise were selected by ordering the granules from lowest to highest variance and selecting granules manually for different regions in nighttime data, during the same period of the year as the data used in this study. This meant the noise added to the resampled models was better correlated to the location of each resampled model granule and could include large noise structures compared to previous methods used (e.g. in Hindley et al., 2021; Okui et al., 2023). However this method of selecting noise granules would be too time consuming for a longer data set, so using machine learning to identify whether granules contain only noise or GWs could be a better approach.

A limitation of this work is that only GWs in the 1 km IFS run with wavelengths in the AIRS's observational filter can be compared to AIRS observations. The AIRS retrieval data has a low vertical resolution. In future work, data from instruments with different observational filters, such as limb sounders or satellites using GPS radio occultation, could be used to validate some of the resolved GWs in the 1 km IFS run with wavelength ranges outside of AIRS's observational filter. Limb sounders have a low horizontal resolution, but higher vertical resolution than nadir sounders like AIRS.

Whilst the location and timing of the GWs agree well in the resampled 1 km IFS run and AIRS observations, the mean amplitudes were found to be significantly lower in the resampled 1 km IFS run, by a factor of ∼2.4, but higher than in the lower resolution ERA5 reanalysis. As a result of this, the horizontal momentum flux is also lower in the resampled models compared to the observations. Kruse et al. (2022) found that GW amplitudes in a lower resolution run of the IFS, with a grid-spacing, of ∼9 km, were lower than in AIRS observations. High amplitude GWs, seen in the AIRS observations, are not found in the resampled 1 km IFS data (Figure 8), but are present in the 1 km IFS run before resampling suggesting these waves have wavelengths outside of AIRS's observational filter.

Amplitude cutoffs were used to reduce areas of noise included in the AIRS retrieval and resampled model data for the kernel distribution functions (Figure 8) and bivariate histograms (Figure 9), but this could not remove all areas of noise without also removing areas of low amplitude GWs. This means that these results will be affected by the remaining noise. These cutoffs were chosen by finding the 70^{th} percentile of all nighttime data during the first 10 days of November 2018 for each data set. Due to the lower wave amplitudes in the resampled models, the 70^{th} percentiles were lower than for the AIRS retrieval data, so more areas of noise are included in the results for the resampled models.

The two methods used to resample the 1 km IFS lead to quite similar results and worked effectively to smooth the model data to AIRS's resolution, resulting in more similar distributions of GW horizontal and vertical wavelengths, compared to before the observational filter was applied (Figure 8c–f). However, the peaks of the distributions of vertical wavelength for the resampled models were found to be around 2–3 km lower than for AIRS (Figure 8f). This could be a result of the AIRS resolution values, used to smooth the model data, being overestimated, but could also be due to differences in the vertical wavelengths in the resampled models compared to the observations. These results will also be affected by the noise in the data.

6 Summary and Conclusions

In this study, gravity wave (GW) properties in a ∼1.4 km gravity-wave-resolving run (TCo7999 resolution) of the IFS are compared to AIRS observations over Asia and surrounding regions, using nighttime data during the first 10 days of November 2018. The results show a good level of fidelity for the model by comparison to the observations, but with important differences, discussed below.

Two different methods were used to resample the 1 km IFS run to facilitate this comparison, the first method by smoothing to AIRS' resolution and then interpolating to the measurement location, and the second by oversampling the model and then producing a weighted average of the oversampled points. Although small differences are seen, they generally produce quite similar results. Since method 2 is significantly more computationally expensive, method 1 may be better suited for comparing models to AIRS. This result does not necessarily hold in the general case: the large- and vertically-deep volume of nadir measurements, such as those from AIRS, is likely less sensitive to footprint positioning and morphology than finer-vertical wider-horizontal measurements such as from limb sounders, and this will be investigated further in future work.

Based on these results, the output from the ERA5 reanalysis was also resampled as AIRS using method 1, to see how well the 1 km IFS run resolves GWs in comparison to this lower resolution (and slightly chronologically older) model with assimilative capabilities. Noise derived from wave-free AIRS observations was also added to the simulated data to produce a fairer comparison with the very noisy observations, following experience in Okui et al. (2023) which showed the significant effect such noise has on 1:1 comparisons. Finally, the 2D+1 S-Transform analysis of Wright et al. (2021) was used to find the wave properties for each data set.

The results of this analysis lead to the following conclusions:

1. GWs in the 1 km IFS run can be seen at similar locations and times, and with similar wave morphology to AIRS, suggesting that the model works well in this regard. ERA5 waves are in general less morphologically consistent with observations, and in particularly often have inconsistently long horizontal wavelengths, but do occur at similar locations and times to the observations in many if not most cases.
2. Measured amplitudes and momentum fluxes are significantly lower in both resampled models than in AIRS data, with ERA5 amplitudes slightly lower (and thus less observationally-consistent) than those in the 1 km model. This difference is large, with a long tail of high-amplitude AIRS measurements (Figure 8b) which in turns drives a similar difference in momentum fluxes. Investigation of the raw model data shows that many high-amplitude waves in the 1 km IFS run have wavelengths too horizontally-short for AIRS to observe (see e.g. Figure 8), which are thus not seen in the resampled model. Given that the overall amplitude and momentum flux distributions (Figure 8a,g) in the raw models are broadly similar to AIRS, this may suggest that wave activity in the model has plausible total amplitudes and fluxes, but skewed to much shorter wavelengths than in the true GW spectrum.
3. Vertical wavelengths in both ERA5 and the 1 km IFS run are significantly shorter than in AIRS observations, even after resampling to match the observational resolution. This difference is typically $\sim$2–3 km, i.e. approximately 10-20% of the observed wavelengths. A similar effect is seen in horizontal wavelength, with resampled model wavelengths showing a sharp peak at short wavelengths while observed horizontal wavelengths peak at a slightly higher value but exhibit a much longer and flatter distribution. Both of these conclusions are difficult to decouple from the effects of (both real and synthesised) measurement noise, and further work is needed to address this question more carefully.

