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

Emily Jane Lear[1], Corwin James Wright[1], Neil Hindley[1], Inna Polichtchouk[2], and Lars Hoffmann[3]

[1]University of Bath
[2]ECMWF
[3]Forschungszentrum Jülich

May 02, 2024

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 is resampled as AIRS using two different methods to allow for comparison of gravity wave properties
- Gravity waves can be seen in the resampled IFS run and AIRS at similar times and locations
- Mean amplitudes in the resampled IFS run are found to be significantly lower than in the observations by a factor of ∼2.77

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 become 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), is compared to observations from the Atmospheric Infrared Sounder (AIRS) instrument, on NASA's Aqua satellite, to test how well the model resolves GWs that AIRS can observe. In this analysis, nighttime data are used from the first 10 days of November 2018 over part of 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 are 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, transport energy and momentum and affect the dynamics of the atmosphere. Gravity waves in a high resolution run of the European Centre for Medium Range Weather Forecasts Integrated Forecasting System (IFS) weather model are compared to those in observations from the AIRS (Atmospheric Infrared Sounder) instrument on NASA's Aqua satellite, to test how well these waves are resolved in the model. Nighttime data is compared over part of Asia and surrounding regions, during the first 10 days of November 2018. Since the high resolution IFS run has a higher vertical resolution, and a significantly higher horizontal resolution than the satellite observations, the model is resampled as if the satellite was viewing the model atmosphere. This removes gravity waves with horizontal and vertical wavelengths outside of the ranges that can be seen in the observations, allowing the data sets to be compared. Gravity waves formed by wind flowing over mountain ranges, can be seen at similar times and in similar locations in the IFS run and observations, but wave amplitudes in the resampled IFS run are 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 that 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 non-operational configurations, the resolution is sufficiently high that 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 ECMWF ERA5 reanalysis, with a horizontal resolution of 31 km, to understand the impact of the increased resolution of the km-scale IFS on resolved GW properties compared to operational configurations. Unlike the kilometre-scale run of the IFS, ERA5 is generated using data assimilation.

The ECMWF IFS, run at different resolutions, has been validated by comparisons to observations in previous work. 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. Temperature variances in the IFS run at resolutions of 9 km and 4 km were also compared to AIRS observations globally in August 2016 (Stephan et al., 2019). The results of this study showed that the spatial structure of these temperature variances was similar for the model and observations in the mid- to high latitudes of the Southern Hemisphere. GW potential energy was found to be underestimated in the middle atmosphere in three IFS versions run at $\sim$9 km resolution, compared to data from the Compact Rayleigh Autonomous Lidar (CORAL) at Río Grande in the lee of the Southern Andes (Gisinger et al., 2022). GW momentum fluxes in the ECMWF operational analysis, produced using the IFS and 4D variational data assimilation, at a grid spacing of approximately 16 km (T1279 resolution) 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 analysis was also found to have lower wave amplitudes by a factor of 2–3 compared to AIRS observations, using data from 2003 to 2012 (Hoffmann et al., 2017).

However, making comparisons between observed and simulated GWs is not straightforward. This is because no instrument 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 can limit 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 data set 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 two different sampling methods are investigated to create this data set: 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).

A further issue with model and observation comparisons of GWs is instrument noise. Sources of noise include instrument radiometric noise, planetary waves that are not fully removed by detrending to find the perturbations from the background atmosphere, mesoscale convective systems and turbulence. Gravity waves with amplitudes too far below the noise level to be distinguished from the noise cannot be compared to models and the noise cannot be removed from the observations. In this study, noise is added to the resampled model data, to avoid comparing low amplitude waves that could not be observed by the AIRS satellite instrument.

The 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 part of continental Asia and surrounding regions, so this region is selected to perform the comparison (see Figure 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). This may suggest a strong role for the GWs with long vertical and short horizontal wavelengths that 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 then discussed in Section 5, and the summary and conclusions are presented in Section 6.

2 Data

2.1 AIRS

Stratospheric temperature data are used from the AIRS instrument on NASA's Aqua satellite (Hoffmann & Alexander, 2009). Aqua's orbit is sun-synchronous and near-polar, with a period of 98.8 minutes. This allows AIRS to obtain data with near daily global coverage. 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 a viewing angle of +49.5° to -49.5° across track, with 90 elements and a swath width of ∼1780 km and has a horizontal resolution of ∼13.5 km × 13.5 km at nadir which reduces to 41 km × 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 are 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 12 AIRS CO_2 emission channels near 15 µm for daytime and nighttime, and an additional 23 channels near 4 µm 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 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 ∼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.

2.2 High Resolution IFS Simulation (TCo7999)

The high resolution run of the ECMWF IFS used in this study 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 hybrid sigma/pressure (model) levels, from 0.01 hPa down to the surface and the spacing between the levels increases with altitude (Wedi et al., 2020). GWs with the smallest wavelengths are likely to be strongly damped by numerical diffusion in the IFS (Polichtchouk et al., 2023). 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. 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. 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 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 ∼9 to ∼1 km and the deep convection parameterization and found GWs are still under-resolved at a grid spacing of ∼9 km, compared to GWs at the ∼1 km resolution.

In this study, 3 hourly 1 km IFS temperature data are interpolated onto a regular longitude-latitude grid, with a resolution of 0.1° × 0.1°, with a regular distance spacing of ∼11.1 km at the equator. This reduced resolution is chosen to make the data easier to use and does not affect the results, as this is still a significantly higher than the horizontal resolution of the AIRS retrieval in the region investigated.

2.3 ERA5

The ECMWF ERA5 atmospheric reanalysis is global and is run from 1940 to the present (Copernicus Climate Change Service, 2023) at a horizontal resolution of 31 km. ERA5 uses 4D-Var 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 have been regridded to a regular latitude-longitude grid with a resolution of 0.25° (Copernicus Climate Change Service, 2023). This data is hourly, but since the 1 km IFS run is 3 hourly, ERA5 data is only used at every 3 hours during the time period investigated, to avoid time differences between the two models affecting the results.

3 Methods

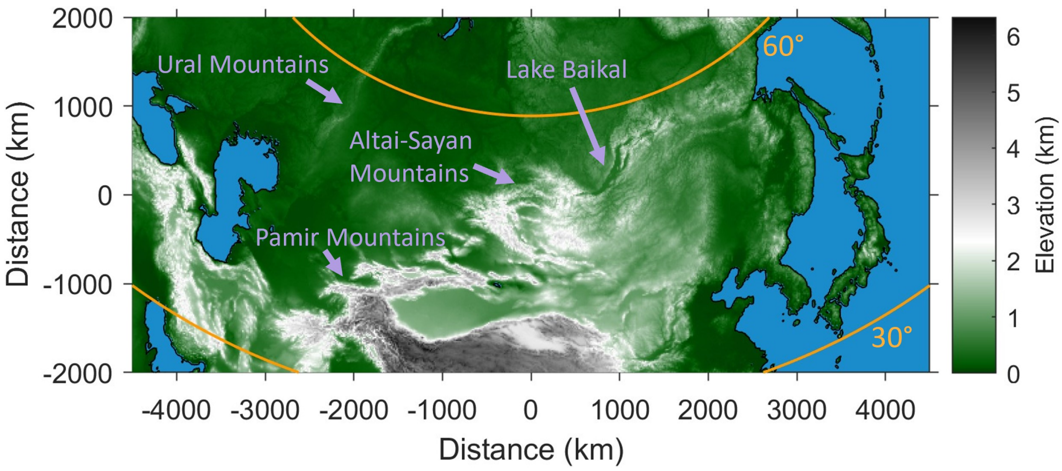

Figure 1. Map of topography in the region investigated on a regular distance grid centred at 52° latitude with a 15 km grid-spacing, 94° longitude. Coastlines are shown in black. Yellow-orange lines show the location of 30°N and 60°N latitude on the regular distance plot. Regions with mountain ranges are labelled with light purple arrows. The longitude and latitude ranges of the data shown are 22.9°E - 165.2°E and 23.0° N – 70.0°N, respectively. The elevation data used to plot this map are from the ETOPO Global Relief Model (NOAA National Centers for Environmental Information, 2022) at a resolution of 60 arc-seconds.

