



RESEARCH ARTICLE

10.1029/2022MS003585

Key Points:

- In a library of weather research and forecasting model simulations, we compare methods for estimating 3D gravity wave drag force that are un- and under-resolved by general circulation models
- For drag associated with vertical fluxes, different methods agree on time- and zonal-mean but not on instantaneous spatiotemporal patterns
- Drag associated with horizontal fluxes is significant but is very sensitive to the estimation methodology

Supporting Information:

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

Correspondence to:

Y. Q. Sun,
ys91@rice.edu

Citation:

Sun, Y. Q., Hassanzadeh, P., Alexander, M. J., & Kruse, C. G. (2023). Quantifying 3D gravity wave drag in a library of tropical convection-permitting simulations for data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003585. <https://doi.org/10.1029/2022MS003585>

Received 14 DEC 2022

Accepted 5 MAY 2023

Author Contributions:

Conceptualization: Y. Qiang Sun, Pedram Hassanzadeh, M. Joan Alexander, Christopher G. Kruse
Data curation: Y. Qiang Sun, Christopher G. Kruse
Formal analysis: Y. Qiang Sun
Funding acquisition: Pedram Hassanzadeh, M. Joan Alexander
Investigation: Y. Qiang Sun

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Quantifying 3D Gravity Wave Drag in a Library of Tropical Convection-Permitting Simulations for Data-Driven Parameterizations

Y. Qiang Sun¹ , Pedram Hassanzadeh¹ , M. Joan Alexander² , and Christopher G. Kruse² 

¹Rice University, Houston, TX, USA, ²NorthWest Research Associates, Boulder, CO, USA

Abstract Atmospheric gravity waves (GWs) span a broad range of length scales. As a result, the un-resolved and under-resolved GWs have to be represented using a sub-grid scale (SGS) parameterization in general circulation models (GCMs). In recent years, machine learning (ML) techniques have emerged as novel methods for SGS modeling of climate processes. In the widely used approach of supervised (offline) learning, the true representation of the SGS terms have to be properly extracted from high-fidelity data (e.g., GW-resolving simulations). However, this is a non-trivial task, and the quality of the ML-based parameterization significantly hinges on the quality of these SGS terms. Here, we compare three methods to extract 3D GW fluxes and the resulting drag (Gravity Wave Drag [GWD]) from high-resolution simulations: Helmholtz decomposition, and spatial filtering to compute the Reynolds stress and the full SGS stress. In addition to previous studies that focused only on vertical fluxes by GWs, we also quantify the SGS GWD due to lateral momentum fluxes. We build and utilize a library of tropical high-resolution ($\Delta x = 3$ km) simulations using weather research and forecasting model. Results show that the SGS lateral momentum fluxes could have a significant contribution to the total GWD. Moreover, when estimating GWD due to lateral effects, interactions between the SGS and the resolved large-scale flow need to be considered. The sensitivity of the results to different filter type and length scale (dependent on GCM resolution) is also explored to inform the scale-awareness in the development of data-driven parameterizations.

Plain Language Summary Gravity waves (GWs) present a challenge to climate prediction: waves on scales of O(1)–O(100) km can neither be systematically measured with conventional observational systems, nor properly represented (resolved) in operational climate models, which have a typical grid spacing on the order of 100 km. Therefore, in these climate models, small-scale GWs must be parameterized, or estimated, based on the resolved (large-scale) flow. The primary effects of these small-scale waves on the resolved flow is the so-called sub-grid scale drag (Gravity Wave Drag [GWD]), resulting from the propagation and breaking of these waves. Existing GW parameterizations in general circulation models are all highly simplified; for example, they only account for vertical propagation of GWs. With growing computing power, a promising alternative approach is to use machine learning to develop data-driven parameterizations. However, this requires to first generate reliable high-resolution computer simulations and then extract GWD from these simulations. This study follows these steps, compares different extraction methods, and describes some challenges and pathways to make advances. Furthermore, our results suggest that the horizontal propagation of GWs should be included in parameterizations too, however, extra care is needed in order to extract the resulting GWD from high-resolution data.

