

Geophysical Research Letters®



RESEARCH LETTER

10.1029/2025GL115683

Key Points:

- A machine-learning-based gravity wave (GW) detection method has been developed for satellite observations and numerical weather models
- Our method identifies both orographic waves and convective waves, allowing like-for-like comparison between observations and models
- Observation-to-model comparison using this method provides the necessary underpinning for accurately tuning GW model representations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Okui,
okui@eps.s.u-tokyo.ac.jp

Citation:

Okui, H., Wright, C. J., Berthelemy, P. G., Hindley, N. P., Hoffmann, L., & Barnes, A. P. (2025). A convolutional neural network for the detection of gravity waves in satellite observations and numerical simulations. *Geophysical Research Letters*, 52, e2025GL115683. <https://doi.org/10.1029/2025GL115683>

Received 28 FEB 2025

Accepted 30 MAY 2025

Author Contributions:

Conceptualization: Haruka Okui, Corwin J. Wright

Data curation: Haruka Okui, Corwin J. Wright, Peter G. Berthelemy, Neil P. Hindley, Lars Hoffmann

Formal analysis: Haruka Okui

Funding acquisition: Haruka Okui, Corwin J. Wright, Peter G. Berthelemy, Neil P. Hindley

Investigation: Haruka Okui

Methodology: Haruka Okui, Corwin J. Wright, Peter G. Berthelemy, Neil P. Hindley, Andrew P. Barnes

Project administration: Corwin J. Wright

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A Convolutional Neural Network for the Detection of Gravity Waves in Satellite Observations and Numerical Simulations

Haruka Okui^{1,2} , Corwin J. Wright¹ , Peter G. Berthelemy¹, Neil P. Hindley¹ , Lars Hoffmann³ , and Andrew P. Barnes⁴ 

¹Centre for Climate Adaptation and Environment Research, University of Bath, Bath, UK, ²Now at Department of Earth and Planetary Science, The University of Tokyo, Tokyo, Japan, ³Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany, ⁴Department of Computer Science, University of Bath, Bath, UK

Abstract Comparisons between observed and model-resolved gravity waves (GWs) are crucial for evaluating general circulation model (GCM) simulation accuracy and understanding wave characteristics. However, observational noise often obscures waves, complicating such comparisons. To address this, we have developed a GW detection method using a convolutional neural network (CNN). The CNN is trained on Atmospheric Infrared Sounder (AIRS) temperatures with labels indicating wave presence based on Berthelemy et al. (2025, <https://doi.org/10.5194/egusphere-2025-455>). Their method detects noise-induced pixel-to-pixel variations in horizontal wavelengths; in contrast, the CNN robustly identify waves even when applied to smoothly varying model data. Using this method, we compare stratospheric GWs in boreal winters between AIRS observations and a high-top GW-permitting GCM, Japanese Atmospheric GCM for Upper Atmosphere Research (JAGUAR). The results agree well and exhibit similar interannual variability, with discrepancies also identified, including a more zonally elongated distribution of tropical GWs in JAGUAR. This method is broadly applicable to the future use of satellites for guiding wave-resolving atmospheric model development.

Plain Language Summary Gravity waves (GWs), small-scale waves driven by buoyancy, play an important role in atmospheric dynamics by transporting momentum throughout the atmosphere. Two key approaches to study the global characteristics of GWs are high-resolution satellite observations and numerical simulations using climate models. However, observational noise can make it difficult to quantitatively compare these two data sets. To reduce the impact of noise, we adopt a machine-learning approach. We train a convolutional neural network (CNN) using satellite temperature measurements produced with the spectral method of Berthelemy et al. (2025, <https://doi.org/10.5194/egusphere-2025-455>), where pixel-to-pixel changes in horizontal wavelengths due to noise are used to identify waves. Once trained, we apply the CNN to evaluate stratospheric GWs simulated by a high-resolution climate model with those observed by a satellite instrument. The observed and simulated waves show good agreement, both exhibiting similar year-to-year variations. The CNN enables fair and computationally efficient validation of climate models based on satellite observation.

1. Introduction

Atmospheric gravity waves (GWs) play a range of crucial roles in the climate system (Achatz et al., 2024; Fritts, 1984; Fritts & Alexander, 2003). They are one of the key factors controlling the temperature, wind, and chemistry of the middle atmosphere, which also influence the troposphere. However, due to their small spatial and temporal scales, accurate representation of GWs remains challenging for weather and climate models.

In recent years, an increasing number of general circulation models (GCMs) have become available which explicitly resolve GWs. ERA5 has been widely used for studying relatively large-scale GWs (Gupta, Sheshadri, Alexander, & Birner, 2024; Pahlavan et al., 2023; Podglajen et al., 2020; Yoshida et al., 2024), while kilometer-scale model runs have been performed using models such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (Gupta, Sheshadri, & Anantharaj, 2024; Lear et al., 2024; Rhode et al., 2024) and the Goddard Earth Observing System (GEOS-5) atmospheric GCM (Holt et al., 2017). High-top GCMs covering the entire middle atmosphere have also been incorporated into GW research, including the High Altitude Mechanistic general Circulation Model (HIAMCM, Becker et al., 2022) and the Japanese Atmospheric GCM for Upper Atmosphere Research (JAGUAR, Watanabe & Miyahara, 2009, see also Okui

Resources: Haruka Okui, Corwin J. Wright, Peter G. Berthelemy, Lars Hoffmann
Software: Haruka Okui, Corwin J. Wright, Peter G. Berthelemy, Neil P. Hindley
Supervision: Corwin J. Wright
Validation: Haruka Okui
Visualization: Haruka Okui
Writing – original draft: Haruka Okui
Writing – review & editing: Corwin J. Wright, Peter G. Berthelemy, Neil P. Hindley, Lars Hoffmann, Andrew P. Barnes

et al. (2021)). Despite improving skills of these high-resolution GCMs, uncertainties persist in processes such as convection, radiation, and turbulence, which are linked to GW generation and dissipation. In long-running climate simulations, GW parameterizations also remain essential. For a realistic model climatology, these parameters still require tuning. However, since GW parameterizations are often implemented to alleviate model biases (e.g., Kim et al., 2003), and wave forcings frequently compensate for one another (e.g., Cohen et al., 2013), GWs in “best-tuned” models do not necessarily match observed GWs. Therefore, comparing observed and model-simulated GWs provides valuable insight into both GW representation and overall model accuracy.