Data from the first 10 days of November 2018 for AIRS, the 1 km IFS run and the ERA5 reanalysis are used, as the 1 km IFS run was initialized on the 1st of this month at 00:00 UTC. This time period is selected, due to the expectation that the 1 km IFS's background temperature and wind structure, which affect the generation and propagation of GWs, will remain similar to observations in this period (Polichtchouk et al., 2022). 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 are located in a part of 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 are also used in Hoffmann et al. (2013) to identify individual GW events. AIRS granules with any data points located in the region shown, are selected for this study and the models are resampled as these AIRS granules, but only data points which are within the region are included in the results.

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 (from around 21–54 km) and the noise is lower (from around 21–39 km) for nighttime data in polar winter, mid latitudes and the tropics, in comparison with altitudes outside of these ranges. The results are presented only for nighttime data, due to the the higher vertical resolution and lower retrieval error of the nighttime AIRS retrieval compared to the daytime retrieval.

3.1 Resampling Methods

In this study, the observational filter of the AIRS retrieval is 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 are selected from the 3 hourly 1 km IFS run at the closest time to the measurement time of each AIRS granule, and are resampled as that granule. ERA5 data are 1 hourly, but are selected at the same 3 hourly time as the 1 km IFS run for resampling as each AIRS granule, so that comparison between the results for the resampled models are not affected by time differences. The models are not interpolated to the AIRS measurement times as this would smooth out small scale structures such as GWs (Wright & Hindley, 2018). Two different methods are used to resample the 1 km IFS run as AIRS. As the vertical resolution of the AIRS retrieval varies with latitude, different values are used to smooth to AIRS vertical resolution depending on whether most of the data points in the model, interpolated to an AIRS granule location, are in the tropics (<30° latitude), mid-latitudes (30°–60° latitude) or polar region (>60° latitude). The yellow-orange lines in Figure 1 are at the boundaries of these regions. The first resampling method is run on a desktop computer, whereas the second is more computationally expensive (Wright & Hindley, 2018) and requires the use of high performance computing.

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 were previously interpolated onto a regular longitude-latitude grid with a spacing of 0.1°, are selected at the closest 3 hourly time to each AIRS granule. These 1 km IFS data are 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 are 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 are interpolated onto the location of the AIRS granule. The data are then interpolated to a regular distance spacing in the vertical of 0.1 km from 26 to 55 km altitude, so they 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 is 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 are used (shown in Figure 2, of Hindley et al. (2019)) for each of the granules, depending on whether they are located mostly in the tropics, mid-latitudes or polar region. The nearest horizontal levels to each altitude are then found and stored.

ERA5 is also resampled using this method, but since these data have a lower horizontal resolution than AIRS, they are not interpolated to a regular distance grid and smoothed to the horizontal resolution of AIRS before they are 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 is 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 leads to improvements in brightness temperature measurements derived from 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 are referred to as IFS 2.

3.1.3 Temperature Divergence of the Resampled Models and AIRS

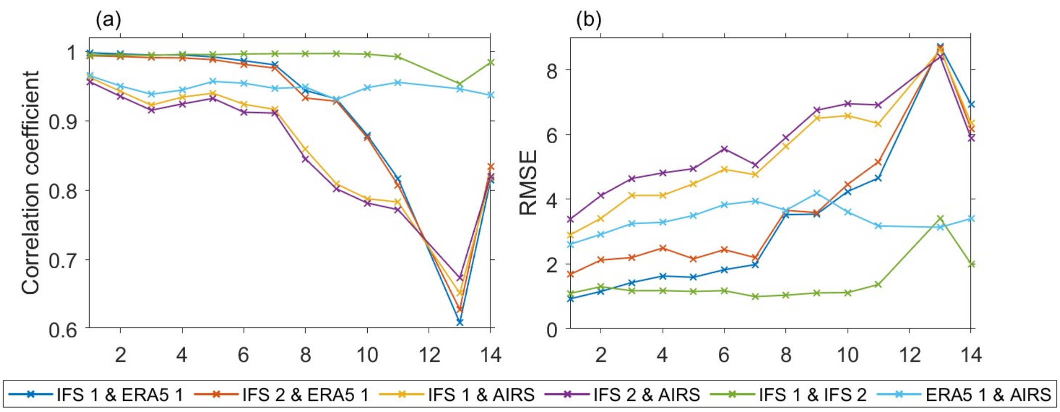

Figure 2. Point-wise correlation coefficients and RMSE (Root-Mean-Square Error) between the temperature at 39 km altitude in AIRS, and the 1 km IFS run and ERA5 resampled as AIRS, for each night during the first 14 days of November 2018, within the region shown in Figure 1. The x-axes show the day in November 2018 (UTC) of the start of each night. Lines between the data points are not used to imply linearity and are shown to make it clearer to see where all the correlation and RMSE points are for each pair of data sets.

Figure 2 shows the point-wise correlation (Figure 2a) and RMSE (Root-Mean-Square Error) (Figure 2b) between the temperature at 39 km altitude in AIRS and the resampled models. 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 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 1 km IFS data. The correlation then increases on the 14th night. As ERA5 assimilates data from observations, including AIRS, the resampled ERA5 data do 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 RMSE (Figure 2b) is greatest between IFS 1 and 2 and AIRS and increases up to night 11. The RMSE also increases between IFS 1 and 2 and ERA5 1 up to night 11. On the 13th night there is a peak in the RMSE 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 RMSE's shown in Figure 2, data are 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 and resampled model data are 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 number of across-track and along-track points remain the same after the data have 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). There is no background removal in the along-track direction as AIRS is travelling meridionally, so there are large differences in temperature in this direction.

Separately, temperature perturbations are 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 are 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 is found by smoothing both data sets using a Gaussian filter with a convolution kernel size of 11 × 11 points and a standard deviation of 7.15 points. This is then subtracted from the temperature data to find the perturbations.

3.3 Adding AIRS Retrieval Noise to the Resampled Models

Since the signal in the AIRS retrieval data cannot 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 are sorted from lowest to highest variance of the temperature perturbations, and checked in this order to find granules without subjectively clearly visible waves.

In previous studies, noise has been added to model data resampled as AIRS to reduce the effect of noise on the comparison. However, there were some issues with the noise addition method used highlighted in these studies. 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 temperature perturbations found for the model resampled as AIRS. This method was also used in Okui et al. (2023), but using one AIRS granule containing no visible waves. However, since the noise added in these studies is uncorrelated pixel-scale noise, 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 lower background amplitudes in the resampled model than in AIRS. They suggest that this is mainly due to lower noise amplitudes in the added noise, and therefore the method used may not be suitable for adding AIRS retrieval noise at global scales. As a result of these problems, a different noise addition method is used in this study, in which the AIRS perturbations are not randomized at each altitude, so the structure of the noise is not lost and the variation of noise with latitude is also taken into account.

30 nighttime AIRS granules containing no subjectively visible waves are selected in total, with 10 granules with >50% of the data points mostly in the tropics (<30° latitude), mid-latitudes (30°–60° latitude) and polar region (>60° latitude), respectively. The boundary lines for these regions are shown in Figure 1. Granules are chosen with over 10% of the data points in the region investigated. This percentage is chosen to allow at least 10 AIRS granules clearly containing only noise during the first 14 days of November 2018 for the mid-latitudes and tropics to be found. As there are not enough nighttime polar AIRS granules containing only noise in the first two weeks of November 2018, granules are chosen during the first 2 weeks of November in years from 2016 to 2020 for the polar region. The AIRS noise should vary mostly meridionally, so the noise granules are selected for the latitude regions described, and any noise granules selected that contain data points mostly outside the region investigated should have little effect on the results. The temperature perturbations of each of the AIRS noise granules (shown in Supplementary Figures S1–S3) are found using the method described above. The data in the arrays of noise temperature perturbations are reversed in the along- and across-track directions separately, and saved so that there are 30 arrays of noise in total for each group, to increase the number that can be selected.