1. Introduction

Atmospheric gravity waves (GWs), with horizontal scales from ~ 1 –1,000 km, play an important role in the transport of momentum from the surface and lower troposphere to the upper troposphere and middle atmosphere (Fritts & Alexander, 2003, and references therein). Once excited by various sources (e.g., convective systems, fronts, flow over topography), GWs propagate both vertically and laterally, transporting momentum and energy away from their sources (Bretherton, 1969; Fritts & Alexander, 2003; Palmer et al., 1986; Plougonven & Zhang, 2014). One challenge for climate and weather prediction is that the entire spectrum of GWs cannot be adequately resolved in current general circulation models (GCMs), which have a typical horizontal grid spacing of around 20–100 km (Eyring et al., 2016; Fritts & Nastrom, 1992; Gettelman et al., 2019). The effects of small-scale GWs are therefore parameterized based on the large-scale state of the atmosphere resolved by the

Methodology: Y. Qiang Sun, Pedram Hassanzadeh, Christopher G. Kruse
Project Administration: Pedram Hassanzadeh
Resources: Y. Qiang Sun, Pedram Hassanzadeh
Software: Y. Qiang Sun, Christopher G. Kruse
Supervision: Pedram Hassanzadeh, M. Joan Alexander
Validation: Y. Qiang Sun
Visualization: Y. Qiang Sun
Writing – original draft: Y. Qiang Sun, Pedram Hassanzadeh, M. Joan Alexander
Writing – review & editing: Y. Qiang Sun, Pedram Hassanzadeh, M. Joan Alexander, Christopher G. Kruse

GCM and other information of the sub-grid scale (SGS) sources. After decades of development, GW parameterization (GWP) is now a critical component of GCMs to enable them to reproduce realistic atmospheric circulation mean, variability, and response to climate change (e.g., Alexander et al., 2010; Beres, 2004; Y. Kim et al., 2003; Lott et al., 2012; McFarlane, 1987; Plougonven & Zhang, 2014; Richter et al., 2010; Scinocca & McFarlane, 2000). For example, generating quasi-biennial oscillation (QBO) in GCMs requires skillful GWPs (e.g., Richter et al., 2020).

Numerous assumptions are used in the current physics-based GWP schemes. For example, one typically used simplification is the “single-column approximation,” where the horizontal propagation of GWs is neglected so that all GWs stay in the same GCM column and will not directly affect the momentum and energy balance of neighboring columns. Other simplifications are also widely adopted, including but not limited to, steady-state approximation (neglecting of transient effects such as non-dissipative GW–mean-flow interactions), monochromatic and linear (ignoring potential triad wave-wave interactions), saturation assumption of GWs (limits the source and dissipation amplitudes), and assumptions of balanced (hydrostatic and geostrophic) resolved flows (Achatz et al., 2017; Bölöni et al., 2016; Wei et al., 2019). In addition to these assumptions, the representation of GW sources (e.g., small-scale convection) in GCMs is also challenging. Many efforts have been made in addressing these drawbacks of GWPs in state-of-the-art GCMs, for example, by relaxing some simplifications in more complex frameworks (Bölöni et al., 2021; Y. H. Kim et al., 2021). While adding realistic complexity to current physics-based GWPs improves their performances, more parameters are involved in general, which means additional tuning (Gettelman et al., 2019). The shortcomings of current GWPs is a major cause of uncertainties in future changes in stratospheric variability, most notably, the QBO, and the resulting surface impact (Richter et al., 2020, 2022; Sigmond & Scinocca, 2010).

Recently, Machine Learning (ML) techniques have emerged as alternative tools for developing parameterizations for climate models. They have been used in parameterizing a variety of SGS processes with promising results (e.g., Bolton & Zanna, 2019; Chattopadhyay et al., 2020; Gentine et al., 2021; Guan et al., 2022; Kashinath et al., 2021; Maulik et al., 2019; Rasp et al., 2018; Schneider et al., 2017; Yuval & O’Gorman, 2020). Matsuoka et al. (2020) were among the first to apply ML to GWs. Focused on the orographic GWs over the Hokkaido region of Japan, they trained a convolutional neural network to connect the large-scale tropospheric state and the small-scale GW wind fluctuations in the lower stratosphere. Recently, Amiramjadi et al. (2022) also found success in reconstructing the non-orographic GWs in the ERA5 data set with a random forest regressor. Both of these studies identified fluctuations associated with GWs using a simple moving-box average and demonstrated the feasibility of using ML to represent GWs. However, these studies only focused on learning GWs or momentum fluxes at one level (100 hPa), without further calculating the GWD, which is required to develop GWPs for GCMs. A number of other studies have also shown the power of ML for GWP through emulating current GWP schemes (Chantry et al., 2021; Espinosa et al., 2022). These emulation efforts provide valuable insight on various promises and challenges of using ML for GWPs, though a number of key challenges, for example, related to GWD extraction and lateral GW propagation, cannot be investigated through emulation (see below).

