AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL104668

# Accelerating Atmospheric Gravity Wave Simulations Using Machine Learning: Kelvin-Helmholtz Instability and Mountain Wave Sources Driving Gravity Wave Breaking and Secondary Gravity Wave Generation

**Wenjun Dong[1,2,3], David C. Fritts[1,2], Alan Z. Liu[1], Thomas S. Lund[2], Han-Li Liu[3], and Jonathan Snively[1]**

[1]Center for Space and Atmospheric Research (CSAR), Department of Physical Sciences, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, [2]Global Atmospheric Technologies and Sciences (GATS), Boulder, CO, USA, [3]High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO, USA

**Key Points:**

- A machine learning model (CAM-Net) tailored for nonlinear Gravity wave (GW) simulations is developed
- CAM-Net can achieve a several order-of-magnitude acceleration relative to physics-based model without sacrificing accuracy
- CAM-Net opens a new window to improve the parameterization of primary and secondary GWs in the global atmospheric models

**Supporting Information:**

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

**Correspondence to:**

W. Dong, and A. Z. Liu,
wenjun@gats-inc.com;
Liuz2@erau.edu

**Citation:**

Dong, W., Fritts, D. C., Liu, A. Z., Lund, T. S., Liu, H.-L., & Snively, J. (2023). Accelerating atmospheric gravity wave simulations using machine learning: Kelvin-Helmholtz instability and mountain wave sources driving gravity wave breaking and secondary gravity wave generation. *Geophysical Research Letters*, *50*, e2023GL104668. https://doi.org/10.1029/2023GL104668

Received 23 MAY 2023
Accepted 13 JUL 2023

**Author Contributions:**

**Conceptualization:** Wenjun Dong, David C. Fritts, Alan Z. Liu
**Data curation:** Wenjun Dong, Thomas S. Lund
**Formal analysis:** Wenjun Dong, David C. Fritts, Alan Z. Liu, Thomas S. Lund
**Funding acquisition:** David C. Fritts, Alan Z. Liu
**Investigation:** Wenjun Dong, David C. Fritts, Alan Z. Liu, Thomas S. Lund, Han-Li Liu



**Abstract** Gravity waves (GWs) and their associated multi-scale dynamics are known to play fundamental roles in energy and momentum transport and deposition processes throughout the atmosphere. We describe an initial machine learning model—the Compressible Atmosphere Model Network (CAM-Net). CAM-Net is trained on high-resolution simulations by the state-of-the-art model Complex Geometry Compressible Atmosphere Model (CGCAM). Two initial applications to a Kelvin-Helmholtz instability source and mountain wave generation, propagation, breaking, and Secondary GW (SGW) generation in two wind environments are described here. Results show that CAM-Net can capture the key 2-D dynamics modeled by CGCAM with high precision. Spectral characteristics of primary and SGWs estimated by CAM-Net agree well with those from CGCAM. Our results show that CAM-Net can achieve a several order-of-magnitude acceleration relative to CGCAM without sacrificing accuracy and suggests a potential for machine learning to enable efficient and accurate descriptions of primary and secondary GWs in global atmospheric models.

**Plain Language Summary** Atmospheric gravity waves (GWs) are well described by the Navier-Stokes equations, but solving these equations including small scale remains daunting, limited by the very high computational cost of resolving the smallest spatial-temporal features in a global context. To address this challenge, we developed a machine learning model called CAM-Net. Our model demonstrates that neural networks can be trained on high-resolution compressible atmospheric model data and then used to simulate GW evolution. Importantly, initial results show that using such trained model can achieve computational savings of >1,000 times compared to a physics-based simulation while still achieve highly accurate results. These findings are exciting, as they suggest that CAM-Net can overcome the limitations of current GW parameterizations and provide a promising avenue for studying the effects of sub-grid-scale processes in atmospheric science and properly incorporating them in global models. The development of CAM-Net opens up major new opportunities for improving effective model resolution, accuracy, and efficiency.

## 1. Introduction

Gravity waves (GWs) play prominent roles throughout Earth's atmosphere. They are generated at lower altitudes by various primary sources including airflow over topography (i.e., mountain waves, MWs), convection, and jet streams (Fritts & Alexander, 2003). Accounting for smaller-scale GW transports and influences remains a challenging problem due to the complex physics involved and the need for high-resolution simulations to describe detailed responses where these are important. Multiple parameterization schemes spanning 40 years have aimed to account for GW pseudo-momentum deposition for various GW sources, discretely or spectrally (i.e., linearly or nonlinearly), from the surface into the thermosphere (e.g., Fritts & Lu, 1993; Holton, 1982; Lindzen, 1981; Palmer et al., 1986). More recent schemes have built on these earlier efforts and insights (Alexander & Dunkerton, 1999; Amemiya & Sato, 2016; Eckermann et al., 2015; Gettleman et al., 2019; Hines, 1997; Miyoshi & Yiğit., 2019; Ribstein et al., 2022; Warner & McIntyre, 1996; Yiğit et al., 2008).

Importantly, all these various schemes are based on simplified, often linear or weakly nonlinear, mathematical models and/or empirical relations that are significant approximations having limited quantitative predictive

**Methodology:** Wenjun Dong, Thomas S. Lund
**Project Administration:** David C. Fritts, Alan Z. Liu
**Resources:** Wenjun Dong
**Software:** Wenjun Dong, Thomas S. Lund
**Supervision:** David C. Fritts, Alan Z. Liu
**Validation:** Wenjun Dong, David C. Fritts, Alan Z. Liu, Han-Li Liu, Jonathan Snively
**Visualization:** Wenjun Dong
**Writing – original draft:** Wenjun Dong
**Writing – review & editing:** Wenjun Dong, David C. Fritts, Alan Z. Liu, Han-Li Liu, Jonathan Snively

abilities. As such, they introduce significant model uncertainties and biases in predictions of middle and upper atmosphere responses (Pedatella et al., 2014). Parameterizations addressing GWs that are partially resolved (the "gray zone") that maintain physical consistency between the resolved and parameterized dynamics are promising (Vosper, 2015; Vosper et al., 2016), but can also be challenging (Liu, 2019). Such efforts attempt to represent the complex and highly nonlinear physics of GWs. However, these assumptions and simplifications limit the representation of many observed GW characteristics, which are physically well understood (e.g., Eckermann et al., 2015; Eyring et al., 2006; Hertzog et al., 2008; Stephan et al., 2016).