Satellite observations are a primary tool for understanding the global distribution of GW parameters (e.g., Alexander et al., 2010; Geller et al., 2013). Their extensive spatial coverage enables the global derivation of key indicators for evaluating and improving GW representation in GCMs. For instance, Ern et al. (2018) examined a GW momentum flux climatology created using atmospheric infrared emissions observed by two limb-sounding satellite instruments, while Hindley et al. (2020) presented a climatology of stratospheric GWs based on three-dimensional (3-D) temperature retrievals from the nadir-sounding Atmospheric Infrared Sounder (AIRS).

Despite their importance for model validation, the characteristics of GWs derived from observations are strongly dependent on instrument characteristics (Alexander, 1998; Wright et al., 2016). Limb-sounding satellite instruments provide high vertical but low horizontal resolution, while nadir-sounding instruments offer low vertical but high horizontal resolution. This has important implications for quantifying the role of GWs in the atmospheric system. Long-vertical waves measured by nadir-sensing instruments tend to be more important to vertical momentum transport, while the short-vertical waves measured by limb-sensing instruments can travel significantly longer horizontal distances. To ensure fair comparisons between observed and model-simulated GWs (hereafter referred to as observation-to-model comparisons), observational filters matching instrument resolutions must be applied to model data.

Additionally, and importantly, instrument noise can have a large influence. For AIRS temperature retrievals, measurement noise ranges from 1.4 to 2.1 K at altitudes between 20 and 60 km (Hoffmann & Alexander, 2009). Given that stratospheric GWs typically exhibit temperature amplitudes of a few to several Kelvins, this level of noise can blur a substantial portion of wave signal. Retrieval errors arise from multiple sources, including instrument radiometric noise, deviations from retrieval constraints, and the effects of non-local thermal equilibrium (non-LTE) in daytime data (Hoffmann & Alexander, 2009). Hindley et al. (2019) showed latitudinal and seasonal variations of AIRS noise. Such noise variabilities complicate GW analysis based on satellite observations.

Accordingly, a key challenge in observation-to-model comparisons is reducing the influence of observational noise. One approach is to add randomized observational noise directly to model temperature fields before comparison (Hindley et al., 2021), however, Okui et al. (2023) found that even after adding randomized AIRS-like noise, model data could still exhibit a lower noise level than the observations. This is likely because only uncorrelated pixel-scale noise was included in the added noise. Lear et al. (2024) improved this by manually selecting AIRS measurements with no visible waves from nighttime measurements during the same period of the year and latitude regions, preserving structural noise, and adding them to model data. Although their method could represent noise more realistically, small-amplitude GWs were also unintentionally removed by post-processing using the commonly used amplitude-based noise removal method (hereafter referred to as the “amplitude method”). They also suggested that machine learning could streamline the time-consuming process of selecting measurements containing noise only.

Due to the random nature of instrument noise, spurious small-scale wavelengths have random orientations. Therefore, the presence of a near-uniform wavelength distribution can act as a way to discriminate between waves and noise. Berthelemy et al. (2025) used this insight to develop a novel wave-detection method to identify regions where wavelength estimates produced using a 2-D S-Transform (ST, Stockwell et al., 1996; see also Hindley et al. (2016)) show only small spatial variations. This approach, referred to as the “neighborhood method” due to its analysis of local wavelength distributions, was shown to detect waves in AIRS temperature measurements more accurately than the amplitude method. However, this method cannot be applied to model fields because they contain no noise. This limits its usefulness for observation-to-model comparisons. Furthermore, the method can be computationally expensive, as it requires spectral analysis using the 2-D ST.

Here, we present a universal detection method applicable for both model-simulated GWs and those in satellite observations using a convolutional neural network (CNN) for semantic segmentation, based upon the U-net

architecture. This builds on the previous work of Coney et al. (2023), who trained a U-net CNN to identify trapped lee waves over Britain and Ireland in Met Office operational forecasts using hand-labeled wave masks. Instead of hand-labeled regions, wave regions found by the neighborhood method of Berthelemy et al. (2025) are used as a training data set for our CNN. In this study, we successfully apply the CNN to both observations and model temperature fields and perform a global observation-to-model comparison of stratospheric GWs.

The data sets used in this study are described in Section 2. The CNN and analysis methods are detailed in Section 3. Section 4 presents the results from the detection and comparisons between AIRS observations and model simulations. In Section 5, we summarize the findings and discuss potential applications of the CNN-based GW detection.

2. Data

2.1. 3-D Temperature Measurements Obtained by AIRS

AIRS (Chahine et al., 2006) is a nadir-sounding hyperspectral radiometer aboard NASA's Aqua satellite. It operates in a near-polar, sun-synchronous orbit with an approximate orbital period of 100 min. Continuously scanning $\pm 49.5^\circ$ from the nadir in the across-track direction, it measures radiances in 2,378 infrared spectral channels. Vertical profiles of parameters including temperature are obtained over an $\sim 1,800$ -km (90-footprint) wide swath. The across- and along-track spacings vary from $\sim 13.5 \times 13.5$ km at nadir to $\sim 41 \times 21.4$ km at the swath edges. AIRS measurements are divided into 240 segments per day along the satellite track, with each segment corresponding to 6 min of observations, referred to as a “granule.”

In this study, we use 3-D temperature measurements derived from the retrieval scheme proposed by Hoffmann and Alexander (2009). This method utilizes two channels within a CO_2 -sensitive spectral range to generate temperature profiles with a 3-km vertical spacing. The effective vertical resolution ranges from 7 to 14 km within the altitude range of $z = 20$ –60 km, depending on both altitude and latitude (Hindley et al., 2019). Nighttime retrievals exhibit higher vertical resolution and lower noise levels compared to daytime retrievals, owing to reduced effects of non-LTE.

2.2. High-Resolution JAGUAR Resampled as AIRS

The high-resolution version of the JAGUAR model (Watanabe & Miyahara, 2009) is a global spectral model with a triangularly truncated spectral resolution of 639, corresponding to a minimum resolvable horizontal wavelength of ~ 60 km. The vertical resolution is 300 m in the middle atmosphere, and the model top is set at $z = 150$ km. No sponge layers are employed in the uppermost part of the model. Instead, the horizontal diffusion coefficient gradually increases above the mesosphere. GW parameterizations are not used. Given these characteristics, the model is considered suitable for GW analysis throughout the entire neutral atmosphere, from the surface up to approximately $z = 110$ km.

Okui et al. (2021, 2022) performed hindcast simulations with the model by nudging its spectral components with horizontal wavenumbers higher than 21 to the JAGUAR-Data Assimilation System (JAGUAR-DAS) Whole neutral Atmosphere ReAnalysis (JAWARA, Koshin et al., 2025; Sato, 2025). The spectral nudging was first conducted for 3 days, followed by a 4-day free-run simulation. This 7-day cycle was repeated every 4 days. In the present study, only outputs from the free-runs are analyzed.