One key challenge for the data-driven approach is the availability of sufficient observationally constrained data of GW momentum transport for training the ML algorithms. With limited availability of observations of GWs and the challenges associated with sparsity and noise, high-resolution GW-resolving model simulations must play a critical role in generating the training data. A number of case studies have verified that high-resolution models are able to capture the key characteristics of observed GWs (Bramberger et al., 2020; Kruse et al., 2022). The second key challenge in the most common data-driven approach (the so-called “supervised” or “offline” learning) is the need to extract, from the high-resolution simulations, the true GWD due to the un- and under-resolved GWs. Hereafter, we refer to these collectively as the *SGS drag*. This SGS GWD is what has to be added to a low-resolution GCM to properly account for the un- and under-resolved GWs (note that this depends on the GCM’s *effective resolution*; more discussions to follow later). In the GW modeling community, a number of methods have been used in the past to separate GWs from the large-scale flow and quantify the SGS fluxes or GWD (e.g., Amiramjadi et al., 2022; Kruse & Smith, 2015; Matsuoka et al., 2020; Polichtchouk et al., 2022; G. J. Shutts & Vosper, 2011; Stephan et al., 2019; Žagar et al., 2015). Some studies pursued a simple box-averaging method (e.g., Matsuoka et al., 2020) or a sharp cut-off spectral filter (e.g., Polichtchouk et al., 2022). There are also more rigorous methods to separate the balanced large-scale components and the unbalanced GWs based on linear wave theory and Helmholtz decomposition (e.g., Callies et al., 2014; Žagar et al., 2015). Stephan et al. (2019) computed the resolved GW pseudo-momentum fluxes in month-long global convection-permitting simulations

with two other methods. These methods showed overall good agreement on the general shape of the longitudinal profiles of GW fluxes, but systematic differences were found for the amplitudes of the pseudo-momentum fluxes even after averaging over the 30-day period, implying the importance of the extraction method if we were to use these high-resolution data for training the ML algorithms.

The third key challenge is related to the 3D propagation of GWs and the resulting 3D SGS GWD. The aforementioned studies focused only on the vertical momentum fluxes of GWs, as the current operational GWP schemes ignore lateral propagation of the waves and the resulting lateral fluxes and their contribution to the total GWD. However, there is growing evidence that horizontal propagation of GWs has to be considered in GWP to produce a realistic atmospheric circulation (e.g., Ehard et al., 2017; Muraschko et al., 2015; Sato et al., 2009). Only few studies have discussed the lateral momentum fluxes in high-resolution simulations (Eckermann et al., 2015; Jiang et al., 2019). More recently, through analyzing the lateral momentum flux in the high-resolution simulations over the Drake Passage, Kruse et al. (2022) showed that the meridionally propagating mountain waves significantly enhanced the zonal drag. Additionally, their work suggested that not accounting for these meridional fluxes would result in GWD in the wrong direction at and below the polar night jet.

In this paper, we aim to quantify the 3D GWD from high-resolution simulations, a critical step for developing data-driven GWP schemes in future work. Using data from a library of 20 convection-permitting (3 km) tropical WRF simulations, we

1. Compare 3 methods that are commonly used in the GW and large-eddy simulation (LES) literature to quantify the SGS fluxes and drags. These methods are (a) Helmholtz decomposition, (b) Spatial filtering to compute the full SGS stress and the resulting GWD, and (c) Same as (b) but only for the Reynolds stress.
2. Quantify the contribution of both vertical and horizontal fluxes of horizontal momentum to the total GWD to investigate if the latter should be included in SGS parameterizations too.