This work highlights the importance of carefully applying the observational filter of the observing platform to models before comparing GWs in simulations to those in observations, which is shown to be necessary for producing a meaningful comparison in this study. This is important for accurate testing of how well GWs are resolved in high resolution models, with further implications for parameterization development, as this increasingly frequently uses high-resolution models of this nature as a 'truth' for tuning purposes.

Open Research

The AIRS temperature data used in the study were computed from AIRS radiances using the retrieval scheme described in Hoffmann and Alexander (2009). The 3D AIRS temperature retrieval can be obtained at `https://datapub.fz-juelich.de/slcs/airs/gravity_waves`. The ECMWF ERA5 reanalysis data at 0.25° resolution can be downloaded from the Copernicus Climate Data Store at `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview` (Copernicus Climate Change Service, 2023). For the 1 km IFS run, the size of the raw model output on the native grid is a few hundred TB, so it is not possible for all of the data to be made available. However, the post processed data will be retained and is available on request. Code written in MATLAB (available at `https://uk.mathworks.com/products/`

`matlab.html`) was used to resample the models as AIRS, analyse the gravity wave properties and produce the figures. The MATLAB code used is available at `https://github.com/Emily-Lear/Comparing_the_1_km_IFS_run_to_AIRS_observations.git`.

Acknowledgments

The high resolution IFS simulation (TCo7999) used in this study was performed using the resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility supported under contract DE-AC05-00OR22725. The grants which supported this work include Royal Society Research Grant RGF\R1\180010 supporting E. J. Lear and C. J. Wright, Royal Society Research Fellowships UF160545 and URF\R\221023 supporting C. J. Wright, and NERC Grants NE/W003201/1 and NE/S00985X/1 supporting C. J. Wright and N. Hindley. C Wright, N Hindley and I Polichtchouk's contribution to this work was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #567. The authors would like to thank L. A. Holt for helpful discussions related to the interpretation of the 8µm AIRS brightness data.

References

Albers, J. R., & Birner, T. (2014). Vortex Preconditioning due to Planetary and Gravity Waves prior to Sudden Stratospheric Warmings. *J. Atmos. Sci.*, *71*, 4028–4054. doi: 10.1175/JAS-D-14-0026.1

Alexander, M. J., & Barnet, C. (2007). Using satellite observations to constrain parameterizations of gravity wave effects for global models. *J. Atmos. Sci.*, *64*, 1652-1665. doi: 10.1175/JAS3897.1

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., ... Watanabe, S. (2010, JUL). 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*(650, A), 1103-1124. doi: {10.1002/qj.637}

Alexander, M. J., Gille, J., Cavanaugh, C., Coffey, M., Craig, C., Eden, T., ... Dean, V. (2008, May). Global estimates of gravity wave momentum flux from high resolution dynamics limb sounder observations. *Journal of Geophysical Research*, *113*(D15), 0148-0227. Retrieved from `https://doi.org/10.1029/2007jd008807` doi: 10.1029/2007jd008807

Alexander, P., Schmidt, T., & de la Torre, A. (2018). A method to determine gravity wave net momentum flux, propagation direction, and "real" wavelengths: A gps radio occultations soundings case study. *Earth and Space Science*, *5*(6), 222-230. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017EA000342` doi: https://doi.org/10.1002/2017EA000342

Aumann, H., Chahine, M., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., ... Susskind, J. (2003, Feb). Airs/amsu/hsb on the aqua mission: design, science objectives, data products, and processing systems. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 253-264. doi: 10.1109/TGRS.2002.808356

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., ... Yamashita, Y. (2011). Multimodel climate and variability of the stratosphere. *J. Geophys. Res.*, *116*(D5), n/a–n/a. (D05102) doi: 10.1029/2010JD014995

Carslaw, K. S., Wirth, M., Tsias, A., Luo, A. B. P., Dörnbrack, A., Leutbecher, M., ... Peter, T. (1998). Increased stratospheric ozone depletion due to mountain-induced atmospheric waves. *Nature*, *391*, 675–678. doi: https://doi.org/10.1038/35589

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Bblaisdell, J.,

. . . Zhou, L. (2006, 07). AIRS: Improving Weather Forecasting and Providing New Data on Greenhouse Gases. *Bulletin of the American Meteorological Society*, *87*(7), 911-926. doi: 10.1175/BAMS-87-7-911

Copernicus Climate Change Service. (2023). *ERA5 hourly data on pressure levels from 1979 to present.* European Centre for Medium-Range Weather Forcasts (ECMWF),`https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview`. ([Accessed 15 Aug 2023])

ECMWF. (2020). *L137 model level definitions.* `https://confluence.ecmwf.int/display/UDOC/L137+model+level+definitions`. ([Accessed 28 Feb 2022])

ECMWF. (2021). *ERA5.* `https://confluence.ecmwf.int/display/CKB/ERA5`. ([Accessed 15 Aug 2023])

ECMWF. (2023). *IFS Documentation.* `https://www.ecmwf.int/en/publications/ifs-documentation`.