We compare GWs in the JAGUAR model with those observed by AIRS, using hindcast simulation outputs from 15 December to 28 February (hereafter DJF) in 2016/2017, 2017/2018, and 2018/2019. Following the approach of Okui et al. (2023), we first interpolate the model data onto a 1 km-spacing height grid by treating geopotential height as an approximation of geometric height, which is appropriate in the stratosphere. JAGUAR temperature fields are then sampled at the locations of the AIRS footprints. A vertical low-pass filter that depends on altitude, latitude, and day-night conditions is applied to simulate the AIRS vertical resolution before the fields are sampled at vertical levels with 3-km spacing to match the AIRS retrieval sampling. This processed JAGUAR temperature data set is referred to as “JAGUAR-as-AIRS.” The subsequent analysis steps are common to both AIRS and JAGUAR-as-AIRS data, as described in Section 3.1.

3. Methods

3.1. Data Preprocessing

Before fitting the CNN (see Section 3.2) and performing spectral analysis (see Section 3.3), each 90×135 element “granule” (across-track $\times$ along-track) of AIRS measurements and JAGUAR-as-AIRS is interpolated onto a regular grid oriented along the satellite track, with a horizontal spacing of 20×20 km. The regridded granules are then detrended by removing fourth-order polynomial fits applied along the across-track direction in order to exclude large-scale structures such as background temperature gradients and Rossby waves. The sensitivity of the detrended AIRS temperature perturbations to different wavelengths is shown in Figure 2c of Hindley et al. (2019).

3.2. Convolutional Neural Network for Semantic Segmentation

The CNN for semantic segmentation developed in this study employs the U-Net architecture (Ronneberger et al., 2015), which is widely used for segmentation tasks. The layer structure of the CNN is depicted in Figure S1 in Supporting Information S1. The CNN consists of seven blocks of layers distributed across four hierarchical levels. The kernel size in the convolution layers is 3×3 . To introduce nonlinearity and constrain the output range, a rectified linear unit activation function is applied after each convolution layer. In the encoder path, down-sampling is performed through 2×2 maximum pooling, while the decoder path uses transposed convolution with 2×2 kernels for upsampling.

Accuracy, that is, the proportion of correct classifications, is used as the evaluation metric, and binary cross entropy is adopted as the loss function. Due to the significant class imbalance in the training data set, with the number of pixels without waves (N_{nowave}) far exceeding that of pixels containing waves (N_{wave}), their contributions to the loss function are weighted by $0.5N_{\text{nowave}}^{-1}$ and $0.5N_{\text{wave}}^{-1}$, respectively (King & Zeng, 2001). In the output layer, a sigmoid activation function converts the output of each neuron into a probability ranging from 0 (no waves) to 1 (waves). In the result section, pixels with probabilities above a threshold of 0.5 are classified as “wave regions.”

For training the CNN, we use wave masks calculated from regridded and detrended granules of AIRS retrieved temperatures at $z = 39$ km using the neighborhood method of Berthelemy et al. (2025). The training data set used in this paper covers December 2013–February 2014 and June–August 2014, periods of seasonally intense GW activity. As a sensitivity test, we also trained the CNN using annual data from December 2013 to November 2014 and found no significant improvement in wave detection performance. A total of 43,440 granules are used, with 80% allocated to the training set and 20% to the test set, following the holdout method (Kohavi, 1995).

To standardize the input temperature fields, the Z-score normalization is applied. In the observational data, noise also contributes to the standard deviation σ_o of temperature perturbations. Thus, directly dividing by σ_o results in amplitudes too small for the CNN to detect wave-like structures effectively. Instead, input granules are normalized by $0.5 \sigma_o$, where σ_o is estimated from 240 granules for each day. We have confirmed that normalizing nighttime and daytime data separately does not yield significant differences in wave detection results.

Temperature fields of a numerical model such as JAGUAR, in contrast, do not contain the same level of noise as observations. The variance in observations σ_o^2 is the sum of variances due to waves and noise, whereas JAGUAR-as-AIRS theoretically contains only variance due to waves. To ensure consistent normalization between the observational and model data sets, we normalize the perturbations in JAGUAR-as-AIRS data using the following procedure. First, we estimate σ_o^2 from AIRS observations separately for wave and non-wave regions classified by the neighborhood method. The variance for non-wave regions, attributed almost entirely to noise, is 83.8% of that in wave regions. This result suggests that waves account for 16.2% of the total variance in wave regions in the AIRS observations. Then, we normalize JAGUAR-as-AIRS temperatures by $0.5\sigma_m/(0.162)^{0.5}$, where the standard deviation σ_m is estimated from 240 JAGUAR-as-AIRS granules for each day. This normalization step assumes that model-simulated wave amplitudes are comparable to those in the observations. When this assumption holds true, the normalization ensures that wave amplitudes in both data sets fall within a similar range, allowing the CNN to effectively focus on relevant features.

3.3. 2D + 1 S-Transform

To estimate amplitudes of detected waves, we perform spectral analysis using the 2D + 1 S-Transform (ST) described by Wright et al. (2021). The 2D + 1 ST is based on the multi-dimensional ST developed by Hindley et al. (2019), which provides localized wave amplitudes and 3-D wavelengths at every location in the input data. Specifically, the 2D + 1 ST method applies a horizontal 2-D ST for each layer of the data, then computes phase differences between groups of adjacent layers to obtain a better estimate of vertical wavelengths, which is ideal for AIRS measurements with a relatively short vertical extent ($z = 20\text{--}60$ km) compared to their horizontal extent ($\sim 2,000$ km).

Here, we select $z = 39$ km as the altitude of temperature perturbations for which we perform CNN-based wave detection and spectral analysis. This altitude corresponds approximately to the center of the altitude range where the AIRS vertical resolution is still high, and noise levels are relatively low (Hindley et al., 2019).

4. Results and Discussion

4.1. Case Studies

Figure 1 shows maps of temperature perturbations and CNN-detected waves (i.e., temperature perturbations masked in regions with probabilities of wave occurrence below 0.5) in AIRS and JAGUAR-as-AIRS data for three sample cases. Owing to the strong class imbalance and inapplicability of the neighborhood method to JAGUAR-as-AIRS, CNN performance cannot be reliably evaluated using statistical metrics (c.f., Figure S2 in Supporting Information S1). Instead, we focus here on visually demonstrating wave detection results.