The recent boom in hardware and software developments relevant to machine learning (ML) has motivated some efforts to examine the possible benefits that ML can bring to GW parameterization (e.g., Chantry et al., 2021; Espinosa et al., 2022). Chantry et al. (2021) trained a neural network on an upgraded version of an existing parameterization scheme that yielded improved results describing GW drag in a numerical weather prediction system. In another study, Espinosa et al. (2022) applied an artificial neural network to emulate the pseudo-momentum forcing described in a traditional GW parameterization in an idealized climate model. By coupling the climate model with their ML-based GW parameterization, they were able to accurately reproduce the quasi-biennial oscillation. However, these ML-based GW parameterizations rely on traditional GW parameterizations, hence inherit their assumptions and simplifications.

The dynamics of GWs are governed by the Navier-Stokes equations. In recent years, several ML-based solvers for partial differential equations (PDEs) have been proposed to approximate or improve various numerical methods. The most explored of these can be divided into two categories: physics-informed neural networks (PINN, e.g., Maziar et al., 2019; Wandel et al., 2022) and neural operators (NOs, e.g., Lu et al., 2019; Li et al., 2020; Xiong et al., 2023). PINN uses a neural network as the solution function and optimizes a loss function to minimize violation of the given equation. However, it experiences difficulties in propagating information from initial or boundary conditions to unseen parts of the interior and to future times. NOs are better suited for solving PDEs and have been successfully used in flow prediction (e.g., Li et al., 2020; Lu et al., 2019; Xiong et al., 2023). However, they require large volumes of simulation data. Recently, physics-informed NOs, for example, both the physics-informed Deep Operator Network proposed by Goswami et al. (2022) and physics-informed Fourier NO proposed by Li et al. (2021) employ both data and physics losses on operator learning to overcome the shortcomings of purely PINN or data-driven learning.

In this paper, we focus on the potential of ML to improve the efficiency of GW simulations while capturing their highly nonlinear dynamics with high fidelity, specifically including instabilities, breaking, and SGW generation. A detail description of our method is provided in Section 2, Section 3 describes our results, and our discussion and summary are provided in Section 4.

## 2. Method: Compressible Atmospheric Model Network (CAM-Net)

CAM-Net is a hybrid machine learning model that combines data-driven and physics-informed approaches. We employ the Complex Geometry Compressible Atmospheric Model (CGCAM) to create the training datasets for CAM-Net. CGCAM is a finite volume model that has been used extensively to study GW dynamics and their instabilities in the Mesosphere and Lower Thermosphere (MLT) at very high resolution (see, e.g., Dong et al., 2020, 2021, 2022, 2023; Fritts et al., 2020, 2021, 2022a, 2022b; Lund et al., 2020).

CAM-Net is based on the Adaptive Fourier NO (AFNO) proposed by Guibas et al., 2021. AFNO is a Fourier transform-based token-mixing scheme with a vision transformer backbone (Dosovitskiy et al., 2020). The Fourier architecture of AFNO applies a fast Fourier Transform (FFT) to the data and applies its fully connected layers in Fourier space before performing an inverse FFT back to real space. The introduction of vision transformer enables it to model long-range dependencies well and yields a state-of-the-art high-resolution model that resolves fine-grained features and scales well with resolution and size of data set.

The power of AFNO stems from its ability to combine linear integral operators, implemented through the Fourier transform, with non-linear activation functions, enabling it to learn highly non-linear operators. Although AFNO truncates higher frequency modes in the Fourier layer, Li et al. (2020) argue that the entire operator can still approximate functions with the full frequency range, due to the function being represented in a high-dimensional channel space. In our case of predicting multi-scale GW dynamics, the Fourier layer truncation of high-frequency information resulted in poor small-scale structure prediction. To address this, we added a convolutional layer,

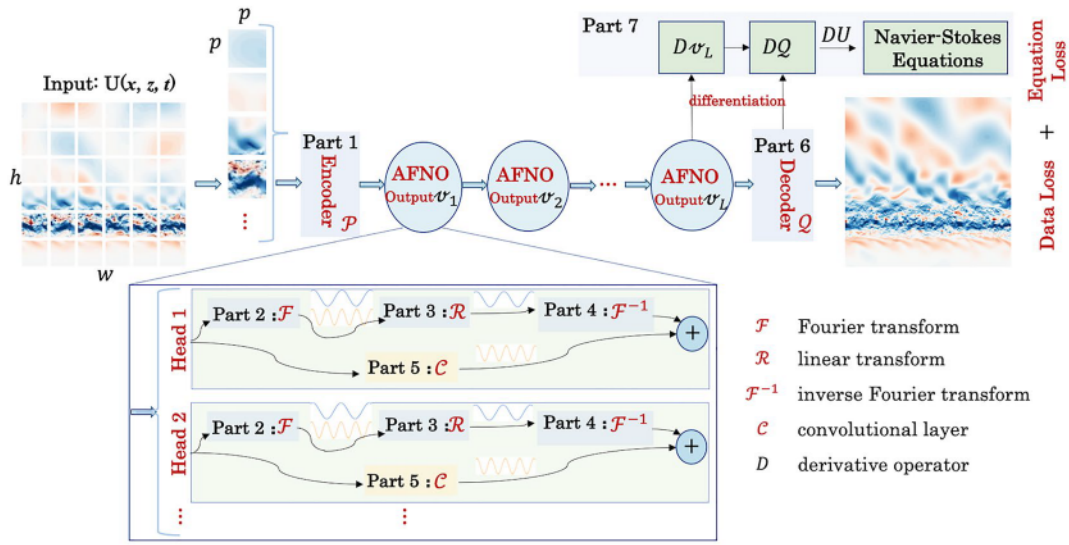


**Figure 1.** CAMNet architecture that utilizes the modified Adaptive Fourier Neural Operator (AFNO) and follows a patch-based approach. The input frame is divided into a $h \times w$ grid of patches, each of size $p \times p$, and encoded in a higher dimensional space with position embedding is added to form a sequence of tokens. These tokens are then mixed spatially using AFNO, which is repeated for L layers, and then a decoder reconstructs the patches for the next frame. A single AFNO layer is composed of multiple heads for parallel processing. Parts 2–4 of the layer applies the Fourier transform $\mathcal{F}$ to the input, followed by a linear transform $\mathcal{R}$ that acts on the lower-frequency Fourier modes and filters out the higher-frequency modes, this is then followed by an inverse Fourier transform $\mathcal{F}^{-1}$. Part 5 stands for a convolutional layer that is used to capture the higher-frequency modes that are missed by Part 2–4. Both AFNO outputs $v_L$ and reconstructed $U$ are functions, and their derivatives $Dv_L$ and $DU$ can be computed at any query points $x$ and $z$. The solutions are constrained by the Navier-Stokes equations that govern GW dynamics. For more information, refer to the corresponding texts.

which is able to amplify high-frequency components and complement the information truncated by the Fourier layer.