Item 1 is crucial because any data-driven method, ML or otherwise, is as good as the data used for the training. Note that the challenges associated with extracting the SGS terms for ML training are not limited to the GW applications, and are in fact relevant to many climate/turbulence processes and currently an active area of research (e.g., Beck & Kurz, 2021; Grooms et al., 2021; Guan et al., 2022; Zanna & Bolton, 2021; Zhou et al., 2019). Moreover, understanding 3D GWD from high-resolution simulations could potentially help with improving the physics-based GWP schemes as well (Kruse et al., 2022).

Before moving to the next section, we highlight that a successful data-driven GWP for a typical low-resolution GCM is expected to represent the GWD missing in such a GCM compared to a GW-resolving model. This missing drag is a result of un-resolved and under-resolved GWs, which as mentioned earlier, we collectively refer to their drag as SGS GWD.

In the rest of the paper, we will first introduce the high-resolution data and the 3 methods used for SGS GWD extraction. We will then compare the results for the SGS vertical momentum flux and GWD, followed by similar analyses for the SGS GWD associated with lateral momentum fluxes. Concluding remarks and discussions are in the last section.

2. Data and Methods

2.1. Weather Research and Forecasting Model Setup and Data

All data used in this study are generated using WRF, with initial conditions from reanalysis data and boundaries nudged toward reanalysis data. For the purpose of this work, the WRF model is modified according to Kruse et al. (2022) to support a deep configuration that runs up to 1 Pa (~80 km). Two key modifications of the released WRF version 4.1 model are made to achieve the high model-top here. First, low-order interpolators are used to avoid the over-/under-shooting of default higher-order interpolators, preventing the intersection of analysis levels near complex topography after horizontal interpolation. Second, the default lateral relaxation is replaced with grid-point nudging confined to the lateral boundaries for the model to run stably (details of nudging are discussed later below).

For now, the library only includes the tropical regions (see the domains in Figure 1). We have conducted a total of 20 simulations in 6 domains, where the dates of the week-long runs are chosen to sample the seasonal cycle,