In Case 1 (Figures 1a–1d), featuring waves over the Scandinavian Peninsula on 20 January 2017, two distinct wave packets are detected: one over the peninsula and another over the Norwegian Sea. The wave packet over the peninsula exhibits larger amplitudes, reaching a maximum of ~ 12 K, and exhibits slightly longer horizontal wavelengths ($\sim 300\text{--}500$ km) than the one over the ocean. These waves are likely orographic waves generated by flows over the Scandinavian Mountains. Similar wave packets also appear in JAGUAR-as-AIRS, with wavelengths comparable to those in the observations. On the other hand, wave-like structures that are not present in the AIRS measurements are also observed in JAGUAR-as-AIRS. In both data sets, the CNN successfully detects these waves while appropriately excluding Southern Europe, where wave-like structures are absent.

Case 2 features waves observed on 14 January 2018 (Figures 1e–1h) around Tropical Cyclone Berguitta, moving through the Western Indian Ocean. To the east of the cyclone's center, marked by a cross in the center of the panels, the CNN detects a small-scale wave packet in both the AIRS and JAGUAR-as-AIRS data. On the opposite side, a larger-scale wave with a phase structure resembling a plane wave dominates. Given its upward propagation, this wave has a westward intrinsic phase velocity. The AIRS measurements capture this wave more distinctly than JAGUAR-as-AIRS.

Case 3 shows temperature perturbations around Tropical Cyclone Cebile on 2 February 2018 (Figures 1i–1l). A nearly concentric wave is observed and detected by the CNN east of the cyclone's center, with a horizontal wavelength of ~ 200 km in both data sets. Additionally, a wave packet with longer south-eastward wavelengths appears south of the cyclone. This wave packet, which is more pronounced in JAGUAR-as-AIRS, is also successfully detected by the CNN.

Across all cases, the CNN demonstrates a capability for GW detection. Wavelengths, phase structures, and noise levels in temperature perturbations appear to have only a limited impact on detection accuracy, particularly for GWs with large amplitudes. For details on amplitudes and horizontal wavelengths of detected waves, see also Figure S3 in Supporting Information S1.

4.2. Global Maps

Figures 2a–2h show the global distribution of averaged CNN outputs, which can be interpreted as probabilities for wave occurrence. On average, GWs are more likely to be detected in JAGUAR-as-AIRS, which exhibits significantly higher probabilities than the AIRS results at Northern Hemisphere (NH) midlatitudes ($5^{\circ}\text{N--}45^{\circ}\text{N}$) during DJF 2018/2019. This difference is also evident, though less pronounced, in 2016/2017 and in the 3-year DJF average (Figures 2g and 2h).

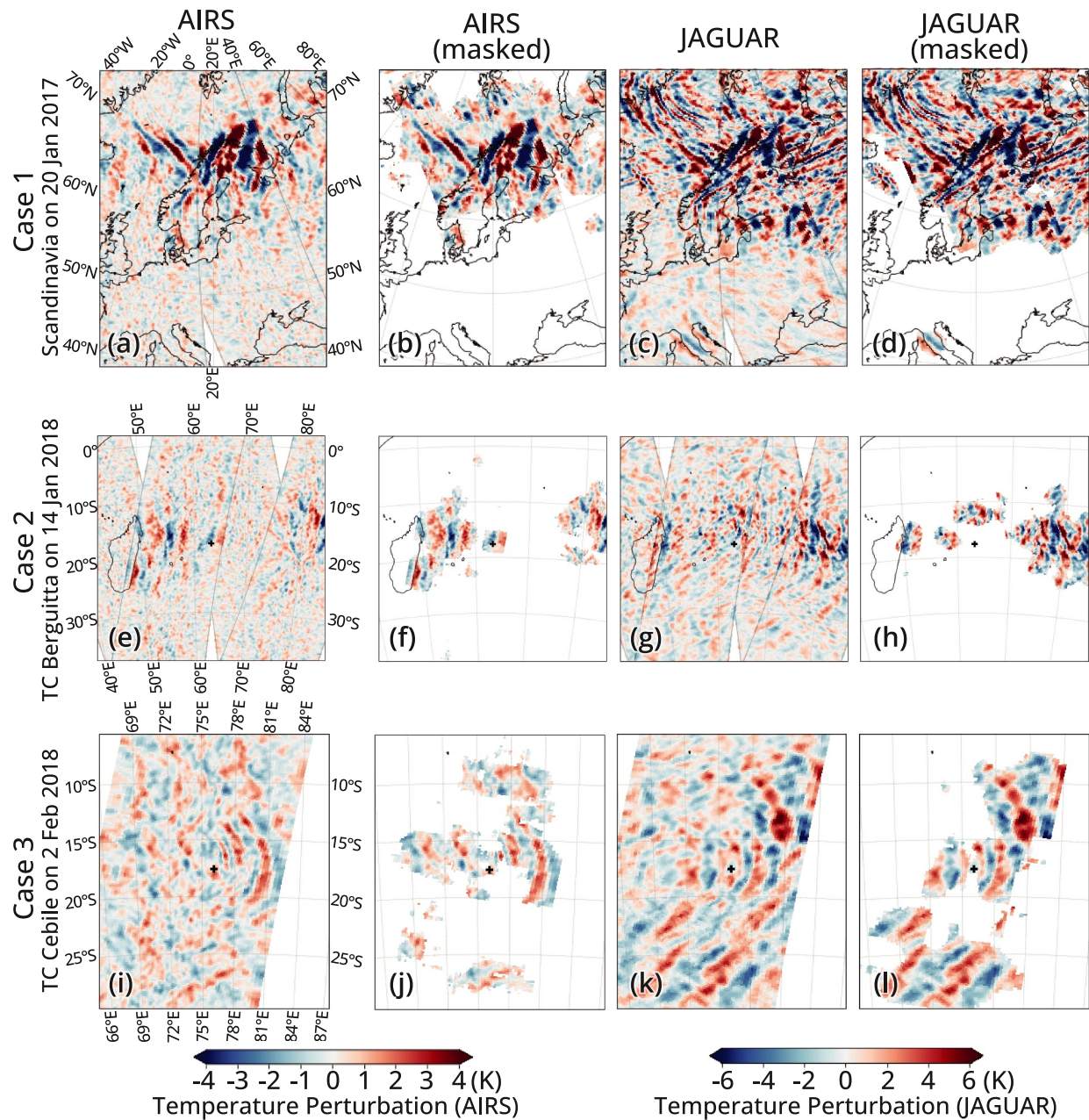


Figure 1. Maps of temperature perturbations (left) and those with higher probability than 0.5 (right), which are detected gravity waves, at $z = 39$ km. Panels (a, b, e, f, i, and l) (c, d, g, h, k, and l) show results for the Atmospheric Infrared Sounder (AIRS) (Japanese Atmospheric GCM for Upper Atmosphere Research [JAGUAR]-as-AIRS) data. Cases 1, 2, and 3 shown in panels (a–d, e–h, and i–l) feature waves over the Scandinavian Peninsula on 20 January 2017, Tropical Cyclone Berguitta on 14 January 2018, and Tropical Cyclone Cebile on 2 February 2018, respectively. Cross marks in the center of Figures 2e–2l indicate the centers of the tropical cyclones. Note that the color bars for the AIRS and JAGUAR-as-AIRS data differ, allowing both detected and excluded perturbations to be clearly shown in each data set.