CAM-Net further extends the AFNO architecture by incorporating physics information from the Navier-Stokes equations, which govern the GW dynamics. These equations are used to create a loss function that captures the violation of these laws, and Fourier derivatives (Li et al., 2021) are used to compute the derivatives for the physics constraints. The physical constraints reduce the demand for training datasets and improve the generalization and physical validity of CAM-Net learning compared to purely data-driven methods.

The CAM-Net architecture is shown in Figure 1. The CAM-Net is currently trained on horizontal wind $U$, so CAM-Net's input is $U$ ($x$, $z$, $t$). The CAM-Net architecture consists of 7 parts. Below, we present a detailed computational implementation of each part. Similar to the iterative update strategy of each Fourier layer in FNO (Li et al., 2020), the improved iterative update strategy of each Fourier layer in CAM-Net can be expressed as follows:

$$v_1 = \overbrace{\mathcal{P}}^{\text{Part 1}} U(x,z,t) \tag{1}$$

$$v_{l+1} = \sigma\left(\overbrace{\mathcal{C}}^{\text{part 5}} v_l + \overbrace{\mathcal{F}^{-1}}^{\text{part 2}} \overbrace{\mathcal{R}}^{\text{part 3}} \overbrace{\mathcal{F}}^{\text{part 4}} v_l\right), for\ l = 1, 2, \cdots, L \tag{2}$$

$$U(x,z,t+1) = \overbrace{Q}^{\text{part 6}} v_L \tag{3}$$

where $\mathcal{P}$ and $Q$ are encoder and decoder that are realized by two neural networks that projects $U(x,z,t)$ to hidden representation $v_l$ and projects the representation back to the solution $U(x,z,t+1)$. $\sigma$ is a nonlinear activation function. The additional term $C$ is a convolutional layer that acts on $v_l$. $v_l$ denotes the output of the $l$th Fourier layer of

AFNO. $\mathcal{F}$ denotes the Fourier transform that acts on $v_l$, $\mathcal{R}$ is a linear transform layer that acts on $\mathcal{F}(v_l)$ to handle its low-frequency modes. $\mathcal{F}^{-1}$ is an inverse Fourier transform that acts on $\mathcal{R}(\mathcal{F}(v_l))$. The details of each part are provided below:

*Part* 1: Encoder $\mathcal{P}$ The encoder is implemented using the token embedding layer in the Vision Transformer architecture proposed by Dosovitskiy et al. (2020). This layer applies a linear projection to each patch to obtain a fixed-sized vector, which is then concatenated with positional embeddings representing the spatial location of the patch. The resulting sequence of vectors serves as the input to the sequent neural network layers.

*Part* 2–4: Fourier Transform $\mathcal{F}$, Linear Transform $\mathcal{R}$, and Inverse Fourier Transform $\mathcal{F}^{-1}$: As stated in Li et al., 2020, since the inputs and outputs of PDEs are continuous functions, it is more efficient to represent them in Fourier space and perform global convolution. This is due to the quasi-linear computational complexity and global properties of Fourier transform. The convolution in the spatial domain is equivalent to the pointwise multiplication in the Fourier domain. To capture global features in input data, a Fourier transform is first applied to the inputs, followed by a Linear Transform $\mathcal{R}$ that acts on the lower-frequency Fourier modes by assigning weights to them. These weights will be updated during the training. Finally, an inverse Fourier transform is performed to obtain the output.

*Part* 5: High-frequency Information Compensation $C$: In each Fourier layer, we utilize a convolutional layer to extract high-frequency information because it can amplify high-frequency components. We train a $\boldsymbol{C}$ on the outputs of Part 1 to extract their high-frequency information. As a complement to Parts 2–4, $\boldsymbol{C}$ enables the forward prediction of high-frequency information.

*Part* 6: Decoder $Q$: Given two future states independently predicted by Parts 2–4 and Part 5, we combine them and train a non-linear decoder using a multi-Layer perceptron layer with a tanh activation function to transform the AFNO outputs back into $U$.

*Part* 7: Physics Informed Loss $\mathcal{L}_p$: As CAM-Net is currently trained on horizontal wind $U$, we utilize only the momentum flux equation for the physics-informed part, assuming a constant density $\rho$ over time. Thus, the equation can be simplified to

$$\frac{\partial(\rho U)}{\partial x} + \frac{\partial(\rho W)}{\partial z} = 0 \tag{4}$$

In Equation 4, the vertical wind $W$ is obtained from CGCAM simulations at each prediction time step, and $\rho$ is set to its initial values. The only physics-informed variable to be calculated in CAM-Net is the derivative of $U$ with respect to $x$.

The loss function of CAM-Net for optimizing Equations 1–3 is defined as

$$\mathcal{L} = \alpha\mathcal{L}_{\text{data}} + \beta\mathcal{L}_p = \alpha\left|U - \hat{U}\right| + \beta\left|\frac{\partial(\rho U)}{\partial x} + \frac{\partial(\rho W)}{\partial z}\right| \tag{5}$$

where $\alpha$ and $\beta$ control the weights of data-driven and physics-informed part in loss functions, respectively.

We performed two cases to evaluate the performance of CAM-Net, and the relevant model parameters are presented in Table S1 in Supporting Information S1.