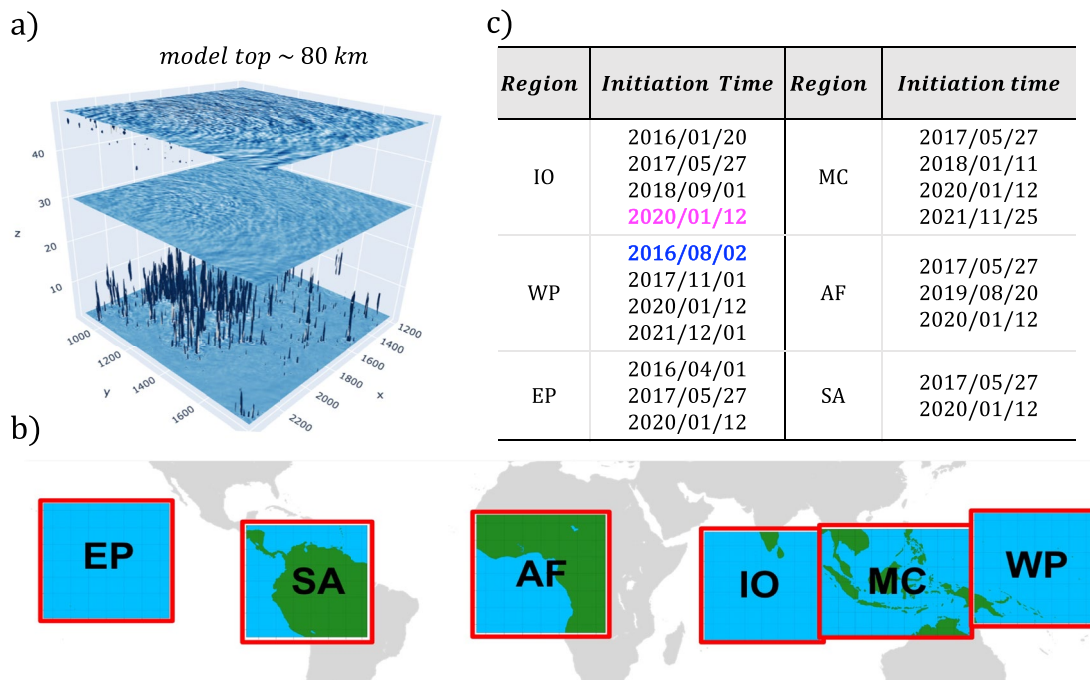


Figure 1. (a) A snapshot of vertical velocity in a weather research and forecasting model (WRF) simulation. (b) Tropical regional domains where the WRF simulations are conducted. (c) The list of 20 weeks-long WRF simulations conducted in this regional tropical channel setup, among which two representative cases are chosen (blue color represents Case 1, and magenta color represents Case 2). The two cases in two different regions are also chosen to have different quasi-biennial oscillation winds and different seasons.

QBO phases, and precipitation distribution (Figure 1c). Two of these simulations, one from the 2016 summertime all-ocean West Pacific, which is in the westerly phase of QBO, and one from the 2020 wintertime land-ocean Indian Ocean, which is in the easterly phase of QBO, are chosen as representative cases 1 and 2, respectively. The first day of all simulations is treated as spin-up periods and not used in analyses. The horizontal domain size is $3,600 \times 3,600$ km, based on a tradeoff between the minimum size needed for meaningful quantification of lateral propagation and maximum size allowed given the computing capabilities. The simulations are done at 3 km grid spacing. There are 180 vertical model levels in total, with a stretched grid. The vertical grid spacing is less than 50 m near the lower boundary and gradually increases to a maximum of 500 m near the model top. For these tropical simulations, we largely use the “Tropical” WRF physics suite (e.g., Qiao et al., 2019), but with a different surface layer scheme. The parameterization set includes the WRF Single-Moment 6-class (WSM6) microphysics scheme (Hong & Lim, 2006), the Yonsei University planetary boundary layer scheme (Hong et al., 2006), the Rapid Radiative Transfer Model for longwave and shortwave radiation (Iacono et al., 2008; Pincus et al., 2003), and the revised surface layer scheme developed in Jiménez et al. (2012). Note that no cumulus scheme is used given the 3 km grid spacing and, most importantly, no GWP is used.

Both the initial condition and the boundary condition come from the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis data (ERA5). We use the same method as Kruse et al. (2022) to nudge the simulation boundaries toward the ERA5 data: all grid points within 150 km of the boundary are nudged toward the ERA5 data at a time scale of 3 hr, hence limiting the GW signals there. Therefore, the model output data near the domain boundary (<300 km) are neglected when conducting analysis for the GWs. The analysis domains are hence $3,000 \times 3,000$ km.

In addition to traditional prognostic variables (e.g., u , v , w , T , p , q), we also modify the WRF model to output diagnostic variables such as 3D reflectivity and 3D diabatic heating, which are the key sources for the GWs in the tropics. The output frequency is every 15 min in order to capture the life cycle of the convective cells.

Table 1
The 2D Low-Pass Spatial Filters Used in This Study (Equation 1)

Filter	Kernel (physical space) $G(\mathbf{r}, \Delta)$	Transfer function (spectral space) $\hat{G}(\mathbf{k}, \Delta) = \int_{-\infty}^{\infty} e^{i2\pi\mathbf{k}\cdot\mathbf{r}} G(\mathbf{r}, \Delta) d\mathbf{r}$	Length-scale (km) Δ
Gaussian	$\frac{6}{\pi\Delta^2} \exp\left(-\frac{6 \mathbf{r} ^2}{\Delta^2}\right)$	$\exp\left(-\frac{ \mathbf{k} ^2\Delta^2}{24}\right)$	700 or 200
Top-hat (box)	$\begin{cases} \frac{1}{\Delta^2}, & \text{if } (r_x, r_y) \leq \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases}$	$\frac{\sin(\frac{1}{2}k_x\Delta)\sin(\frac{1}{2}k_y\Delta)}{(\frac{1}{2}k_x\Delta)(\frac{1}{2}k_y\Delta)}$	700 or 200
Sharp-spectral	$\frac{\sin(\frac{\pi r}{\Delta})}{\pi r}$	$\begin{cases} 1, & \text{if } (k_c - \mathbf{k} \geq 0), k_c = \frac{\pi}{\Delta} \\ 0, & \text{otherwise} \end{cases}$	700 or 200

Note. All filters are implemented in spectral space using their transfer function (e.g., Guan et al., 2022). Here, $\mathbf{r}$ and $\mathbf{k}$ are coordinates in physical space and spectral space, respectively, with $\mathbf{r} = (r_x, r_y)$, and $\mathbf{k} = (k_x, k_y)$. $\hat{(\cdot)}$ is the Fourier transform, and Δ is the filter size as in Equation 1.