During DJF 2016/2017 (Figures 2a and 2b) and 2017/2018 (Figures 2c and 2d), high probabilities are concentrated along the stratospheric eastward jet over the northernmost Atlantic Ocean. In contrast, during DJF 2018/2019 (Figures 2e and 2f), probabilities at mid- and high latitudes in the NH are significantly lower than in the other 2 years. A strong sudden stratospheric warming occurred from late December to January in this season (Rao et al., 2019), leading to a significant reduction in GW activity in the NH stratosphere (Okui et al., 2023).

In the Southern Hemisphere (SH) subtropics, specifically at 10°S–30°S, high probabilities are observed around the east coasts of South America, Africa, and the Maritime Continent (Figures 2g and 2h). These peaks align with

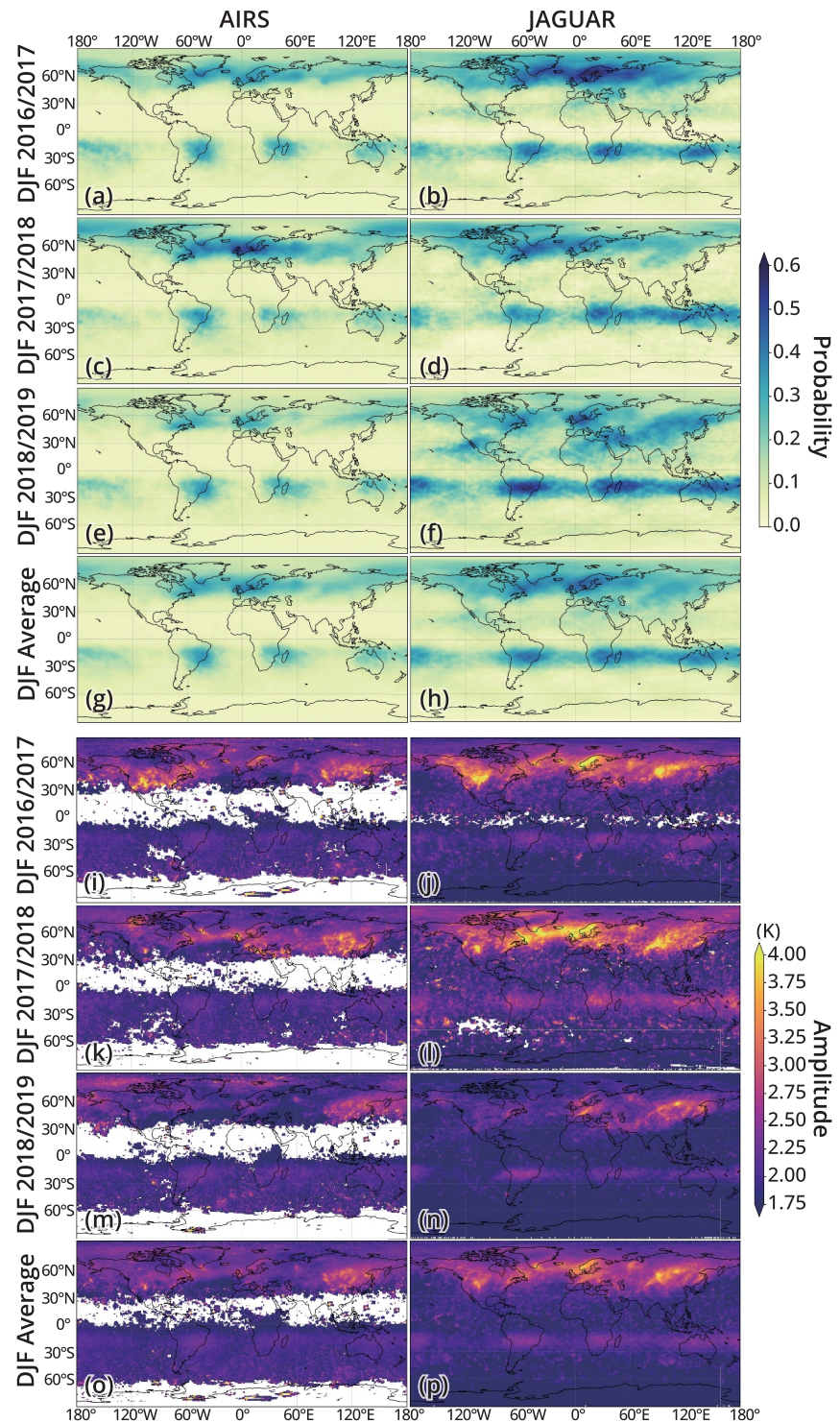


Figure 2. (a–h) Probabilities of wave occurrence and (i–p) amplitudes of detected waves at $z = 39$ km averaged over DJF 2016/2017, 2017/2018, 2018/2019, and these three DJFs. The left-hand panels show the results from Atmospheric Infrared Sounder (AIRS) observations, and the right-hand panels show the results from Japanese Atmospheric GCM for Upper Atmosphere Research-as-AIRS.

regions of heavy precipitation, including the South Atlantic convergence zone (SACZ, e.g., Carvalho et al., 2004), South Indian convergence zone (SICZ, e.g., Cook, 2000), and South Pacific convergence zone (SPCZ, e.g., Vincent, 1994). The SACZ and SICZ have a northwest-southeast orientation, and the SPCZ extends south-eastward over the South Pacific. Interestingly, AIRS wave probabilities align more clearly with these orientations of the convergence zones (Figure 2g), whereas in the JAGUAR-as-AIRS data, the distribution appears more zonally uniform. Additionally, GWs in the model data are more widespread along the intertropical convergence zone as shown in Figure 2h. These differences may stem from potential biases in the model representation of convection and convective GWs. Another possible explanation is that model GWs may propagate over longer horizontal distances and converge more strongly toward the core of the stratospheric summer jet than in the real atmosphere. GWs in JAGUAR-as-AIRS data exhibit longer horizontal wavelengths (as shown in Figure S3b in Supporting Information S1). Such waves typically have higher horizontal-to-vertical ratios of their group velocities, which may contribute to the broader zonal distribution of GWs along the summer jet.