## 3. Results

This section describes our initial efforts using CAM-Net to accelerate simulations of GWs arising from two very different sources constrained to a two-dimensional (2-D) domain. Case 1 describes the generation of initial GWs by large-scale, shear-induced Kelvin-Helmholtz Instability (KHI) and the successive generation of SGWs at much larger scales that readily propagate to much higher altitudes. Case 2 describes MWs arising from flow over idealized terrain, their attainment of large amplitudes, breaking, and generation of SGWs that likewise attain very high altitudes. CAM-Net wind fields and spectra are compared with high-resolution 2-D CGCAM simulations in both cases.

### 3.1. Case 1: Gravity Waves Emitted From Kelvin–Helmholtz Instability

We explore here the potential of CAM-Net for modeling GWs emitted from KHI described by Dong et al. (2023). We use CGCAM to generate training and testing data for CAM-Net. The initial background winds (see Figure S1 in Supporting Information S1) are specified as

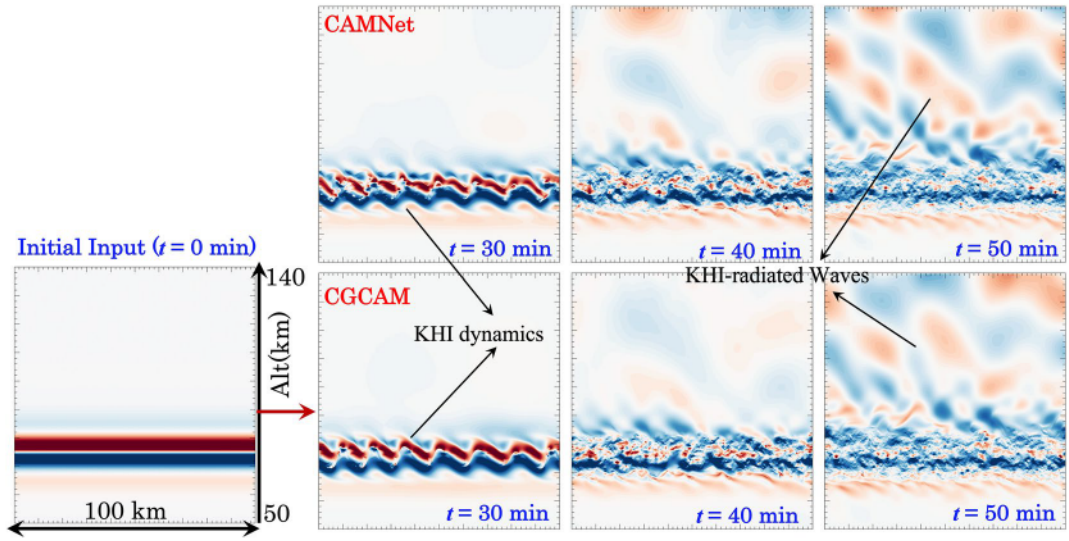


**Figure 2.** KHI and GWs predicted by CGCAM and CAM-Net.

$$U(z) = U_0 \cos\left[\frac{\pi(z - z_0)}{15\,\text{km}}\right] \tan h\left(\frac{z - z_0}{h}\right) \tag{6}$$

CGCAM simulations are performed for a computational domain having dimensions 180 km × 180 km ($x$,$z$) with resolutions of 50 m at the shear center, with exponential mesh stretching approaching the upper and lower boundaries to reduce computational demands. Periodic boundary conditions are used at the lateral boundaries. An isothermal no-stress wall condition is used at the lower boundary, and a characteristic radiation condition is used at the upper boundary. The vertical boundary conditions are supplemented with sponge layers having 20-km depths to further ensure no reflected GWs. After excluding irrelevant data in the sponge layers, the variable $U$ are stored on a grid of dimension of 2,000 × 1,000.

Given the initial conditions, CAM-Net is required to simulate the future states of variable $U$ at $t \in \{1,2,3,\cdots,50\}$ min for a suite of initial conditions. A total of 200 cases are generated by varying $U_0$ and $h$ in Equation 6, and the corresponding outputs of CGCAM serves as the true reference solutions for each case. The CGCAM simulations for each case were run for 50 min at an interval of 1 min. The 200 CGCAM cases were then split into a training set of 180 cases and a testing set of 20 cases. All samples have a grid of 2,000 × 1,000. The CAM-Net training is implemented in a multi-GPU environment with 4 V100 GPUs.

Figure 2 displays the variable $U$ employed for both the CAM-Net and CGCAM simulations during the model test, and we observe a high level of consistency between two models. The ML model CAM-Net can capture small- and large-scale structures qualitatively, with clear evidence of KHI and KHI-radiated GWs seen in both CGCAM and CAM-Net. The initial strong shear produces deep and broad KH billows that break down after ~30 min, leading to the emergence of small-amplitude GWs above the KHI altitude. By 50 min, increasing GWs are seen to propagate with group velocities along the iso-phase lines that extend away from the KHI shear layer. Both models suggest that GWs are continuously emitted from the KHI dynamics, with well-defined spatial structures having orientations and spatial scales that agree closely between the CGCAM and CAM-Net fields.

The spectral properties of GWs can reveal important details about their sources, such as the altitude and vertical extent of the source region, and the dominant wavelengths and frequencies of the GWs generated. The spectral structures of horizontal wind disturbance $u'$ are calculated and found to exhibit high consistency between CGCAM and CAM-Net. The spectra of $u'^2$ computed at the KHI region and higher altitudes at $t$ = 30, 40, and 50 min are displayed in Figure 4. The spectra reflect the characteristics of KHI dynamics. In the KHI region, the onset of strong KHI yields spectral slopes approaching −5/3 corresponding to wavenumbers of ~1–6 rad/km. Spectral amplitudes fall sharply beyond wavenumbers of ~20 rad/km and exhibit steeper slopes approaching −7 in the viscous range. At higher altitudes, the −5/3 slope corresponds to wavenumbers of ~0.8–3 rad/km and −7

slopes correspond to wavenumbers of 3–10 rad/km. These spectra suggest that small-scale structures discussed above are well resolved at these times.

CAM-Net has promising potential as an alternative to CGCAM for simulating KHI and KHI-radiated GWs. The time cost of a single KHI case simulation using CAM-Net was approximately 0.8 s on a single A100 (80GB). This represents a significant acceleration (by >2,000) compared with the 30 min needed by CGCAM when using 36 CPU cores.