2.2. Filtering and Coarse-Graining

Before introducing the 3 GWD extraction methods, we first discuss two operations that are essential for almost any data-driven SGS modeling method: (a) spatial filtering, denoted with $\tilde{(\cdot)}$ and (b) coarse-graining, denoted with $\bar{(\cdot)}$. For any variable $\phi(\mathbf{x}, t)$, spatial filtering is defined as (e.g., Grooms et al., 2021; Guan et al., 2022; Sagaut, 2006)

$$\tilde{\phi}(\mathbf{x}, t) = G * \phi = \int_{-\infty}^{\infty} G(\mathbf{r}, \Delta) \phi(\mathbf{x} - \mathbf{r}, t) d\mathbf{r}, \quad (1)$$

where G is the filter's kernel, Δ is the filter's length scale, $*$ is the convolution operator, and the integration is performed over the entire domain. Table 1 presents a list of commonly used 2D low-pass spatial filters. Then, any variable $\phi(\mathbf{x}, t)$ can be separated into two components

$$\phi = \tilde{\phi} + \phi' \quad (2)$$

where $\tilde{\phi}$ contains the large scales (larger than Δ) and ϕ' contains the small scales (smaller than Δ).

Two key points need to be clarified here. One is that following the convention used in recent literature (e.g., Grooms et al., 2021; Guan et al., 2022; Sagaut, 2006), we define “filtering” as an operation that only separates the scales but does not change the grid resolution (e.g., all 3 terms in Equation 2 remain on the high-resolution grid). “Coarse-graining,” defined later in this section, is the operation that changes resolution, for example, from the WRF's high-resolution to a GCM's low-resolution grid.

Second, it should be highlighted that Equation 2, while appears analogous, is not the same as Reynolds decomposition in this application. This is because spatial filtering (Equation 1) is different from Reynolds averaging; unlike the latter, here, $\tilde{\phi}' \neq 0$ and $\tilde{\tilde{\phi}} \neq \tilde{\phi}$ depending on the choice of the filter function (e.g., Alfonsi, 2009; Clark et al., 1979; Leonard, 1975; Sagaut, 2006). For example, for sharp cut-off spectral filters, $\tilde{\phi}' = 0$ and $\tilde{\tilde{\phi}} = \tilde{\phi}$. However, for other common filters such as Gaussian or Box, these equalities do not hold. The importance of the difference between spatial filtering and Reynolds averaging has been recently pointed out in a number of other studies too; for example, it has been shown to play a significant role in understanding momentum exchange between atmosphere and ocean at small scales (e.g., Aluie et al., 2018; Rai et al., 2021).

A major question in using Equation 2 is the choice of filter type and size (length scale, Δ in Equation 1). As described below, in the 3 methods used here (and generally, in many other methods), Equation 2 might be used to separate GWs from the large-scale flow, or to separate the un-resolved and under-resolved GWs from the resolved GWs, or both. The choice of filter type (e.g., Gaussian, top-hat or box, sharp-spectral) can affect the extracted SGS terms, as already shown in a number of past studies including in the context of geophysical turbulence (e.g., Leonard, 1975; Zanna & Bolton, 2021; Beck & Kurz, 2021). Figure B1 shows an example of the effect of

filter type on the spectrum of zonal wind from our WRF simulations. Different low-pass filters (e.g., top-hat and sharp-spectral) have been used in previous studies to separate the GWs from the large-scale background (Kruse & Smith, 2015; Matsuoka et al., 2020; Polichtchouk et al., 2022), though a systematic study on the effect of filter type and the potential implications for the extracted SGS terms is lacking.