Before adding the noise to the resampled models, the magnitudes of the AIRS noise temperature perturbations are rescaled, due to the gravity wave amplitudes being significantly lower in the resampled models than in the observations. Scaling factors are found by calculating the 95^{th} percentile for each of the resampled model data sets without added noise and dividing these values by the 95^{th} percentile of the AIRS observations in the first 10 days of November 2018. The scaling factors calculated for the 1 km IFS run, resampled using methods 1 and 2 (IFS 1 and IFS 2), are averaged so that noise added to IFS 1 and IFS 2 is scaled by the same factor (0.32) and this does not lead to differences in the results using these methods. A scaling factor of 0.28 is used for ERA5. For each granule of model data resampled as AIRS, a noise temperature perturbation array is chosen randomly from the corresponding group, multiplied by the scaling factor, and added to the resampled model temperature perturbations. If the magnitudes of the noise temperature perturbations are not rescaled before being added to the resampled models, waves with lower amplitudes that may also be observable in the observations, but with higher amplitudes may be too far below the added noise level in the resampled models to be clearly identified.

3.4 2D+1 S-Transform

The S-Transform (ST) is a spectral analysis method 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 ST (Hindley et al., 2016) and the 3D ST(Wright et al., 2017), which are extensions of the 1D ST (Stockwell et al., 1996).

The 2D+1 ST calculates vertical wavelengths using vertical phase shifts between spectral features, which allows it to measure waves more effectively for 3D data with low resolution in one dimension. 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 horizontal data levels and the vertical phase differences between them are calculated (Wright et al., 2021). The vertical wavelengths between the levels can then be found using these phase differences. Unlike the 3D ST, the 2D+1 ST does not quantize vertical wavelengths to Fourier modes, so these wavelengths vary smoothly in the output.

The 2D+1 ST is 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 are calculated using the wave frequencies, in the coordinate frame of the granule, from the 2D+1 ST. The zonal M_x and meridional M_y components of the momentum flux are 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. $\overline{T}$ 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).

N_B is calculated using ERA5 temperature data resampled at each AIRS granule location at the closest 3 hourly time, assuming land surface pressure is equal to 1000 hPa. The values for Nb are then interpolated to the altitudes of AIRS and the resampled models (IFS 1, IFS 2 and ERA5 1). These new values for Nb are used to calculate the momentum flux for all the resampled model datasets and AIRS observations. However, since calculating the N_B values was found to have little effect on the results, the buoyancy frequency is set 0.02 s^{-1} in this study for the models not resampled as AIRS to reduce computational time.

The altitude range selected of the model data resampled as AIRS and of the AIRS data used for the 2D+1 ST analysis is 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 is restricted to select waves with horizontal wavelengths ranging from 60 to 800 km. These values are chosen as 60 is approximately twice the Nyquist frequency and 800 km is below half the track width of each AIRS granule. Data points where the vertical wavelength is below 6 km (twice the vertical spacing) or above 45 km are removed from the data for the wave properties calculated. The 45 km upper limit is selected as this is the approximate height of the stratosphere.

4 Results

4.1 GW Structures and Properties in two Case Studies

Two case studies, in a sub region of the area shown in Figure 1 at 39 km altitude, are presented in order to compare the resampled models and AIRS. These case studies are chosen to show an example with better agreement between the AIRS observations and the resampled IFS (Figures 3 and 4), and an example with worse agreement between the resampled IFS and the observations, further in time from when the IFS was initialised

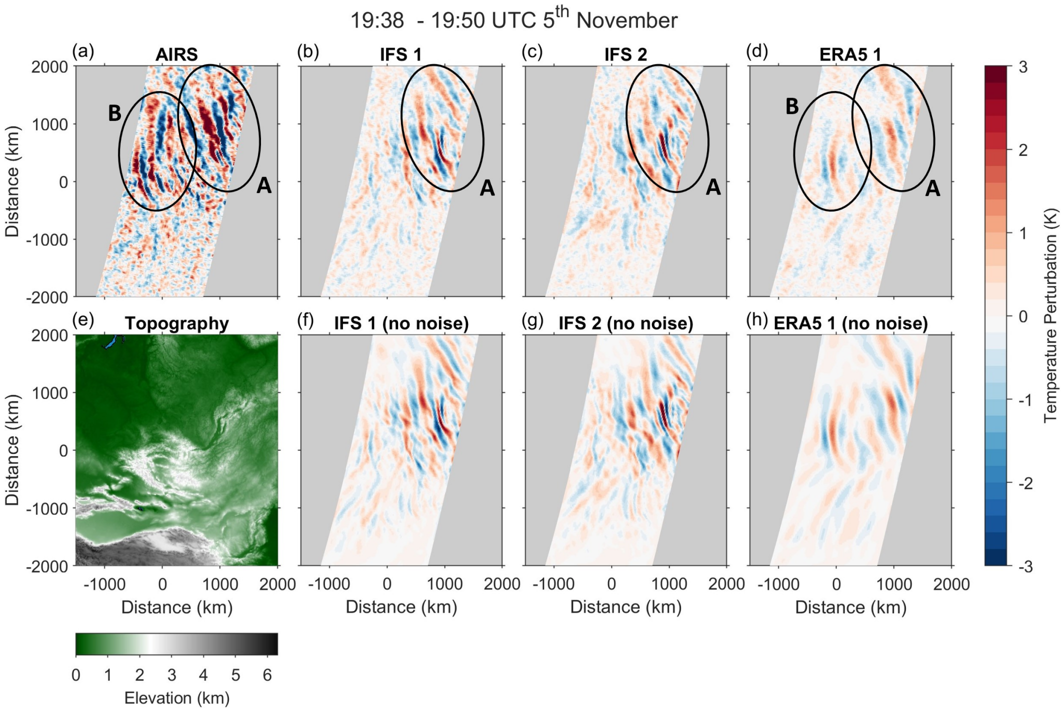

Figure 3. Temperature perturbations at 39 km altitude for granules in a sub region of the area shown in Figure 1, with mean times of 19:38–19:50 UTC for AIRS (a) and at the closest time for the resampled models. Panels (b–d) show the resampled models after adding AIRS noise, with the resampled models before adding noise below them (no noise, panels f–h). The section of the topography from Figure 1 for the sub region is shown in panel (e), with (0,0) at 52° latitude, 94° longitude.

(Figures 5 and 6). The case studies also show examples of the difference in the wave structures in the resampled model data compared to the observations and the temperature perturbations for the resampled models with and without added scaled noise (Figures 3 and 5).

The first case study includes data from AIRS granules with mean times of 19:38–19:50 UTC on the 5th 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 S5.

The large wave on the right, labelled A, is 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, near Lake Baikal (labelled in Figure 1). A curved wave can be seen on the left in AIRS (labelled B, in Figure 3a), which could be convective as it is located close to a region 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, wave (B) can be seen but at a lower amplitude, suggesting the convective source may have been correctly assimilated in the reanalysis or this source could have been correctly simulated in ERA5 due to the atmospheric flow remaining closer to that of the real atmosphere.

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. Discontinuities can be seen in the figure (and Figure 6, which shows the wave properties for the 2nd case study) at the granule edges, as the 2D+1 S-Transform output was found separately for each AIRS or model resampled as AIRS granule. Areas of higher amplitude in Figure 4a–d are seen in similar locations in each data set, but are highest 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). In IFS 1, there is a patch in the zonal and meridional momentum fluxes (Figure 6n and r), in a location where the amplitude is higher, where the direction of these fluxes is reversed compared to the other datasets. This could be a result of a wave in this dataset close to vertical, so the 2D+1 S-Transform is less effective at finding the wave direction.

The second case study is shown for AIRS granules with mean times of 19:14–19:26 UTC on the 9th November 2018, 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, such as near Lake Baikal and the Altai-Sayan mountains (labelled in Figure 1, and not located near 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 9th November, the 1 km IFS run would have diverged further from reality compared to the first case study on the 5th 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 4 and are higher over a larger area than for IFS 1 and 2 (Figure 6b and c) in a similar location to where the AIRS amplitudes are highest (Figure 6a). This is expected as ERA5 assimilates observations, unlike the 1 km IFS run. 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. The horizontal wavelengths in the area with greater amplitudes in ERA5 1 (Figure 6h) appear to also be longer than in AIRS (Figure 6e).

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 generally negative in areas where the amplitude is highest for all the data sets.

4.2 Mean GW Properties in Days 1–10 of November 2018

Figure 7a–d shows the mean nighttime amplitudes during the first 10 days of November 2018 at 39 km altitude 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. These results are also presented Supplementary Figure S6, but without noise added to the resampled models. The mean horizontal wavelengths for nighttime data, with and without scaled AIRS noise added to the resampled models, in the first 10 days of November 2018 are also shown in Supplementary Figures S7 and S9 respectively.