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

Ern, M., Trinh, Q. T., Preusse, P., Gille, J. C., Mlynczak, M. G., III, J. M. R., & Riese, M. (2018, April). GRACILE: a comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth System Science Data*, *10*(2), 857–892. doi: 10.5194/essd-10-857-2018

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

Fritts, D. C., Riggin, D. M., Balsley, B. B., & Stockwell, R. G. (1998). 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*(3), 297–300. doi: 10.1029/97GL03702

Geller, M., Alexander, M. J., Love, P., Bacmeister, J., Ern, M., Hertzog, A., . . . Zhou, T. (2013). A Comparison between Gravity Wave Momentum Fluxes in Observations and Climate Models. *Journal of Climate*, *26*, 6383–6405. doi: 10.1175/JCLI-D-12-00545.1

Harvey, V. L., Randall, C. E., Becker, E., Smith, A. K., Bardeen, C. G., France, J. A., & Goncharenko, L. P. (2019). Evaluation of the mesospheric polar vortices in waccm. *Journal of Geophysical Research: Atmospheres*, *124*(20), 10626-10645. doi: https://doi.org/10.1029/2019JD030727

Hindley, N. P., Smith, N. D., Wright, C. J., Rees, D. A. S., & Mitchell, N. J. (2016, June). A two-dimensional stockwell transform for gravity wave analysis of AIRS measurements. *Atmospheric Measurement Techniques*, *9*(6), 2545–2565. doi: 10.5194/amt-9-2545-2016

Hindley, N. P., Wright, C. J., Gadian, A. M., Hoffmann, L., Hughes, J. K., Jackson, D. R., . . . Ross, A. N. (2021). Stratospheric gravity waves over the mountainous island of south georgia: testing a high-resolution dynamical model with 3-d satellite observations and radiosondes. *Atmospheric Chemistry and Physics*, *21*(10), 7695–7722. Retrieved from `https://acp.copernicus.org/articles/21/7695/2021/` doi: 10.5194/acp-21-7695-2021

Hindley, N. P., Wright, C. J., Hoffmann, L., Moffat-Griffin, T., & Mitchell, N. J. (2020, November). An 18-year climatology of directional stratospheric gravity wave momentum flux from 3-d satellite observations. *Geophysical Research Letters*, *47*(22), e2020GL089557. doi: 10.1029/2020gl089557

Hindley, N. P., Wright, C. J., Smith, N. D., Hoffmann, L., Holt, L. A., Alexander, M. J., . . . Mitchell, N. J. (2019). 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. doi: 10.5194/acp-19-15377-2019

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

Hoffmann, L., & Alexander, M. J. (2010, 10). Occurrence frequency of convective gravity waves during the north american thunderstorm season. *J. Geophys. Res.*, *115*(D20). doi: 10.1029/2010JD014401

Hoffmann, L., Spang, R., Orr, A., Alexander, M. J., Holt, L. A., & Stein, O. (2017). A decadal satellite record of gravity wave activity in the lower stratosphere to study polar stratospheric cloud formation. *Atmospheric Chemistry and Physics*, *17*(4), 2901–2920. Retrieved from https://acp.copernicus.org/articles/17/2901/2017/ doi: 10.5194/acp-17-2901-2017

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

Jewtoukoff, V., Hertzog, A., Plougonven, R., de la Camara, A., & Lott, F. (2015). Comparison of gravity waves in the southern hemisphere derived from balloon observations and the ecmwf analyses. *Journal of the Atmospheric Sciences*, *72*(9), 3449-3468. doi: 10.1175/JAS-D-14-0324.1

Kruse, C. G., Alexander, M. J., Hoffmann, L., van Niekerk, A., Polichtchouk, I., Bacmeister, J. T., ... Stein, O. (2022). Observed and modeled mountain waves from the surface to the mesosphere near the drake passage. *Journal of the Atmospheric Sciences*, *79*(4), 909 - 932. Retrieved from https://journals.ametsoc.org/view/journals/atsc/79/4/JAS-D-21-0252.1.xml doi: https://doi.org/10.1175/JAS-D-21-0252.1

Lane, T. P., Doyle, J. D., Sharman, R. D., Shapiro, M. A., & Watson, C. D. (2009). Statistics and dynamics of aircraft encounters of turbulence over greenland. *Monthly Weather Review*, *137*(8), 2687 - 2702. Retrieved from https://journals.ametsoc.org/view/journals/mwre/137/8/2009mwr2878.1.xml doi: https://doi.org/10.1175/2009MWR2878.1

Lund, T. S., Fritts, D. C., Wan, K., Laughman, B., & Liu, H.-L. (2020). Numerical simulation of mountain waves over the southern andes. part i: Mountain wave and secondary wave character, evolutions, and breaking. *Journal of the Atmospheric Sciences*, *77*(12), 4337 - 4356. doi: https://doi.org/10.1175/JAS-D-19-0356.1

Okui, H., Wright, C. J., Hindley, N. P., Lear, E. J., & Sato, K. (2023). A comparison of stratospheric gravity waves in a high-resolution general circulation model with 3-d satellite observations. *Journal of Geophysical Research: Atmospheres*, *128*(13), e2023JD038795. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023JD038795 doi: https://doi.org/10.1029/2023JD038795

Parkinson, C. (2003). Aqua: an earth-observing satellite mission to examine water and other climate variables. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 173-183. doi: 10.1109/TGRS.2002.808319