Figures 2i–2p shows the global distribution of the averaged amplitude of CNN-detected GWs. In general, the geographical distributions of GW activity are similar between AIRS and JAGUAR-as-AIRS. The results show amplitude peaks consistently in the SH subtropics and Eastern Eurasia from 2016/2017 to 2018/2019 (Figures 2i–2n), the Rocky Mountains, southern Greenland, and the Scandinavian Peninsula in 2016/2017 (Figures 2i and 2j), and the North Atlantic Ocean in 2017/2018 (Figures 2k and 2l). In both data sets, amplitudes in the NH during DJF 2018/2019 (Figures 2m and 2n) are significantly lower compared to those during DJF 2016/2017 (Figures 2i and 2j) and 2017/2018 (Figures 2k and 2l).

Discrepancies are also identified. Regions with missing values correspond to areas where no waves are detected (i.e., the probability never exceeds 0.5) through the analysis period. At NH low latitudes, higher probabilities and smaller non-wave regions are observed in JAGUAR-as-AIRS than the AIRS measurements (Figures 2a–2h). In such regions, amplitudes are generally smaller than ~ 2 K. A possible explanation for this difference is a remaining influence of noise, in which waves with amplitudes comparable to the noise level (1.4–2.1 K, Hoffmann & Alexander, 2009) are less likely to be detected in AIRS data. Another notable discrepancy is that amplitudes along the winter jet are larger in JAGUAR-as-AIRS, but generally not greater by a factor of two. This fact indicates two possible model biases: one being an overestimation of wave activity and the other a bias toward longer vertical wavelengths, which are more readily detectable within the AIRS spectral sensitivity.

When compared with previous studies, the large amplitudes between 10°S and 30°S are consistent with the large climatological zonal momentum flux in this latitude range, as shown by Hindley et al. (2020). The longitudinal distribution of these peaks also broadly agrees with previous studies on the climatology of GW activity, including those of McLandress et al. (2000) and Sato et al. (2016). McLandress et al. (2000), using the Microwave Limb Sounder, estimated noise variance based on the flat structure in temperature variances at low latitudes. Meanwhile, Sato et al. (2016), using AIRS temperature measurements, focused on horizontal wavenumbers where noise influence is considered small. By excluding regions without waves, the CNN-based wave detection enables a more direct observation-to-model comparison without requiring assumptions regarding magnitude or spectral range of noise, as was necessary in previous studies.

5. Conclusions

A CNN for semantic segmentation has been developed to detect GWs both from satellite temperature perturbations and high-resolution GCM simulations. This is the first CNN-based global GW detection approach for the stratosphere, and demonstrates the applicability of this approach to both satellite observations and model simulations, while previous approaches have only been applicable to one or the other.

The CNN was trained using a large data set of wave regions detected by Berthelemy et al. (2025), which separates wave and noise regions based on wavelength consistency within neighboring pixels. Once trained, however, our CNN method does not rely on noise characteristics or spectral properties to identify waves. As a result, the wave detection is more robust across different noise levels, including the absence of any noise. The CNN not only shows similar performance detecting waves in AIRS temperatures to the neighborhood method, but also provides consistent wave detection in model simulation data. Note that, however, it remains inherently impossible to completely eliminate the influence of observational noise.

Detected GWs were compared between AIRS temperature measurements and JAGUAR simulation outputs resampled to match the AIRS sampling locations and vertical resolution, providing an almost “apples-to-apples” comparison of observed and modeled waves. The results show generally good agreement between the data sets in both the spatial distribution of GW activity and their amplitudes, although more waves are detected in the JAGUAR-as-AIRS data than in AIRS observations. An interesting discrepancy is the zonally uniform distribution of GW occurrence along the stratospheric summer jet in JAGUAR, suggesting potential model biases in GWs in the tropics and requiring further investigation.

The CNN-based wave detection offers key advantages, including not only its applicability to both satellite and model data but also its low computational cost. Other methods, such as the amplitude method and neighborhood method, typically incorporate spectral analysis as part of their detection process. The CNN allows for the pre-screening of noise-only granules, enabling spectral analysis to be performed only on data containing waves. This efficiency will facilitate GW analysis using the AIRS data, currently accumulated for 23 years.

The past 4 decades have been a golden age of stratospheric measurements. To fully utilize the extensive data collected during this period, it is essential to conduct statistical and multi-instrumental analyses of GWs in an effective and comprehensive manner, in combination with high-resolution model data. The method presented here is applicable to other nadir sounders, such as the Infrared Atmospheric Sounding Interferometer (Blumstein et al., 2004; Clerbaux et al., 2009; Hoffmann et al., 2014) aboard EUMETSAT's Meteorological Operational (MetOp) satellites and the Cross-track Infrared Sounder (CrIS) (Bloom, 2001; Eckermann et al., 2019) on NOAA's Joint Polar Satellite System satellites; we have tested and confirmed that the method works well for temperature perturbations observed by CrIS, as shown in Figure S4 in Supporting Information S1. Fair and detailed comparisons between these satellite data and GW-resolving model outputs enabled by the CNN will enhance our understanding of stratospheric GWs, providing the ground-truth needed to refine GW parameterizations in future climate-scale models and in turn reveal hidden biases that affect other simulated geophysical systems.

Data Availability Statement

The AIRS temperature retrievals (Hoffmann & Alexander, 2009) are available from https://datapub.fz-juelich.de/slcs/airs/gravity_waves (Hoffmann, 2021). The CNN, a sample code to run the CNN, and processed JAGUAR-as-AIRS data can be downloaded from <https://doi.org/10.5281/zenodo.15382321> (Okui, 2025). Version R2022b of MATLAB was used for spectral analysis of gravity waves (<https://www.mathworks.com/products/matlab.html>). Figures were produced using Python 3.10 (<https://www.python.org/>).

Acknowledgments

HO acknowledges support from a Japan Society for the Promotion of Science (JSPS) Overseas Research Fellowship. PGB was supported by an URSA studentship awarded by the University of Bath and by Royal Society Grant RF\ERE\210079. CJW was supported by Royal Society Research Fellowship URF\R\221023 and NERC Grants NE/V01837X/1, NE/W003201/1, and NE/Z50399X/1. NPH was supported by NERC Grants NE/W003201/1 and NE/Z50399X/1, and NERC Fellowship NE/X017842/1. The high-resolution JAGUAR simulations were performed using the Earth Simulator at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). HO sincerely thanks Kaoru Sato and Shingo Watanabe for their valuable feedback while performing the simulations. This manuscript benefited from discussions at the International Space Science Institute in Bern as part of International Team 567.