Areas of higher amplitude in Figures 7a–d 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 amplitude are located near to mountain ranges, shown as areas of higher elevation in Figure 1 including the mountains near Lake Baikal, the Altai-Sayan Mountains, the Pamir Mountains and the Urals, suggesting that the GWs have orographic sources. There is an area of higher amplitude over the Urals (labelled in Figure1) in Russia which can be seen in AIRS and the resampled IFS (1 and 2), but can only be seen in ERA5 1 at a significantly lower amplitude. 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 2.72 and 2.81 lower than in AIRS respectively, so the mean of the maximum mean amplitudes for the 1 km IFS run is a factor of 2.77 lower than in AIRS, averaging the results from the two resampling methods. The maximum mean amplitude in ERA5 1 (Figure 7b) is a factor of $\sim$3.59 lower than in AIRS. Supplementary Figure S6 shows similar results for the mean amplitudes with no noise added to the resampled models, but Figure 7b–d have higher background amplitudes due to the added noise. The factors lower than in AIRS for the maximum mean amplitudes are similar without noise added to the resampled models, but slightly lower.

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 of Hindley et al. (2020) and due to background wind related processing, including wind filtering and refraction. 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 maximum mean zonal momentum fluxes for the resampled IFS are a factor of 3.87 and 4.23 lower than in AIRS, for IFS 1 and IFS 2 respectively. In ERA5 1, the maximum mean zonal momentum flux is a factor of 9.74 lower than in the satellite observations. For the maximum mean meridional flux, IFS 1 and 2 are factors of 5.48 and 5.35 lower than in AIRS respectively and ERA5 1 is a factor of 17.6 below AIRS.

4.3 Effect of Adding Randomly Selected Noise to the Resampled Models

Figure 8 shows time series for the mean amplitude and momentum flux on each night in the first 10 days of 2018. The values for the resampled models with no noise added and AIRS observations are shown as crosses and the mean values for AIRS are divided by 2. The circles show the median values of the means from 100 ensemble members where granules of AIRS noise temperature perturbations are selected randomly and scaled before being added to the resampled model data. The shading behind the lines with circles shows the range from the minimum to maximum mean found for each night from the 100 noise ensemble members.

Before finding the mean values, areas of noise in the AIRS and resampled model data are reduced by smoothing the amplitude measurements using a 7 by 7 point boxcar filter and removing points in the original unsmoothed data where the amplitude of the smoothed data is below the 70$^{\mathrm{th}}$ percentile of the amplitude in nighttime data in the first 10 days of November 2018 in each data set ($\sim$0.31 for IFS 1 and 2 (no noise), $\sim$0.28 for ERA5 1 (no noise) and $\sim$1.18 K for AIRS). For the ensemble members, the 70$^{\mathrm{th}}$ percentiles of the amplitude were found for each ensemble member separately. The 70$^{\mathrm{th}}$ percentile is chosen for the amplitude cutoff for each dataset optimise for reducing the areas of noise in the data, while removing as little of the areas with low amplitude GWs as possible. The mean values for amplitude and momentum flux are lower for each night in the resampled models with no noise added than the mean values calculated for the ensemble. The range in mean values from the different ensembles is small for both the amplitude and the momentum flux, suggesting that the addition of different random scaled AIRS noise temperature perturbations has only a small effect on the results.

The 95$^{\mathrm{th}}$ percentiles of the amplitudes and momentum fluxes are found for the resampled models in each ensemble member and for the AIRS observations for comparison. The 95$^{\mathrm{th}}$ percentile is chosen rather than the 100$^{\mathrm{th}}$ percentile to avoid including spikes in the AIRS data. The 95$^{\mathrm{th}}$ percentile of the AIRS observations is divided by the minimum, median and maximum 95$^{\mathrm{th}}$ percentiles of the 100 ensemble members. The factors lower than the 9$^{\mathrm{th}}$ percentile for AIRS amplitudes, of the median 95$^{\mathrm{th}}$ amplitude percentiles of the resampled models are found to be 2.87 for IFS 1 (ranging from 2.84–2.89 from the maximum to minimum 95$^{\mathrm{th}}$ percentile), 2.90 for IFS 2 (ranging from 2.88–2.93), and 3.30 for ERA5 1 (ranging from 3.26–3.3). The factors of the median 95$^{\mathrm{th}}$ percentile for the momentum fluxes found are 10.2 for IFS 1 (with a range of 10.1–10.3), 10.2 for IFS 2 (with a range of 10.0–10.4), and 16.5 for ERA5 1 (with a range of 16.2–16.7) lower than in the observations.

4.4 Distributions of GW Properties in AIRS Observations and Models Before and After Resampling as AIRS

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 at 39 km altitude in Figure 9. In the first column (Figure 9a, 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 9b, d, f and h) shows the KDFs for AIRS and the median from 100 noise ensemble members for the resampled models. The shading behind the lines for the median resampled model KDFs, shows the range from the minimum to the maximum probability for the ensemble members. KDFs were chosen to show the distributions of the data rather than probability density functions (PDFs), as plotting PDFs would involve making assumptions about the data distributions and they could not properly show the distribution of the data.

Areas of noise in the data are reduced using the same method as for the mean amplitude and momentum flux time series (Figure 8), by removing areas in the original data where the smoothed amplitude (using a 7 by 7 point boxcar filter) is below the 70$^{\mathrm{th}}$ percentile for AIRS ($\sim$1.18 K) and the 70$^{\mathrm{th}}$ percentiles found for the resampled models in each ensemble member separately. For the KDFs of ERA5 and the 1 km IFS run before being resampled as AIRS, the same method as for AIRS and the resampled models, was used to remove areas of the data where the smoothed amplitude, using the 7 by 7 point boxcar filter, is below the 70$^{\mathrm{th}}$ percentile of AIRS. The AIRS noise is scaled by the ratio of amplitudes between the observations and the resampled models without added noise (see Section 3.3), which have significantly lower wave amplitudes, so lower amplitude cutoffs can be used for the resampled models to avoid removing areas of low amplitude gravity waves.

After being resampled as AIRS, the horizontal and vertical wavelength spectra of the models are more similar to the AIRS data (Figure 9d and f), compared to the spectra before resampling (Figure 9c 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 AIRS than in the resampled models (Figure 9d), but this could be affected by the higher magnitude noise in the observations. Areas containing only noise with no identifiable wave signals have shorter horizontal wavelengths (shown in Figures 4e–h and 6e–h in areas where the amplitude is low). 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 9c and d, with the highest peaks in ERA5 1. 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 9f). This could be due to the values of the AIRS vertical resolution being overestimated, but this could also be due to the vertical wavelengths in the resampled models could being too short, and the different 70$^{\text{th}}$ percentile amplitude cutoffs used for each data set could also affect these results. For the horizontal wavelengths, the variation in the resampled model distribution is greatest at the peaks, but very low at shorter horizontal wavelengths (Figure 9d). The range for the ensemble distributions is most visible for the vertical wavelengths in the resampled models (Figure 9f), suggesting the added noise has a greater effect on this wave property, although the range is still low and is largest at the peaks.

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 9a). 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$3 K, than in IFS 1 and 2, which have KDFs that decrease to $\sim$6 K and $\sim$5 K respectively (Figure 9b). Whilst the horizontal momentum flux KDFs are similar for AIRS and the 1 km IFS run before resampling (Figure 9g), they are generally higher in AIRS than in the resampled models (Figure 9h), as the momentum flux is proportional to vertical wavelengths and the square of the amplitudes. The results shown in Figure 9b, d and f suggest that there are high amplitude GWs with shorter horizontal or vertical wavelengths in the AIRS retrieval data that are not present in the resampled 1 km IFS run. There is no visible shading showing the range of probabilities for the logarithm to base 10 of the amplitude probabilities in Figure 10b, but these ranges were very small before the amplitude probabilities were logged. The range in probability distributions is also low for the momentum flux.

Supplementary Figure S10 shows the kernel distribution functions, as in Figure 9, but with no scaled AIRS noise added to the resampled models. Areas where the smoothed amplitude is below the 70$^{\text{th}}$ percentile for each resampled model data set, with no added noise, are removed using the same method as for the resampled model data set with added scaled AIRS noise. The differences for the results with and without added noise include that the distributions shift to the left of the panel with no noise added (Figure S10d). This is because areas of noise have lower horizontal wavelengths (see Figures 4e–h and 6e–h). The peaks of the momentum flux distributions are also at lower values without added noise (Figure S10h).