Plougonven, R., de la Cámara, A., Hertzog, A., & Lott, F. (2020). How does knowledge of atmospheric gravity waves guide their parameterizations? *Quarterly Journal of the Royal Meteorological Society*, *146*(728), 1529-1543. doi: 10.1002/qj.3732

Polichtchouk, I., Shepherd, T. G., Hogan, R. J., & Bechtold, P. (2018, May). Sensitivity of the brewer-dobson circulation and polar vortex variability to parameterized nonorographic gravity wave drag in a high-resolution atmospheric model. *Journal of the Atmospheric Sciences*, *75*(5), 1525–1543. doi: 10.1175/jas-d-17-0304.1

Polichtchouk, I., van Niekerk, A., & Wedi, N. (2023). Resolved gravity waves in the extratropical stratosphere: Effect of horizontal resolution increase from o(10) to o(1) km. *Journal of the Atmospheric Sciences*, *80*(2), 473 - 486. Retrieved from https://journals.ametsoc.org/view/journals/atsc/80/2/JAS-D-22-0138.1.xml doi: https://doi.org/10.1175/JAS-D-22-0138.1

Polichtchouk, I., Wedi, N., & Kim, Y.-H. (2022). Resolved gravity waves in the tropical stratosphere: Impact of horizontal resolution and deep convection parametrization. *Quarterly Journal of the Royal Meteorological Society*, *148*(742), 233-251. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4202 doi: https://doi.org/10.1002/qj.4202

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

Sato, K., & Hirano, S. (2019). The climatology of the brewer–dobson circulation and the contribution of gravity waves. *Atmospheric Chemistry and Physics*, *19*(7), 4517–4539. Retrieved from https://acp.copernicus.org/articles/19/4517/2019/ doi: 10.5194/acp-19-4517-2019

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

Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996, Apr). Localization of the complex spectrum: the s transform. *IEEE Transactions on Signal Processing*, *44*(4), 998-1001. doi: 10.1109/78.492555

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

Watanabe, S., & Miyahara, S. (2009). Quantification of the gravity wave forcing of the migrating diurnal tide in a gravity wave–resolving general circulation model. *Journal of Geophysical Research: Atmospheres*, *114*(D7), 0148-0227. doi: https://doi.org/10.1029/2008JD011218

Wedi, N. P., Polichtchouk, I., Dueben, P., Anantharaj, V. G., Bauer, P., Boussetta, S., ... Vitart, F. (2020). A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002192. doi: https://doi.org/10.1029/2020MS002192

Wright, C. J., & Banyard, T. P. (2020). Multidecadal measurements of utls gravity waves derived from commercial flight data. *Journal of Geophysical Research: Atmospheres*, *125*(19), e2020JD033445. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD033445 (e2020JD033445 10.1029/2020JD033445) doi: https://doi.org/10.1029/2020JD033445

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

Wright, C. J., Hindley, N. P., Alexander, M. J., Holt, L. A., & Hoffmann, L. (2021). Using vertical phase differences to better resolve 3d gravity wave structure. *Atmospheric Measurement Techniques*, *14*(9), 5873–5886. Retrieved from https://amt.copernicus.org/articles/14/5873/2021/ doi: 10.5194/amt-14-5873-2021

Wright, C. J., Hindley, N. P., Hoffmann, L., Alexander, M. J., & Mitchell, N. J. (2017). 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. doi: 10.5194/acp-17-8553-2017

Wu, D. L. (2004). Mesoscale gravity wave variances from amsu-a radiances. *Geophy. Res. Lett.*, *31*(12), 1944-8007. (L12114) doi: 10.1029/2004GL019562