References

- Achatz, U., Alexander, M. J., Becker, E., Chun, H.-Y., Dörnbrack, A., Holt, L., et al. (2024). Atmospheric gravity waves: Processes and parameterization. *Journal of the Atmospheric Sciences*, 81(2), 237–262. <https://doi.org/10.1175/jas-d-23-0210.1>
- Alexander, M. J. (1998). Interpretations of observed climatological patterns in stratospheric gravity wave variance. *Journal of Geophysical Research*, 103(D8), 8627–8640. <https://doi.org/10.1029/97jd03325>
- Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., et al. (2010). Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quarterly Journal of the Royal Meteorological Society*, 136(650), 1103–1124. <https://doi.org/10.1002/qj.637>
- Becker, E., Vadas, S. L., Bossert, K., Harvey, V. L., Zülicke, C., & Hoffmann, L. (2022). A high-resolution whole-atmosphere model with resolved gravity waves and specified large-scale dynamics in the troposphere and stratosphere. *Journal of Geophysical Research: Atmospheres*, 127(2), e2021JD035018. <https://doi.org/10.1029/2021jd035018>
- Berthelemy, P. G., Wright, C. J., Hindley, N. P., Noble, P. E., & Hoffmann, L. (2025). A novel identification method for stratospheric gravity waves in nadir viewing satellite observations. *EGU sphere*. <https://doi.org/10.5194/egusphere-2025-455>
- Bloom, H. J. (2001). The cross-track infrared sounder (CrIS): A sensor for operational meteorological remote sensing. *IGARSS '01 IEEE 2001 International Geoscience and Remote Sensing Symposium*, 3, 1341–1343. <https://doi.org/10.1109/igarss.2001.976838>
- Blumstein, D., Chalon, G., Carlier, T., Buil, C., Hebert, P., Maciaszek, T., et al. (2004). IASI instrument: Technical overview and measured performances. *SPIE Proceedings*, 5543, 196–207. <https://doi.org/10.1117/12.560907>
- Carvalho, L. M., Jones, C., & Liebmann, B. (2004). The South Atlantic Convergence Zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. *Journal of Climate*, 17(1), 88–108. [https://doi.org/10.1175/1520-0442\(2004\)017%3C0088:TSACZ%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017%3C0088:TSACZ%3E2.0.CO;2)
- Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnett, C., Blaisdell, J., et al. (2006). Airs. *Bulletin of the American Meteorological Society*, 87(7), 911–926. <https://doi.org/10.1175/bams-87-7-911>
- Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., et al. (2009). Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmospheric Chemistry and Physics*, 9(16), 6041–6054. <https://doi.org/10.5194/acp-9-6041-2009>
- Cohen, N. Y., Gerber, E. P., & Bühler, O. (2013). Compensation between resolved and unresolved wave driving in the stratosphere: Implications for downward control. *Journal of the Atmospheric Sciences*, 70(12), 3780–3798. <https://doi.org/10.1175/JAS-D-12-0346.1>