4.5 Point-wise Comparisons of GW Properties

Bivariate histograms, plotted using nighttime data from the first 10 days of November 2018 at 39 km altitude, are shown in Figure 10 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 are also reduced using the method described in Section 4.3. The number of points can be found in Table S1 in the Supplementary Information. 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 70^{th} 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 10. Figure S11 shows the bivariate histograms as in Figure 10, but with no scaled AIRS noise added to the resampled models.

The amplitudes in nighttime AIRS data are significantly higher than in the resampled models (Figure m, q and u), with f_b ranging from 0.952 (AIRS & IFS 1) to 0.977 (AIRS & ERA5 1) (Table S1). Stripes with no data can be seen in the bivariate histograms of the horizontal wavelengths (Figure 10b, f, j, n, r, and v), as the horizontal wavelengths are quantized at longer wavelengths. There are also more points where the AIRS data have a longer vertical wavelength than in the resampled 1 km IFS run (Figure 10o, s and w) with f_b ranging from 0.632 (AIRS & IFS 1) to 0.627 (AIRS & IFS 2). In ERA5, there are more points with longer vertical wavelengths than in AIRS (f_b of 0.471). The histograms also show higher momentum fluxes in AIRS (Figure 10p, t and x) as a result of the higher GW amplitudes and vertical wavelengths. For the momentum flux (Figure 10d, 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 between IFS 1 and 2 (Figure 10d). The data points are also quite spread out for the vertical wavelength plots (Figure 10c, 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 10a–d) and the other wave properties shown. The fraction of points above and below the 1:1 line for IFS 1 & IFS 2 (Figure 10a–d) are similar for all wave properties shown (see Table S1) with the greatest difference in the fractions for the vertical wavelength where f_a is 0.540 and f_b is 0.460.

5 Discussion

The methods used in this study have allowed a more observing-system-aware 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 are 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 means the noise added to the resampled models is better correlated to the location of each resampled model granule compared to previous methods used (e.g. in Hindley et al., 2021; Okui et al., 2023) and could include large noise structures that are present in AIRS noise. However this method of selecting noise granules would be too time consuming for a longer data set, so using machine learning, with a training data set selected by eye, 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 have a low vertical resolution, and relatively low horizontal resolution compared to the 1 km IFS run. 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 are found to be significantly lower in the resampled 1 km IFS run, by a factor of ∼2.77, 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 9), 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 9) and bivariate histograms (Figure 10), 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 added noise is scaled by the amplitude differences (see subsection 3.3 in the Methods section), so lower 70^{th} percentiles can be used as the amplitude cutoffs without removing areas of low amplitude gravity waves.

The two methods used to resample the 1 km IFS lead to quite similar results and work effectively to smooth the model data to AIRS's resolution, resulting in more similar distributions of GW horizontal and vertical wavelengths (Figure 9c–f). However, the peaks of the distributions of vertical wavelength for the resampled models are found to be around 2–3 km lower than for AIRS (Figure 9f). 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 differences in peaks for the vertical wavelength distributions between the resampled models and AIRS are also seen in Supplementary Figure S10, showing that the added noise has little effect on this result.

The results in Figures 7, 9 and 10 are also shown in the Supplementary Information with no noise added to the resampled models in Figures S6, S10 and S11. Comparison of these suggest that while the added noise has some effect on the results this is not large. The largest effect appeared to be on the horizontal wavelength distribution, because lower horizontal wavelengths are calculated in areas of noise. Amplitudes and momentum fluxes are also slightly lower without added noise. An ensemble of added noise with 100 members for each resampled model data set is also analyzed. This analysis shows that the effect of randomly selecting AIRS noise temperature perturbations has little effect on the distributions of the wave properties and on the mean amplitudes and momentum fluxes for each day in the first 14 days of November 2018.

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 a part of 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, followed by 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 measurements with finer vertical and lower horizontal resolution, such as those from limb sounders, and this will be investigated further in future work.

Based on these results, the output from the ERA5 reanalysis is 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 more observing-system-aware 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) is 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 particular 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 9b) which in turn drives a similar difference in momentum fluxes.
3. 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 9), which are thus not seen in the resampled model. Given that the overall amplitude and momentum flux distributions (Figure 9a,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.
4. The effect of adding noise to models resampled as AIRS that is scaled by the differences in amplitudes between the resampled models and observations has also been investigated. The results show that adding this noise does not have a large impact on the results, but this appears to have a larger effect on the results for the horizontal wavelength than for the other wave properties compared. Adding noise leads a higher proportion of shorter horizontal wavelengths in the resampled model data.

Vertical wavelengths in both ERA5 and the 1 km IFS run are found to be significantly shorter than in AIRS observations, even after resampling to match the observational resolution. This difference is typically ∼2–3 km, i.e. approximately 10-20% of the observed wavelengths. This conclusion is difficult to decouple from the effects of noise in the AIRS observations, and further work is needed to address this question more carefully. However, this result is also shown without noise added to the resampled models, suggesting that this added noise has little impact on this result.

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 from `https://datapub.fz-juelich.de/slcs/airs/gravity_waves/data/retrieval/` (Hoffmann, Lars, 2021). The ECMWF ERA5 reanalysis data at 0.25° resolution (Copernicus Climate Change Service, 2023) can be downloaded from the Copernicus Climate Data Store at `https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.bd0915c6?tab=overview`. 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 Lear, Emily (2024).

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 1940 to present* [Dataset]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Retrieved from `{https://doi.org/10.24381/cds.bd0915c6}` ([Accessed 13 Mar 2024]) doi: {10.24381/cds.bd0915c6}

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

Gisinger, S., Polichtchouk, I., Dörnbrack, A., Reichert, R., Kaifler, B., Kaifler, N., … Sandu, I. (2022). Gravity-wave-driven seasonal variability of temperature differences between ecmwf ifs and rayleigh lidar measurements in the lee of the southern andes. *Journal of Geophysical Research: Atmospheres*, *127*(13), e2021JD036270. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JD036270 doi: https://doi.org/10.1029/2021JD036270

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

Hoffmann, Lars. (2021). *AIRS/Aqua Observations of Gravity Waves* [Dataset]. Jülich DATA. Retrieved from {https://doi.org/10.26165/JUELICH-DATA/LQAAJA} ([Accessed 13 Mar 2024]) doi: {10.26165/JUELICH-DATA/LQAAJA}

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

Lear, Emily. (2024). *Code for 'Comparing Gravity Waves in a Kilometre-Scale Run of the IFS to AIRS Satellite Observations and ERA5'* [Code]. GitHub, https://github.com/Emily-Lear/Comparing_the_1_km_IFS_run_to_AIRS_observations/tree/main/Paper.

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

NOAA National Centers for Environmental Information. (2022). *ETOPO 2022 15 Arc-Second Global Relief Model.* NOAA National Centers for Environmental Information (NOAA),https://www.ncei.noaa.gov/products/etopo-global-relief-model. ([Accessed 12 Mar 2023]) doi: {https://doi.org/10.25921/fd45-gt74}

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

Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H. (2019). Intercomparison of gravity waves in global convection-permitting models. *Journal of the Atmospheric Sciences*, *76*(9), 2739 - 2759. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/76/9/jas-d-19-0040.1.xml` doi: 10.1175/JAS-D-19-0040.1