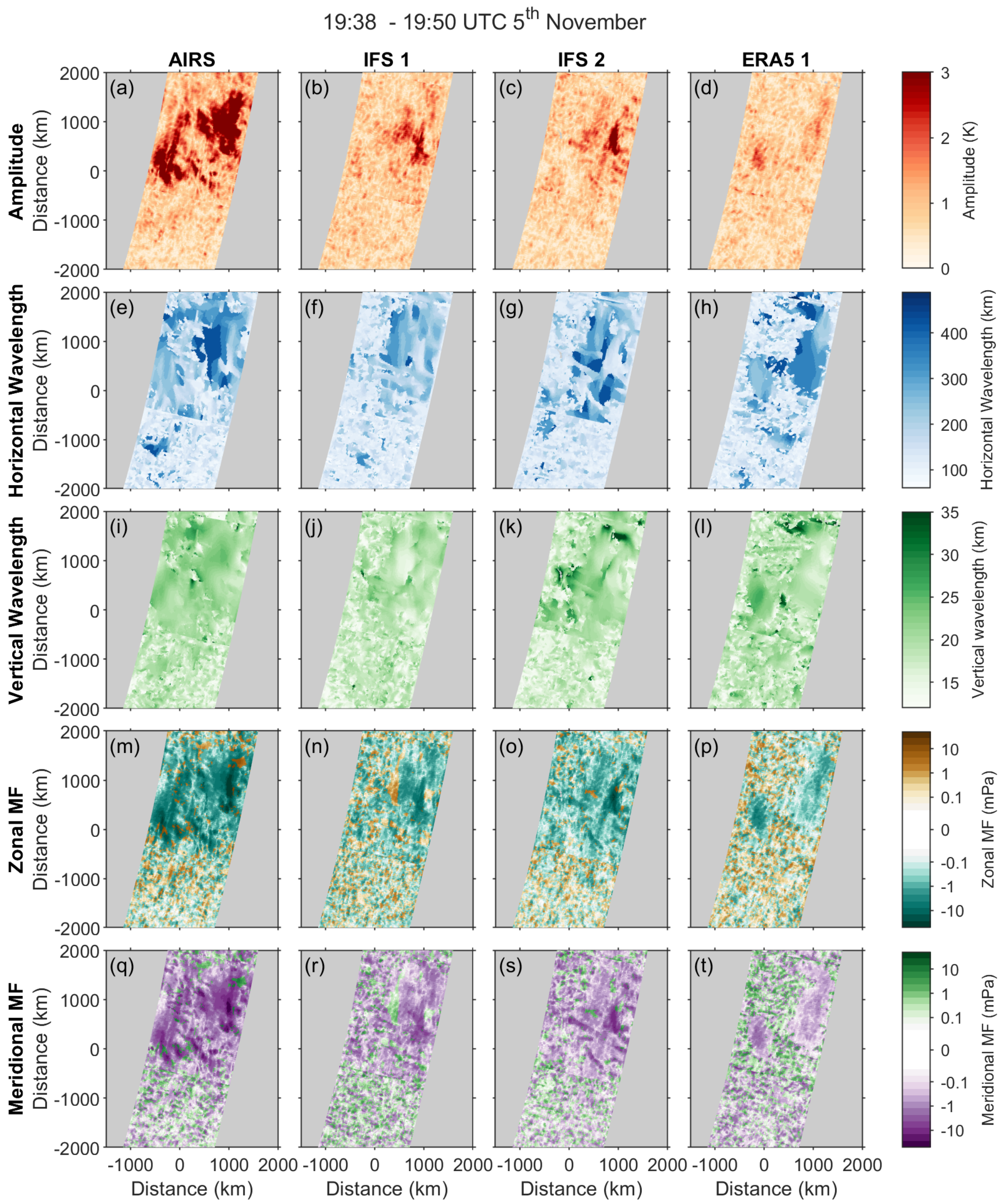

Figure 4. Wave properties for the granules shown in Figure 3, including the amplitude, horizontal and vertical wavelengths and the zonal, and meridional momentum flux (MF). The zonal and meridional momentum flux are shown on a log color scale. Data points were removed where the vertical wavelength is below 6 km or above 45 km.

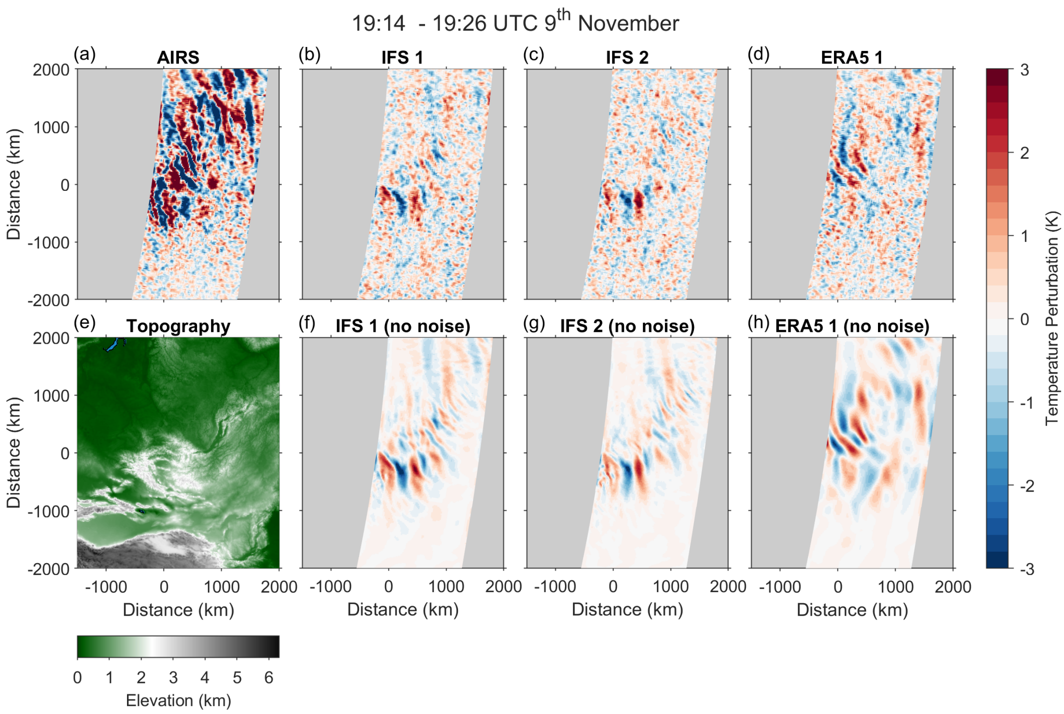

Figure 5. As in Figure 3, but for AIRS granules with mean times from 19:14 to 19:26 UTC on the 9th November 2018 (a) and the resampled models at the closest times (b–d) with and (f–h) without AIRS noise added.