### 3.2. Case 2: Mountain Waves Generation, Propagation, and Breaking

Our intent in Case 2 was to explore CAM-Net capabilities for modeling MW generation, propagation, breaking, and SGW generations. MW breaking is one of the strongest sources of SGWs (Lund et al., 2020). As in Case 1, CAM-Net training employs CGCAM simulation data. The CGCAM simulations cover a computational domain extending 700 × 220 km ($x$ and $z$) at a resolution of 1 and 0.5 km in $x$ and $z$. The lateral boundary is periodic. At the lower boundary, a Gaussian terrain of peak height 4 km and half-width of 30 km is used, and a characteristic radiation condition is used at the upper boundary. Sponge layers of 20 and 50 km are added to the upper and lateral boundaries, respectively, to ensure absorption of outgoing GWs. The variable $U$ are stored on a grid having dimensions of 600 × 400 after irrelevant values in the sponge layers are excluded.

Given the samples of initial conditions, CAM-Net is trained to reproduce the future states of variable $U$ at $t \in \{140, 145, 150, \cdots, 220\}$ min. Note that we start from $t = 140$ min to avoid CAM-Net being trained with non-physical data produced by CGCAM at early simulation times. A total of 200 cases are generated by varying the initial wind field, with a random wind field randomly extracted from the HWM14 at 30°S, 70°W (Andes Lidar Observatory) at 00:00 on 200 days among 365 days. The initial temperature field is simplified as used in Dong et al. (2020). Winds at lower altitudes from HWM14 are consistently lower than actual observations, and thus a correction is needed to enable simulation of MW generation. To account for this discrepancy, we randomly assign wind values ranging from 0 to 30 m/s at these lower altitudes to facilitate the occurrence of MWs. The initial fields were assumed to be uniform over the domain. The initial wind and temperature fields are shown in Figure S2 in Supporting Information S1. CGCAM simulations for each case were run for 220 min at an interval of 5 min. The CGCAM simulations were then split into a training set of 180 cases and a testing set of 20 cases. All samples have a grid of 600 × 400. The training of CAM-Net is implemented in a multi-GPU environment with 4 V100.

Our results demonstrate that CAM-Net is capable of effectively modeling MW generation, propagation, breaking, and SGW generation. We chose a case with an initial wind field that includes a tidal wind field (represented by gray lines in Figure 3) that was not part of the 200 cases used for training and testing. Figure 3 shows $U$ at $t = 150$, 180, 200, 220 min generated by CAM-Net and the corresponding CGCAM results at these times. Considering CAM-Net results first, the earliest responses at $t = 150$ min reveal MW generation at lower altitudes and their extension into the MLT. At $t = 180$ min, initial instabilities are seen at lower altitudes. At $t = 200$ min, there is evidence for strong SGW excitation in the MW breaking regions. The MW field and its associated instabilities and SGWs continue to intensify to $t = 220$ min. SGWs are modulated by tidal winds and have large scales and large influences extending into the thermosphere. The CAM-Net results approximate the CGCAM ground truth remarkably well over 220 min. Additionally, high consistency is found between CGCAM and CAM-Net in the $u'$ spectra, which are shown in Figure 4 (third and second rows).

CAM-Net demonstrates promising potential as a competitive alternative to CGCAM for simulating MW generation, propagation, and breaking. A single MW case simulation using CAM-Net takes approximately 0.5 s on a single A100 (80GB), in this case ~4,000 times faster than the corresponding CGCAM simulation using 36 CPU cores, which took around 40 min.

## 4. Summary

In this paper, we developed a machine learning model solving the compressible Navier-Stokes equations in our Complex Geometry Compressible Atmosphere Model (CGCAM) named CAM-Net. CAM-Net is a hybrid machine learning model that combines data-driven and physics-informed approaches. It is based on the Adaptive Fourier NO (AFNO) with modifications tailored to our simulations. The main improvements include: (a) the