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

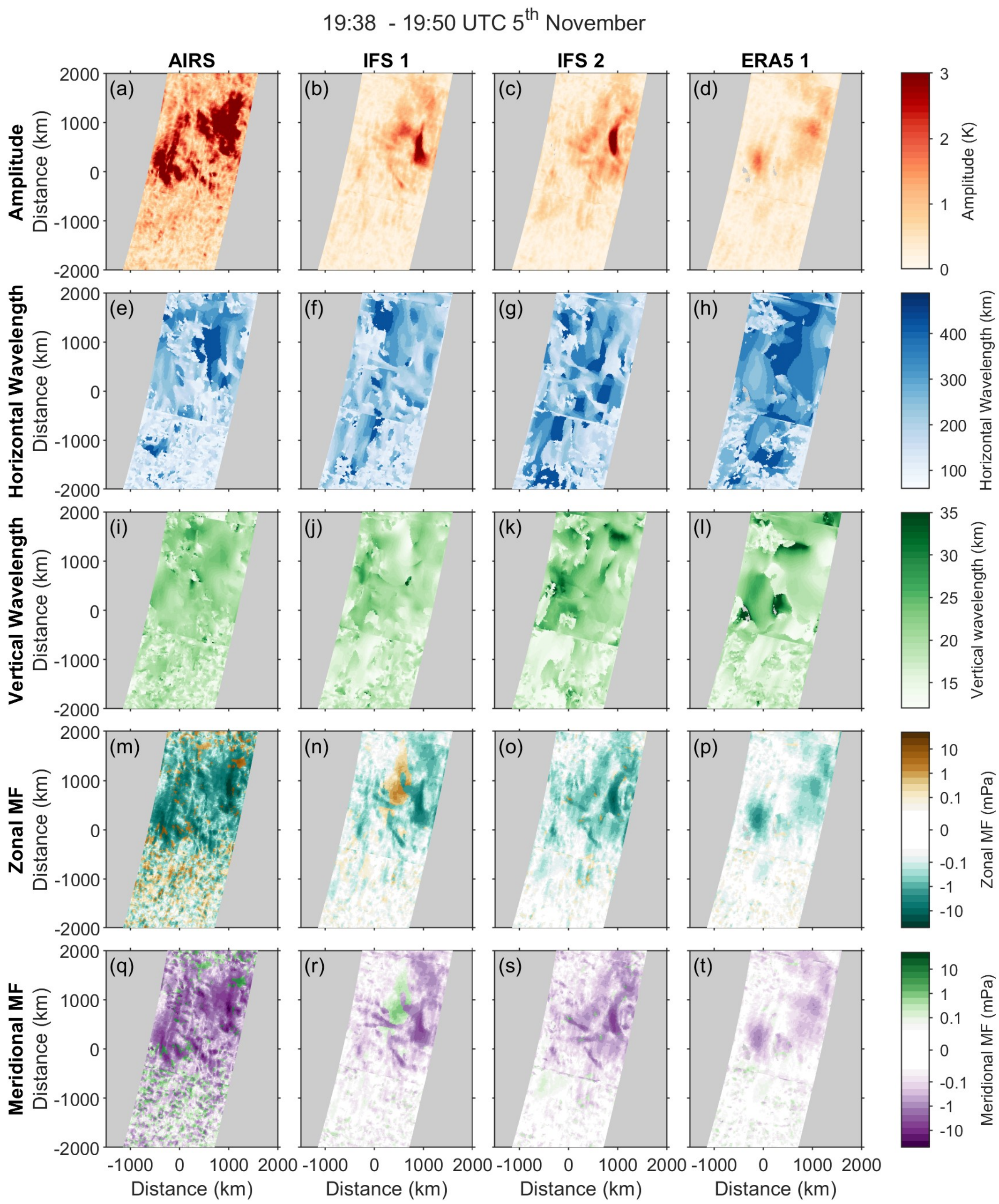

Figure 4. Wave properties at 39 km altitude for the granules shown in Figure 3, including the amplitude, horizontal and vertical wavelengths and the zonal, and meridional momentum flux (MF). The zonal and meridional momentum flux are shown on a log color scale. Data points were removed where the vertical wavelength is below 6 km or above 45 km. Wave properties are only significant in regions where the amplitude is higher for each data set.

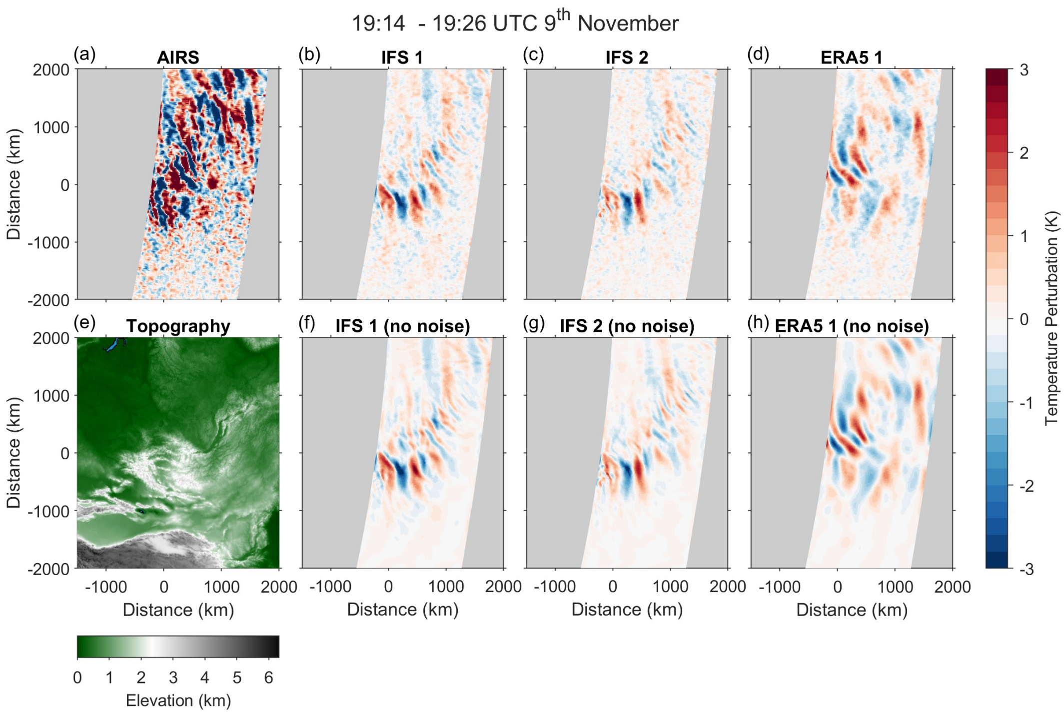

Figure 5. As in Figure 3 at 39 km altitude, but for AIRS granules with mean times from 19:14 to 19:26 UTC on the 9th November 2018 (a) and the resampled models at the closest times (b–d) with and (f–h) without AIRS noise added.