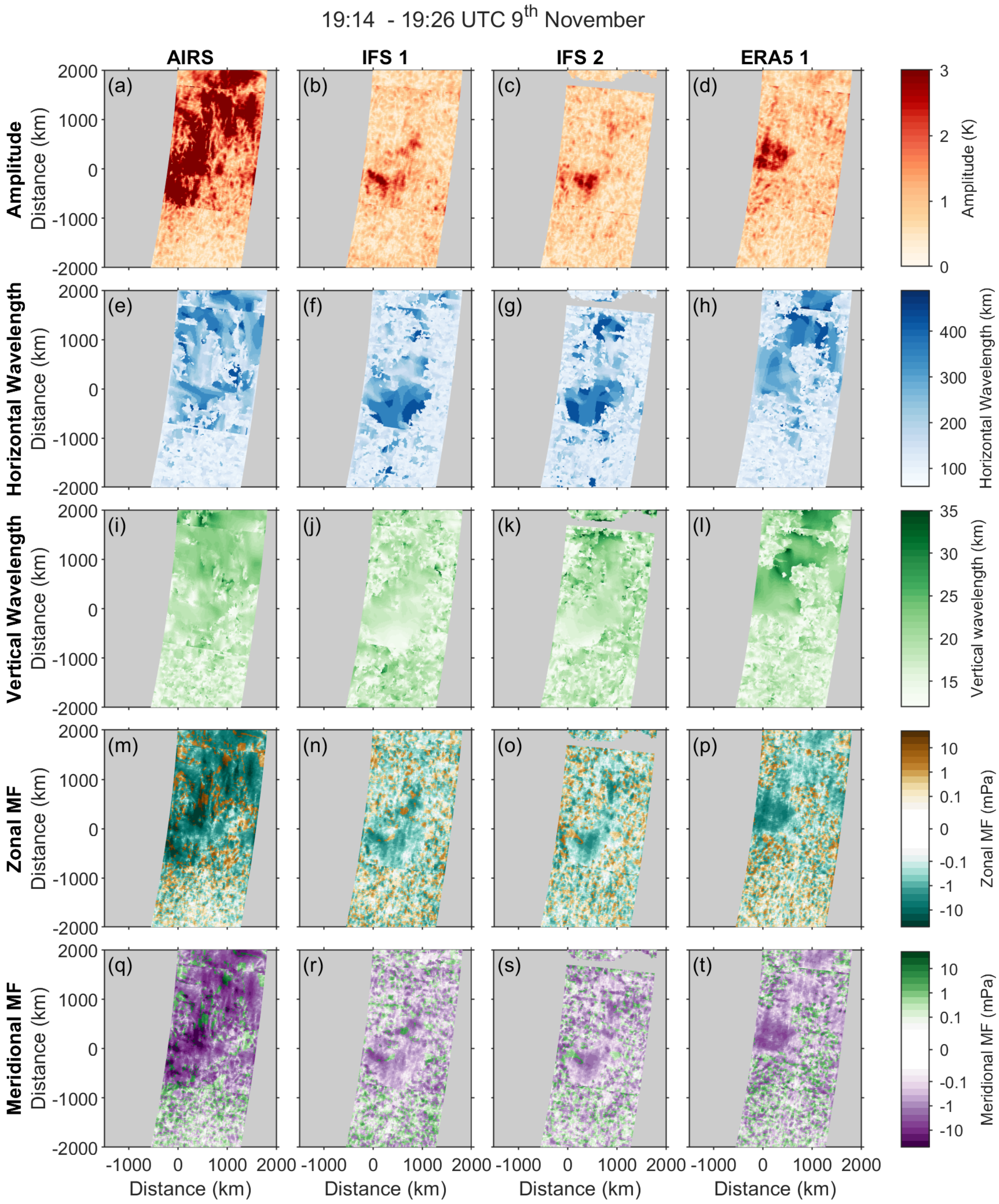

Figure 6. As in Figure 4, but for wave properties for the granules shown in Figure 5. The zonal and meridional momentum flux (MF) are shown on a log color scale.