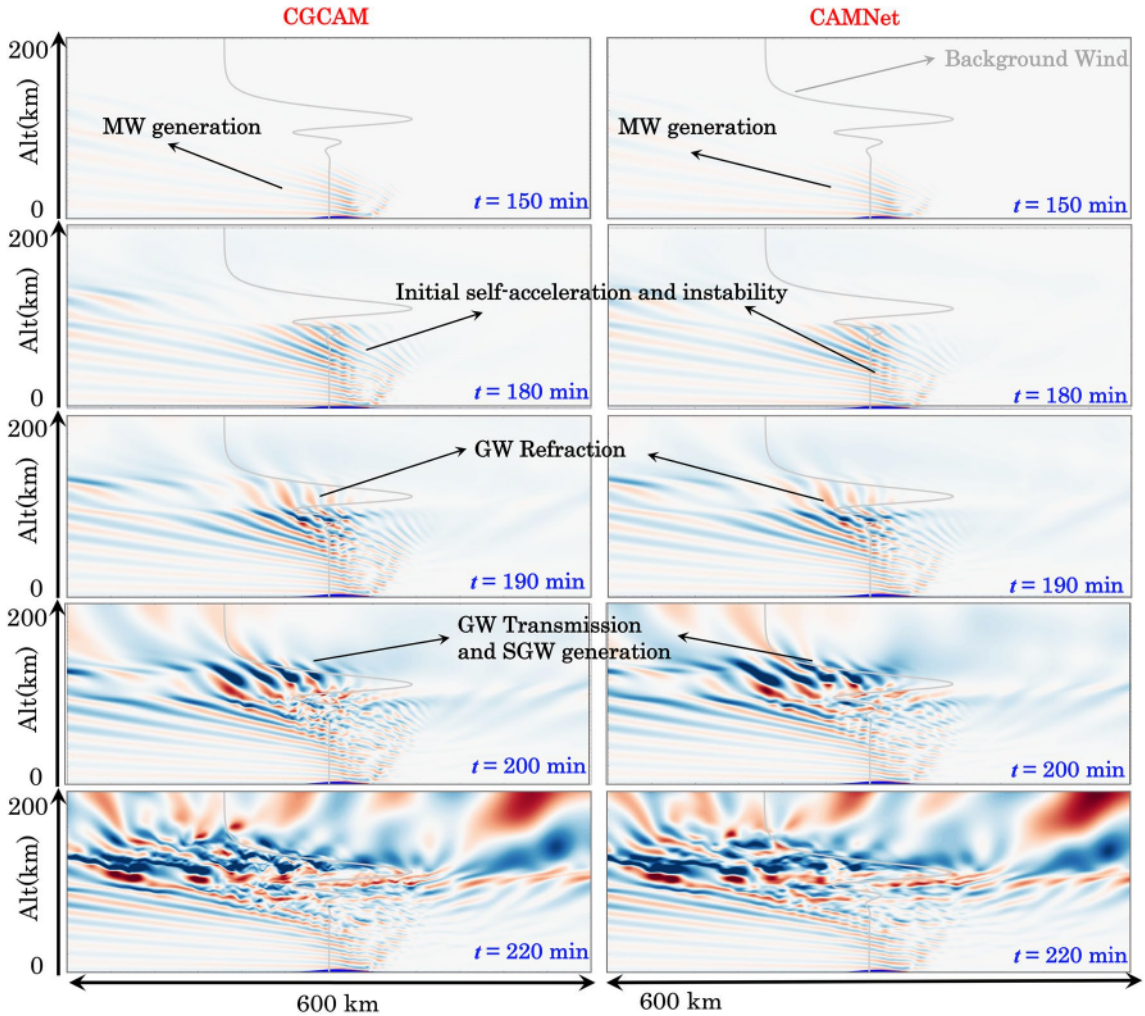


**Figure 3.** MW evolutions predicted by CGCAM (left) and CAM-Net (right). The initial condition (gray lines) is a horizontal wind of 25 m/s at the surface flowing over a Gaussian Mountain with a height of 4 km and a half width of 30 km.

addition of convolutional layer branch to compensate for high-frequency components truncated by the Fourier layers, making the model more robust in resolving multi-scale dynamics, and (b) the incorporation of physical information from the Navier-Stokes equation in CGCAM. The CAM-Net feedback neural network utilizes a loss function that combines both the physical information in Navier-Stokes equations and the data loss from CGCAM simulations. This approach has the potential to reduce the need for extensive training data. However, it is important to consider the influence of the training data and the need for further investigation in more diverse and complex scenarios.

CAM-Net has successfully undergone validation on two compelling cases: KHI and MW Sources Driving GW Breaking and SGW Generation. Both cases have demonstrated CAM-Net's remarkable capacity to significantly accelerate GW simulations. To further enhance the application of CAM-Net's exceptional GW simulation capabilities, CAM-Net will be optimized and extended (a) from single-variable simulation to multi-variable simulation to better capture the complex nonlinear interactions among various GW variables; and (b) from 2D to 3D to model the horizontal evolution characteristics of GWs more accurately, which are only fully displayed in 3D cases.

CAM-Net offers several advantages over traditional and previous ML-based GW parameterization schemes. The training process of CAM-Net does not rely on any existing GW parameterization schemes, thus it is not limited by their assumptions and simplifications. CAM-Net is trained with high-resolution simulation data from CGCAM, which are accurate numerical solutions of the Navier-Stokes equations. Well-trained CAM-Net is capable of

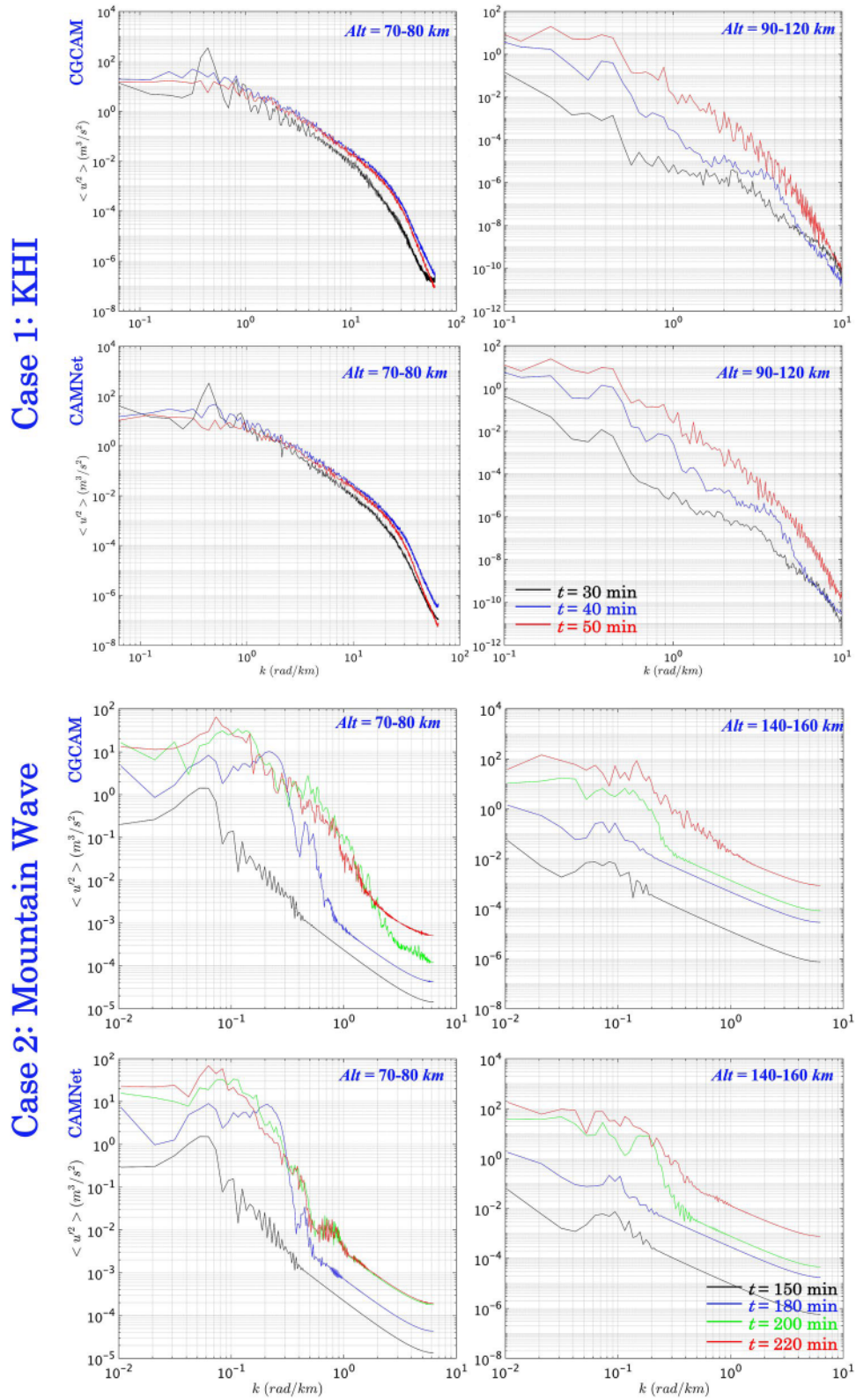


**Figure 4.** Cases 1 (firth and second rows) and 2 (third and fourth rows) spectral characteristics predicted by CGCAM and CAM-Net.