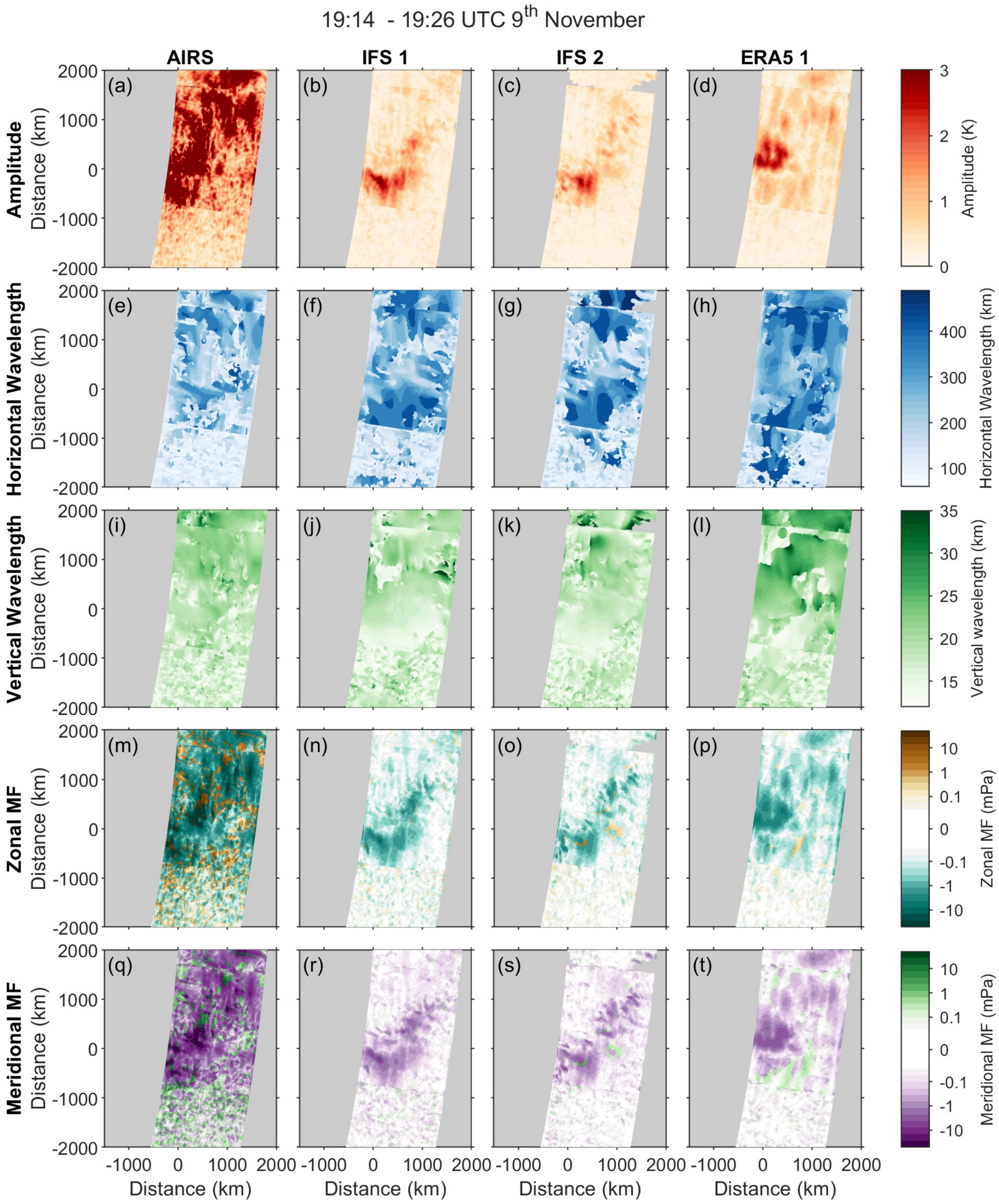

Figure 6. As in Figure 4 at 39 km altitude, but for wave properties for the granules shown in Figure 5. The zonal and meridional momentum flux (MF) are shown on a log color scale.

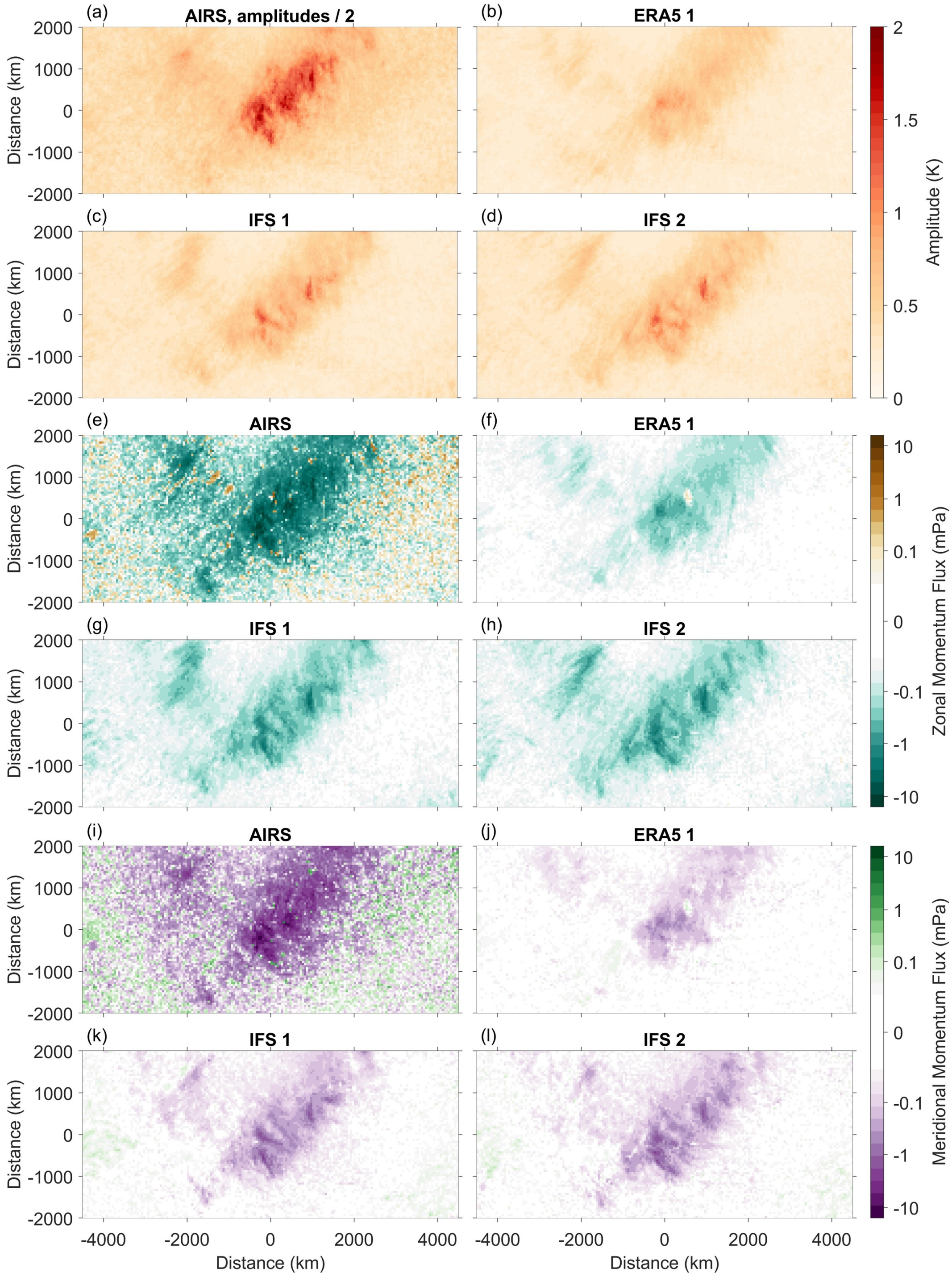

Figure 7. Mean amplitude (a–d) and mean zonal (e–h) and meridional (i–l) momentum flux at 39 km altitude in the region shown in Figure 1 for nighttime data in the first 10 days of November 2018. These are plotted on a regular distance grid with a point spacing of 50 km by 50 km. AIRS amplitudes in (a) are divided by 2. The zonal and meridional momentum flux are shown on a log color scale.

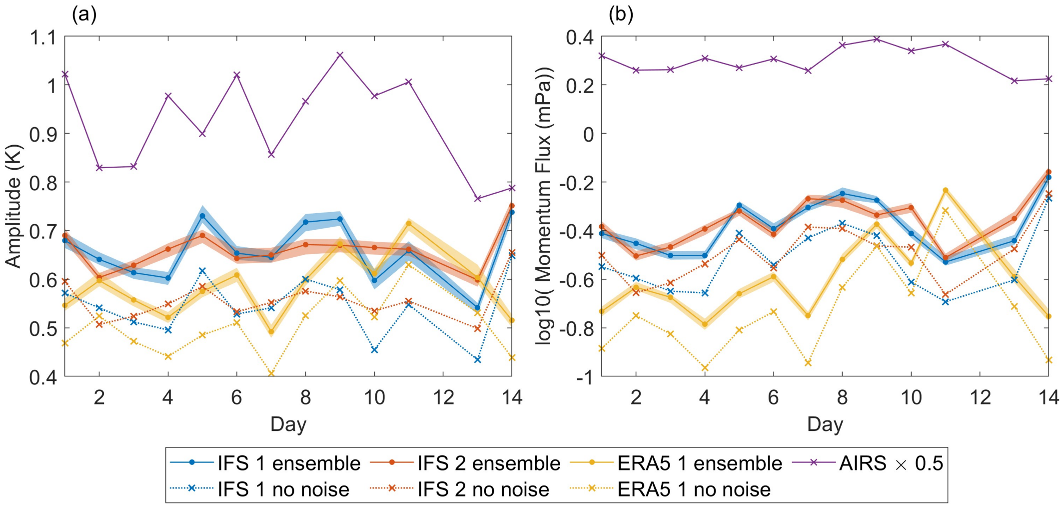

Figure 8. Mean amplitude (a) and the mean of the logarithm to base 10 of the momentum fluxes (b) for each night in the first 10 days of November 2018. The day on the x-axis shows the day of the start of each night in November 2018. Crosses show the mean for AIRS and the resampled models without added scaled AIRS noise. The circles show the median of the means for an ensemble with 100 members where scaled AIRS noise is added to the resampled models and the shading behind the solid lines with cicles shows the range from the minimum to the maximum mean value for each day of the 100 ensemble members. Lines are shown between the data points to make it clearer where all the data points are for each dataset.