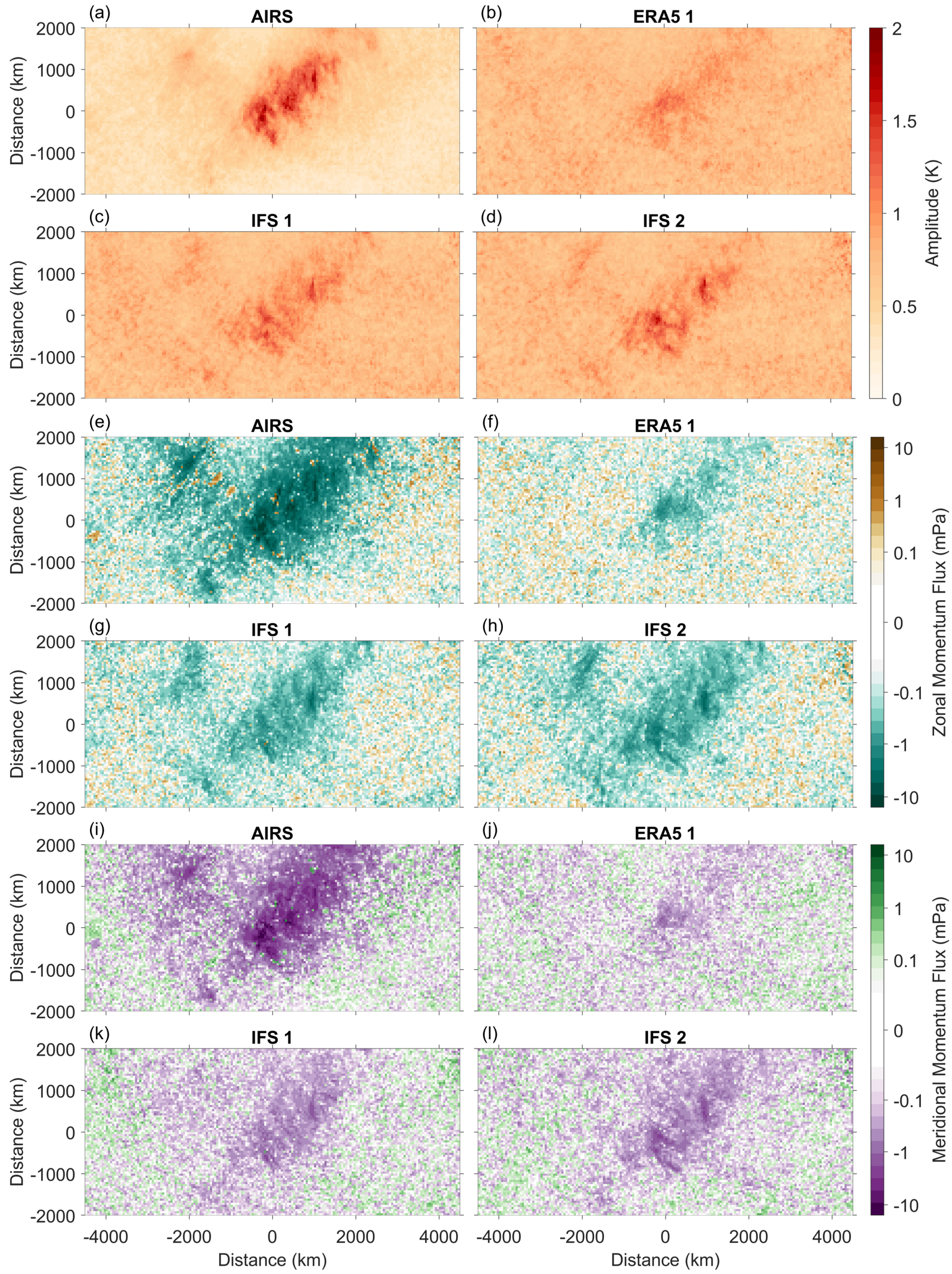

Figure 7. Mean amplitude (a–d) and mean zonal (e–h) and meridional (i–l) momentum flux in the region shown in Figure 1 for nighttime data in the first 10 days of November 2018. This is plotted on a regular distance grid with a point spacing of 50 km by 50 km. AIRS amplitudes in (a) are divided by 2. The zonal and momentum flux are shown on a log color scale.