resolving multi-scale and highly nonlinear GW dynamics, such as instability, GW breaking, and SGW generations, at much faster speed. To the best of our knowledge, CAM-Net is the first ML-based approach that can directly simulate highly nonlinear GW dynamics.

A well-trained CAM-Net might produce the simulations orders of magnitude faster than CGCAM, while maintaining an acceptable level of accuracy. First, high-resolution GW simulations can be generated within seconds, thus enabling estimation of well-calibrated and constrained uncertainties regarding unresolved GW scales with higher confidence compared to current global models that have severely simplified GW parameterization schemes due to computational cost. Second, there is potential to develop a software library of well-trained CAM-Net models to be applied to a broad range of conditions. The well-trained models have the potential to become a viable alternative to current GW parameterizations in global models.

## Data Availability Statement

The data required to reproduce each figure can be found at https://doi.org/10.6084/m9.figshare.22362955. Simulation data is provided in a vtk file format and can be opened by the open-source software ParaView version 5.10, which is available under the BSD license at https://www.paraview.org/download/. The ParaView Guide can be downloaded from http://www.paraview.org/download/.The original AFNO code is available at https://github.com/lonestar686/AdaptiveFourierNeuralOperator.

**Acknowledgments**
Research described here was supported by the Air Force Office of Scientific Research (AFOSR) Grant FA9550-18-1-0009 and NSF Grants AGS-1759471, AGS-2032678, AGS-2131350, and AGS-2128443. We also acknowledge Embry-Riddle Aeronautical University and the National Center for Atmospheric Research for access to supercomputer platforms that allowed the CGCAM and CAMNet simulations reported here. The National Center for Atmospheric Research is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

## References

Alexander, M. J., & Dunkerton, T. J. (1999). A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, *56*(24), 4167–4182. https://doi.org/10.1175/1520-0469(1999)056%3C4167:ASPOMF%3E2.0.CO;2

Amemiya, A., & Sato, K. (2016). A new gravity wave parameterization including three-dimensional propagation. *Journal of the Meteorological Society of Japan*, *94*(3), 237–256. https://doi.org/10.2151/jmsj.2016-013

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002477. https://doi.org/10.1029/2021MS002477

Dong, W., Fritts, D. C., Hickey, M. P., Liu, A. Z., Lund, T. S., Zhang, S., et al. (2022). Modeling studies of gravity wave dynamics in highly structured environments: Reflection, trapping, instability, momentum transport, secondary gravity waves, and induced flow responses. *Journal of Geophysical Research: Atmospheres*, *127*(13), e2021JD035894. https://doi.org/10.1029/2021JD035894

Dong, W., Fritts, D. C., Liu, A. Z., Lund, T. S., & Hiu, H. (2023). Gravity waves emitted from Kelvin-Helmholtz instabilities. *Geophysical Research Letters*, *50*(8), e2022GL102674. https://doi.org/10.1029/2022GL102674

Dong, W., Fritts, D. C., Lund, T. S., Wieland, S. A., & Zhang, S. (2020). Self-acceleration and instability of gravity wave packets: 2. Two-dimensional packet propagation, instability dynamics, and transient flow responses. *Journal of Geophysical Research: Atmospheres*, *125*(3), e2019JD030691. https://doi.org/10.1029/2019JD030691

Dong, W., Fritts, D. C., Thomas, G. E., & Lund, T. S. (2021). Modeling responses of polar mesospheric clouds to gravity wave and instability dynamics and induced large-scale motions. *Journal of Geophysical Research: Atmospheres*, *126*(13), e2021JD034643. https://doi.org/10.1029/2021JD034643

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. https://doi.org/10.48550/arXiv.2010.11929

Eckermann, S. D., Broutman, D., & Knight, H. (2015). Effects of horizontal geometrical spreading on the parameterization of orographic gravity wave drag. Part II: Analytical solutions. *Journal of the Atmospheric Sciences*, *72*(6), 2348–2365. https://doi.org/10.1175/jas-d-14-0148.1

Espinosa, Z. I., Sheshadri, A., Cain, G. R., Gerber, E. P., & DallaSanta, K. J. (2022). Machine learning gravity wave parameterization generalizes to capture the QBO and response to increased CO2. *Geophysical Research Letters*, *49*(8), e2022GL098174. https://doi.org/10.1029/2022GL098174

Eyring, V., Butchart, N., Waugh, D. W., Akiyoshi, H., Austin, J., Bekki, S., et al. (2006). Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *Journal of Geophysical Research*, *111*(D22), D22308. https://doi.org/10.1029/2006JD007327

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

Fritts, D. C., Dong, W., Lund, T. S., Wieland, S., & Laughman, B. (2020). Self-acceleration and instability of gravity wave packets: 3. Three-dimensional packet propagation, secondary gravity waves, momentum transport, and transient mean forcing in tidal winds. *Journal of Geophysical Research: Atmospheres*, *125*(3), e2019JD030692. https://doi.org/10.1029/2019JD030692

Fritts, D. C., & Lu, W. (1993). Spectral estimates of gravity wave energy and momentum fluxes. Part II: Parameterization of wave forcing and variability. *Journal of the Atmospheric Sciences*, *50*(22), 3695–3713. https://doi.org/10.1175/1520-0469(1993)050%3C3695:SEOGWE%3E2.0.CO;2