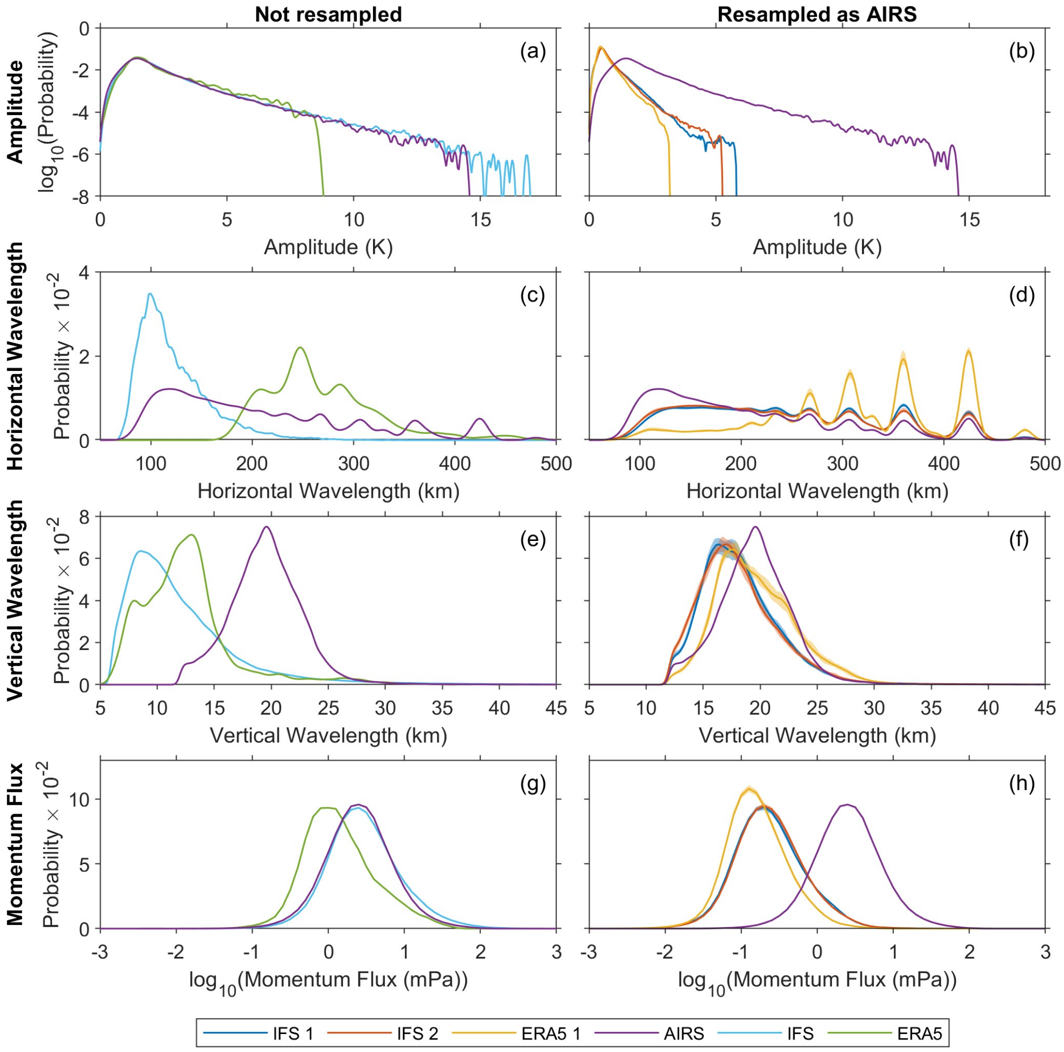

Figure 9. Kernel distribution functions (KDFs) for the wave properties in nighttime data at 39 km altitude for AIRS and the models before resampling (the 1 km IFS run (IFS) and ERA5) (panels a, c, e and g) and for the median distribution from a 100 member added noise ensemble for the resampled 1 km IFS run (IFS 1 and 2) and the median for the added noise ensemble for 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$^{\text{th}}$ 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. The shading behind the lines showing the median distribution for the resampled model ensembles shows the range from the minimum to the maximum probability in the ensemble.

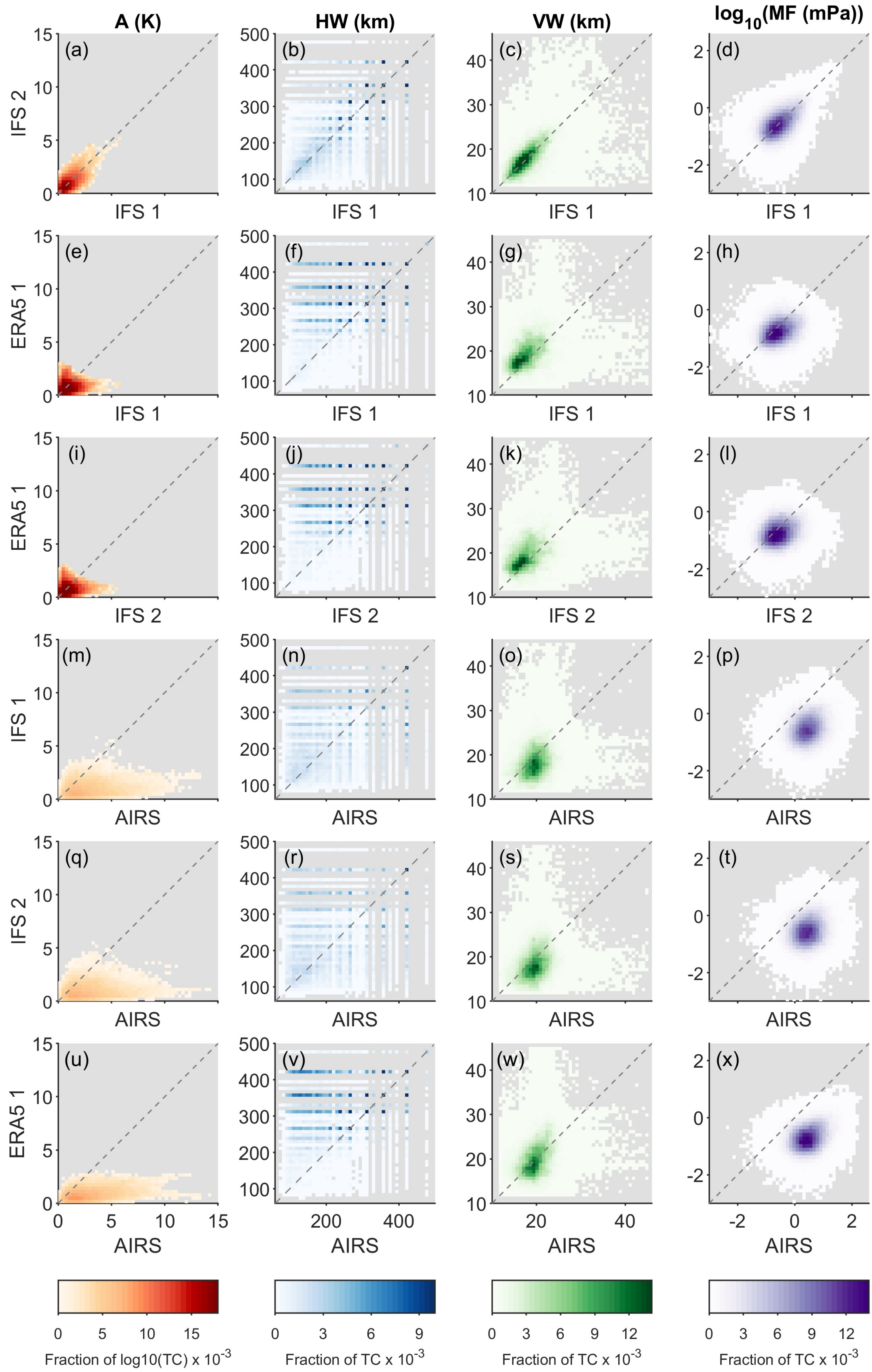

Figure 10. Bivariate histograms of wave properties (amplitude (A), horizontal wavelength (HW), vertical wavelength (VW) and momentum flux (MF)) plotted using nighttime data from the first 10 days of November 2018 in the region shown in Figure 1 at 39 km altitude. The color scales show the fraction of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