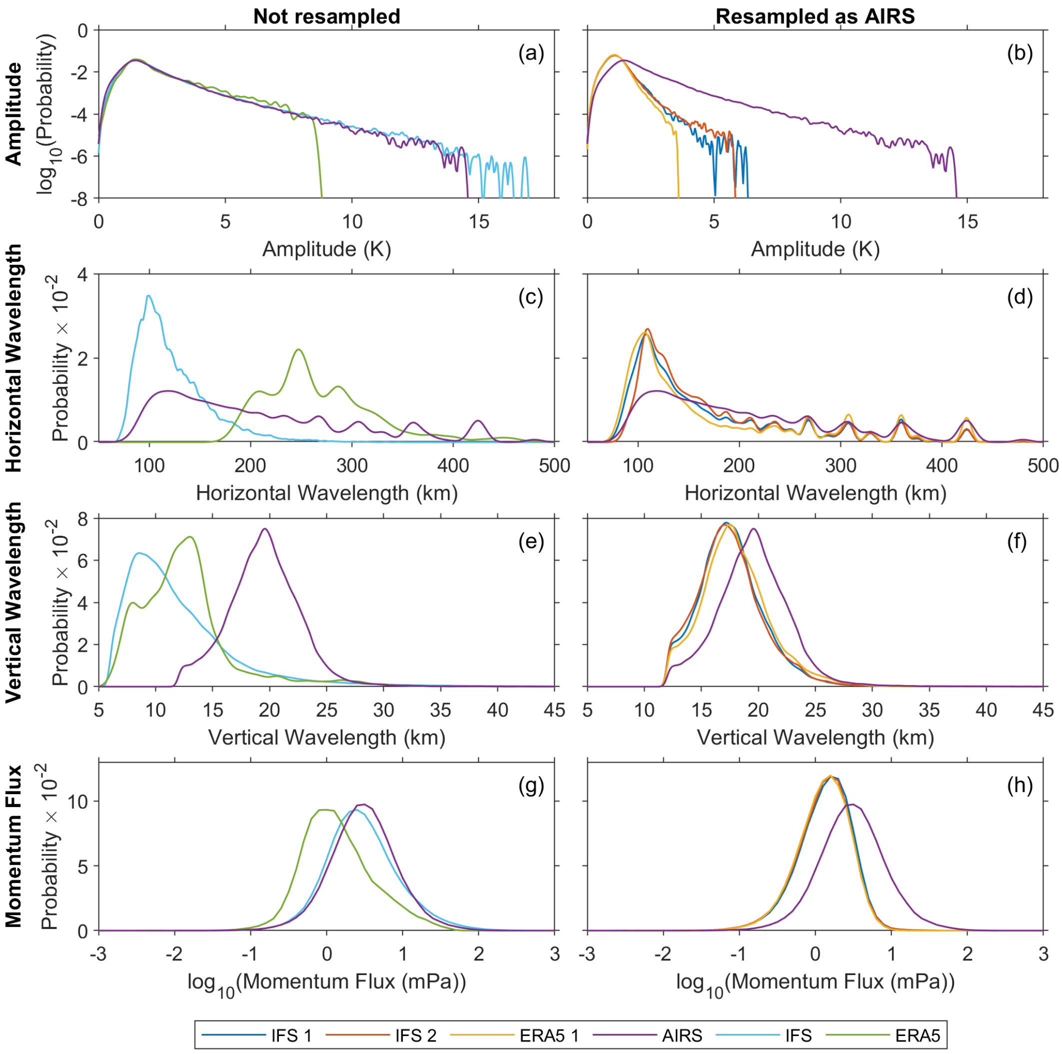

Figure 8. Kernel distribution functions (KDFs) for the wave properties in nighttime data at 39 km altitude for AIRS and the models before resampling (the 1 km IFS run (IFS) and ERA5) (panels a, c, e and g) and for the resampled 1 km IFS run (IFS 1 and 2) and ERA5 resampled as AIRS (ERA5 1) shown with the KDFs for AIRS in panels (b, d, f and h). The KDFs for the amplitude have been logged to base 10. Noise is reduced by using a 70^{th} percentile amplitude cutoff for the resampled models and AIRS, and AIRS's amplitude cutoff is also used for ERA5 and the 1 km IFS run before resampling.

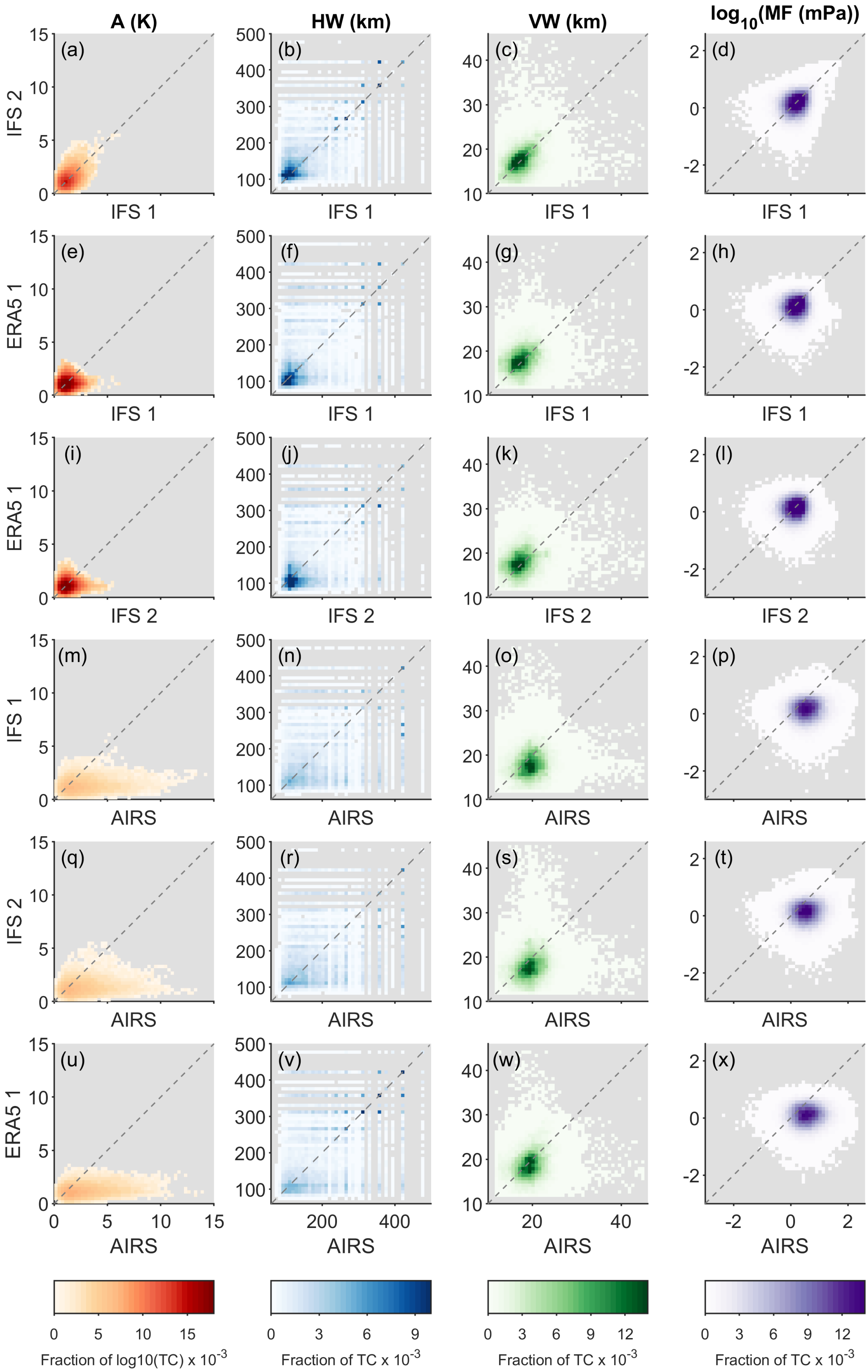

Figure 9. Bivariate histograms of wave properties (amplitude (A), horizontal wavelength (HW), vertical wavelength (VW) and momentum flux (MF)) using nighttime data from the first 10 days of November 2018 in the region shown in Figure 1 at 39 km altitude. The color scales show the fraction of the total bin counts (TC) for each subplot.