Fritts, D. C., Lund, A. C., Lund, T. S., & Yudin, V. (2022a). Impacts of limited model resolution on the representation of mountain wave and secondary gravity wave dynamics in local and global models. 1: Mountain waves in the stratosphere and mesosphere. *Journal of Geophysical Research: Atmospheres*, *127*(9), e2021JD035990. https://doi.org/10.1029/2021JD035990

Fritts, D. C., Lund, A. C., Lund, T. S., & Yudin, V. (2022b). Impacts of limited model resolution on the representation of mountain wave and secondary wave dynamics in local and global models: 2. Mountain wave and secondary wave evolutions in the thermosphere. *Journal of Geophysical Research: Atmospheres*, *127*(9), e2021JD036035. https://doi.org/10.1029/2021JD036035

Fritts, D. C., Lund, T. S., Wan, K., & Liu, H.-L. (2021). Numerical simulation of mountain waves over the southern Andes. Part 2: Momentum fluxes, mean-flow accelerations, and gravity-wave/tidal interactions. *Journal of the Atmospheric Sciences*, 3069–3088. https://doi.org/10.1175/JAS-D-20-0207.1

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The whole atmosphere community climate model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12380–12403. https://doi.org/10.1029/2019JD030943

Goswami, S., Bora, A., Yu, Y., & Karniadakis, G. E. (2022). Physics-informed neural operators. https://doi.org/10.48550/arXiv.2207.05748

Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., & Catanzaro, B. (2021). Adaptive Fourier neural operators: Efficient token mixers for transformers. https://doi.org/10.48550/arXiv.2111.13587

Hertzog, A., Boccara, G., Vincent, R. A., Vial, F., & Cocquerez, P. (2008). Estimation of gravity wave momentum flux and phase speeds from Quasi-Lagrangian stratospheric balloon flights. Part II: Results from the vorcore campaign in Antarctica. *Journal of the Atmospheric Sciences*, *65*(10), 3056–3070. https://doi.org/10.1175/2008JAS2710.1

Hines, C. O. (1997). Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation. *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*(4), 371–386. https://doi.org/10.1016/S1364-6826(96)00079-X

Holton, J. R. (1982). The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *Journal of the Atmospheric Sciences*, *39*(4), 791–799. https://doi.org/10.1175/1520-0469(1982)039%3C0791:TROGWI%3E2.0.CO;2

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. https://doi.org/10.48550/arXiv.2010.08895

Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., et al. (2021). Physics-informed neural operator for learning partial differential equations. https://doi.org/10.48550/arXiv.2111.03794

Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical Research*, *86*(C10), 9707–9714. https://doi.org/10.1029/JC086iC10p09707

Liu, H. (2019). Quantifying gravity wave forcing using scale invariance. *Nature Communications*, *10*(1), 2605. https://doi.org/10.1038/s41467-019-10527-z

Lu, L., Jin, P., & Karniadakis, E. G. (2019). DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. https://doi.org/10.48550/arXiv.1910.03193

Lund, T. S., Fritts, D. C., Wan, K., Laughman, B., & Liu, H. (2020). Numerical simulation of mountain waves over the southern Andes. Part 1: Mountain wave and secondary wave character, evolutions, and breaking. *Journal of the Atmospheric Sciences*, *77*(12), 4337–4356. https://doi.org/10.1175/JAS-D-19-0356.1

Maziar, R., Paris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–770. https://doi.org/10.1016/j.jcp.2018.10.045

Miyoshi, Y., & Yiğit, E. (2019). Impact of gravity wave drag on the thermospheric circulation: Implementation of a nonlinear gravity wave parameterization in a whole-atmosphere model. *Annals of Geophysics*, *37*(5), 955–969. https://doi.org/10.5194/angeo-37-955-2019

Palmer, T. N., Shutts, G. J., & Swinbank, R. (1986). Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1001–1039. https://doi.org/10.1002/qj.49711247406

Pedatella, N. M., Fuller-Rowell, T., Wang, H., Jin, H., Miyoshi, Y., Fujiwara, H., et al. (2014). The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models. *Journal of Geophysical Research*, *119*(2), 1306–1324. https://doi.org/10.1002/2013ja019421

Ribstein, B., Millet, C., Lott, F., & de la Cámara, A. (2022). Can we improve the realism of gravity wave parameterizations by imposing sources at all altitudes in the atmosphere? *Journal of Advances in Modeling Earth Systems*, *14*(2), e2021MS002563. https://doi.org/10.1029/2021MS002563

Stephan, C., Alexander, M. J., & Richter, J. H. (2016). Characteristics of gravity waves from convection and implications for their parameterization in global circulation models. *Journal of the Atmospheric Sciences*, *73*(7), 2729–2742. https://doi.org/10.1175/JAS-D-15-0303.1

Vosper, S. B. (2015). Mountain waves and wakes generated by South Georgia: Implications for drag parameterization. *Quarterly Journal of the Royal Meteorological Society*, *141*(692), 2813–2827. https://doi.org/10.1002/qj.2566

Vosper, S. B., Brown, A. R., & Webster, S. (2016). Orographic drag on islands in the NWP mountain grey zone. *Quarterly Journal of the Royal Meteorological Society*, *142*(701), 3128–3137. https://doi.org/10.1002/qj.2894

Wandel, N., Weinmann, M., Neidlin, M., & Klein, R. (2022). Spline-PINN: Approaching PDEs without data using fast, physics-informed Hermite-spline CNNS. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36(8) pp. 8529–8538). https://doi.org/10.1609/aaai.v36i8.20830

Warner, C. D., & McIntyre, M. E. (1996). On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *Journal of the Atmospheric Sciences*, *53*(22), 3213–3235. https://doi.org/10.1175/1520-0469(1996)053%3C3213:OTPADO%3E2.0.CO;2

Xiong, W., Huang, X., Zhang, Z., Deng, R., Sun, P., & Tian, Y. (2023). Koopman neural operator as a mesh-free solver of non-linear partial differential equations. https://doi.org/10.48550/arXiv.2301.10022

Yiğit, E., Aylward, A. D., & Medvedev, A. S. (2008). Parameterization of the effects of vertically propagating gravity waves for thermosphere general circulation models: Sensitivity study. *Journal of Geophysical Research*, *113*(D19), D19106. https://doi.org/10.1029/2008JD010135