- Coney, J., Denby, L., Ross, A. N., Wang, H., Vosper, S., van Niekerk, A., et al. (2023). Identifying and characterising trapped lee waves using deep learning techniques. *Quarterly Journal of the Royal Meteorological Society*, 150(758), 213–231. <https://doi.org/10.1002/qj.4592>
- Cook, K. H. (2000). The South Indian convergence zone and interannual rainfall variability over southern Africa. *Journal of Climate*, 13(21), 3789–3804. [https://doi.org/10.1175/1520-0442\(2000\)013%3C3789:TSICZA%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3C3789:TSICZA%3E2.0.CO;2)
- Eckermann, S. D., Doyle, J. D., Reinecke, P. A., Reynolds, C. A., Smith, R. B., Fritts, D. C., & Dörmbrack, A. (2019). Stratospheric gravity wave products from satellite infrared nadir radiances in the planning, execution, and validation of aircraft measurements during DEEPWAVE. *Journal of Applied Meteorology and Climatology*, 58(9), 2049–2075. <https://doi.org/10.1175/jamc-d-19-0015.1>
- Ern, M., Trinh, Q. T., Preusse, P., Gille, J. C., Mlynchak, M. G., Russell, J. M. III, & Riese, M. (2018). GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth System Science Data*, 10(2), 857–892. <https://doi.org/10.5194/essd-10-857-2018>
- Fritts, D. C. (1984). Gravity wave saturation in the middle atmosphere: A review of theory and observations. *Reviews of Geophysics* (1985), 22(3), 275–308. <https://doi.org/10.1029/rg022i003p00275>
- Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of Geophysics*, 41(1), 1003. <https://doi.org/10.1029/2001rg000106>
- Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., et al. (2013). A comparison between gravity wave momentum fluxes in observations and climate models. *Journal of Climate*, 26(17), 6383–6405. <https://doi.org/10.1175/jcli-d-12-00545.1>
- Gupta, A., Sheshadri, A., Alexander, M. J., & Birner, T. (2024). Insights on lateral gravity wave propagation in the extratropical stratosphere from 44 years of ERA5 data. *Geophysical Research Letters*, 51(14), e2024GL108541. <https://doi.org/10.1029/2024gl108541>
- Gupta, A., Sheshadri, A., & Anantharaj, V. (2024). Gravity wave momentum fluxes from 1 km global ECMWF Integrated Forecast System. *Scientific Data*, 11(1), 903. <https://doi.org/10.1038/s41597-024-03699-x>
- Hindley, N. P., Smith, N. D., Wright, C. J., Rees, D. A. S., & Mitchell, N. J. (2016). A two-dimensional Stockwell transform for gravity wave analysis of AIRS measurements. *Atmospheric Measurement Techniques*, 9(6), 2545–2565. <https://doi.org/10.5194/amt-9-2545-2016>
- Hindley, N. P., Wright, C. J., Gadian, A. M., Hoffmann, L., Hughes, J. K., Jackson, D. R., et al. (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. <https://doi.org/10.5194/acp-21-7695-2021>
- Hindley, N. P., Wright, C. J., Hoffmann, L., Moffat-Griffin, T., & Mitchell, N. J. (2020). An 18-year climatology of directional stratospheric gravity wave momentum flux from 3-D satellite observations. *Geophysical Research Letters*, 47(22), e2020GL089557. <https://doi.org/10.1029/2020gl089557>
- Hindley, N. P., Wright, C. J., Smith, N. D., Hoffmann, L., Holt, L. A., Alexander, M. J., et al. (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. <https://doi.org/10.5194/acp-19-15377-2019>
- Hoffmann, L. (2021). AIRS/Aqua observations of gravity waves [Dataset]. *Jülich DATA*, V1. <https://doi.org/10.26165/JUELICH-DATA/LQAAJA>
- Hoffmann, L., & Alexander, M. J. (2009). Retrieval of stratospheric temperatures from Atmospheric Infrared Sounder radiance measurements for gravity wave studies. *Journal of Geophysical Research*, 114(D7), D07105. <https://doi.org/10.1029/2008jd011241>
- Hoffmann, L., Alexander, M. J., Clerbaux, C., Grimsdell, A. W., Meyer, C. I., Rößler, T., & Tournier, B. (2014). Intercomparison of stratospheric gravity wave observations with AIRS and IASI. *Atmospheric Measurement Techniques*, 7(12), 4517–4537. <https://doi.org/10.5194/amt-7-4517-2014>
- Holt, L. A., Alexander, M. J., Coy, L., Liu, C., Molod, A., Putman, W., & Pawson, S. (2017). An evaluation of gravity waves and gravity wave sources in the Southern Hemisphere in a 7 km global climate simulation. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2481–2495. <https://doi.org/10.1002/qj.3101>
- Kim, Y., Eckermann, S. D., & Chun, H. (2003). An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather prediction models. *Atmosphere-Ocean*, 41(1), 65–98. <https://doi.org/10.3137/ao.410105>
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137–163. <https://doi.org/10.1093/oxfordjournals.pan.a004868>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, 14(2), 1137–1145.
- Koshin, D., Sato, K., Watanabe, S., & Miyazaki, K. (2025). The JAGUAR-DAS whole neutral atmosphere reanalysis: JAWARA. *Progress in Earth and Planetary Science*, 12(1), 1. <https://doi.org/10.1186/s40645-024-00674-3>
- Lear, E. J., Wright, C. J., Hindley, N. P., Polichtchouk, I., & Hoffmann, L. (2024). Comparing gravity waves in a kilometer-scale run of the IFS to airs satellite observations and ERA5. *Journal of Geophysical Research: Atmospheres*, 129(11), e2023JD040097. <https://doi.org/10.1029/2023jd040097>
- McLandress, C., Alexander, M. J., & Wu, D. L. (2000). Microwave limb sounder observations of gravity waves in the stratosphere: A climatology and interpretation. *Journal of Geophysical Research*, 105(D9), 11947–11967. <https://doi.org/10.1029/2000jd900097>
- Okui, H. (2025). Gravity-wave-detecting CNN and JAGUAR-as-AIRS dataset [Dataset]. *Zenodo*. <https://doi.org/10.5281/ZENODO.15382321>
- Okui, H., Sato, K., Koshin, D., & Watanabe, S. (2021). Formation of a mesospheric inversion layer and the subsequent elevated stratopause associated with the major stratospheric sudden warming in 2018/19. *Journal of Geophysical Research: Atmospheres*, 126(18), e2021JD034681. <https://doi.org/10.1029/2021jd034681>
- Okui, H., Sato, K., & Watanabe, S. (2022). Contribution of gravity waves to universal vertical wavenumber ($\sim m^{-3}$) spectra revealed by a gravity-wave-permitting general circulation model. *Journal of Geophysical Research: Atmospheres*, 127(10), e2021JD036222. <https://doi.org/10.1029/2021jd036222>
- 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. <https://doi.org/10.1029/2023jd038795>
- Pahlavan, H. A., Wallace, J. M., & Fu, Q. (2023). Characteristics of tropical convective gravity waves resolved by ERA5 reanalysis. *Journal of the Atmospheric Sciences*, 80(3), 777–795. <https://doi.org/10.1175/jas-d-22-0057.1>
- Podglajen, A., Hertzog, A., Plougonven, R., & Legras, B. (2020). Lagrangian gravity wave spectra in the lower stratosphere of current (re)analyses. *Atmospheric Chemistry and Physics*, 20(15), 9331–9350. <https://doi.org/10.5194/acp-20-9331-2020>
- Rao, J., Garfinkel, C. I., Chen, H., & White, I. P. (2019). The 2019 New Year stratospheric sudden warming and its real-time predictions in multiple S2S models. *Journal of Geophysical Research: Atmospheres*, 124(21), 11155–11174. <https://doi.org/10.1029/2019jd030826>

- Rhode, S., Preusse, P., Ungermann, J., Polichtchouk, I., Sato, K., Watanabe, S., et al. (2024). Global-scale gravity wave analysis methodology for the ESA Earth Explorer 11 candidate CAIRT. *Atmospheric Measurement Techniques*, 17(19), 5785–5819. <https://doi.org/10.5194/amt-17-5785-2024>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Sato, K. (2025). JAGUAR data assimilation system whole neutral atmosphere reanalysis (JAWARA) [Dataset]. *National Institute of Polar Research*. <https://doi.org/10.17592/002.2025010407>
- Sato, K., Tsuchiya, C., Alexander, M. J., & Hoffmann, L. (2016). Climatology and ENSO-related interannual variability of gravity waves in the southern hemisphere subtropical stratosphere revealed by high-resolution AIRS observations. *Journal of Geophysical Research: Atmospheres*, 121(13), 7622–7640. <https://doi.org/10.1002/2015jd024462>
- Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: The S transform. *IEEE Transactions on Signal Processing*, 44(4), 998–1001. <https://doi.org/10.1109/78.492555>
- Vincent, D. G. (1994). The South Pacific convergence zone (SPCZ): A review. *Monthly Weather Review*, 122(9), 1949–1970. [https://doi.org/10.1175/1520-0493\(1994\)122%3C1949:TSPCZA%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122%3C1949:TSPCZA%3E2.0.CO;2)
- 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*, 114(D7), D07110. <https://doi.org/10.1029/2008jd011218>
- 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. <https://doi.org/10.5194/amt-14-5873-2021>
- Wright, C. J., Hindley, N. P., Moss, A. C., & Mitchell, N. J. (2016). Multi-instrument gravity-wave measurements over Tierra del Fuego and the Drake Passage – Part I: Potential energies and vertical wavelengths from AIRS, COSMIC, HIRDLS, MLS-Aura, SAAMER, SABER and radiosondes. *Atmospheric Measurement Techniques*, 9(3), 877–908. <https://doi.org/10.5194/amt-9-877-2016>
- Yoshida, L., Tomikawa, Y., Ejiri, M. K., Tsutsumi, M., Kohma, M., & Sato, K. (2024). Large-amplitude inertia gravity waves over Syowa station: Comparison of PANSY radar and ERA5 reanalysis data. *Journal of Geophysical Research: Atmospheres*, 129(22), e2023JD040490. <https://doi.org/10.1029/2023jd040490>