The question about filter size Δ is even more challenging when it comes to systems without clear scale separation. While the (low) resolution of the GCMs provides a clear length scale, the issue of “effective resolution” makes this further complicated. Even in a GCM with grid spacing dx , GWs with wavelength larger than $2dx$ are not fully resolved, depending on the specifics of the numerical schemes used in the dynamical core of the targeted GCM. Skamarock (2004), through computing kinetic energy spectra, demonstrated that in WRF, GWs with scales up to $7dx$ remain under-resolved. There are also additional complications. For example, Stephan et al. (2022) argued that the separation scale Δ for balanced and unbalanced motions, based on partitioning of total wave energies, varies with height. Finally, more complications arise on non-uniform grids (e.g., Aluie et al., 2018; Grooms et al., 2021), though this is not a problem in the current study as WRF's grid is uniform.

To systematically quantify the effects of filter type and size, here, we use 3 filter types and two length scales $\Delta = 200$ and 700 km to help with understanding the scale-awareness when building a data-driven GWP in the future (Table 1). Note that these choices of Δ are motivated by assuming that the low-resolution GCM has grid spacing of 100 km ($\sim 1^\circ$ resolution). $\Delta = 200$ km is based on the common choice for Δ in the LES literature, that is, twice the low-resolution model's grid spacing (Guan et al., 2022; Pope, 2000; Sagaut, 2006). $\Delta = 700$ km is based on the effective-resolution study of Skamarock (2004); this is the filter size used for the presented results, unless indicated otherwise.

Once resolved fluxes are quantified point-wise on the original grid, the effective fluxes within a hypothetical GCM grid cell must be computed on a coarse GCM grid. As mentioned before, we refer to this operation as coarse-graining. Admittedly, this terminology has not been uniformly adopted in the literature, though it has been recommended by several recent studies (Grooms et al., 2021; Guan et al., 2022). Also, note that in some studies filtering and coarse-graining are done via just one operation, rather than two separate ones (e.g., Brenowitz & Bretherton, 2018; Yuval & O’Gorman, 2020). With all these issues in mind, here, we employ one commonly used coarse-graining strategy: we simply truncate the wavenumbers greater than the cut-off wavenumber k_{cg} corresponding to the GCM grid spacing (100 km in this case). In other words, for any variable (or flux term) $\phi(\mathbf{x}, t)$, we first compute the Fourier transform ($\hat{\phi}$), then do a truncation in the spectral space to derive the Fourier transform of the coarse-grained variable (or flux term) $\bar{\phi}$:

$$(\bar{\phi}) = (\hat{\phi})_{|k_x| < k_{cg}, |k_y| < k_{cg}}, \quad (3)$$

where $k_{cg} = \pi/\Delta_{GCM}$ and Δ_{GCM} is the targeted GCM grid spacing (e.g., 100 km). An inverse Fourier transform is then conducted to get $\bar{\phi}$.

Note that in this study, for computational efficiency, both filtering and coarse-graining are done in the spectral (Fourier) space, and mirrored tiles are added around the original domain following Sun and Zhang (2016) to reduce problems with non-periodic boundaries.

To better illustrate the effects of these filtering and coarse-graining operations, Figure 2 shows examples of the high-resolution WRF snapshots, and filtered (Gaussian with $\Delta = 700$ km) and coarse-grained 3D velocity fields at 30 km height. The full u , v , w in the $3,000 \times 3,000$ km domain are shown in the left column. After the filtering operation, the velocity fields are separated into the large-scale (second column) and the perturbation (third column) components. We also apply coarse-graining operators to these fields (fourth and fifth columns) to transfer them to a 30×30 grid, similar to that of a GCM with a grid spacing of 100 km. From this plot, we notice systematic differences between horizontal winds and vertical winds. For the horizontal winds u and v , the large-scale background ($\bar{u}$ and $\bar{v}$) are much larger in amplitudes than the small-scale perturbations u' and v' , whereas for the vertical velocity, the large-scale background is almost negligible, with all the signal at small scales w' . Moreover, notice that there can be significant differences between $\bar{\phi}$ and $\bar{\phi}'$, which implies that $\bar{\phi}' \neq 0$ with the Gaussian filtering applied here.

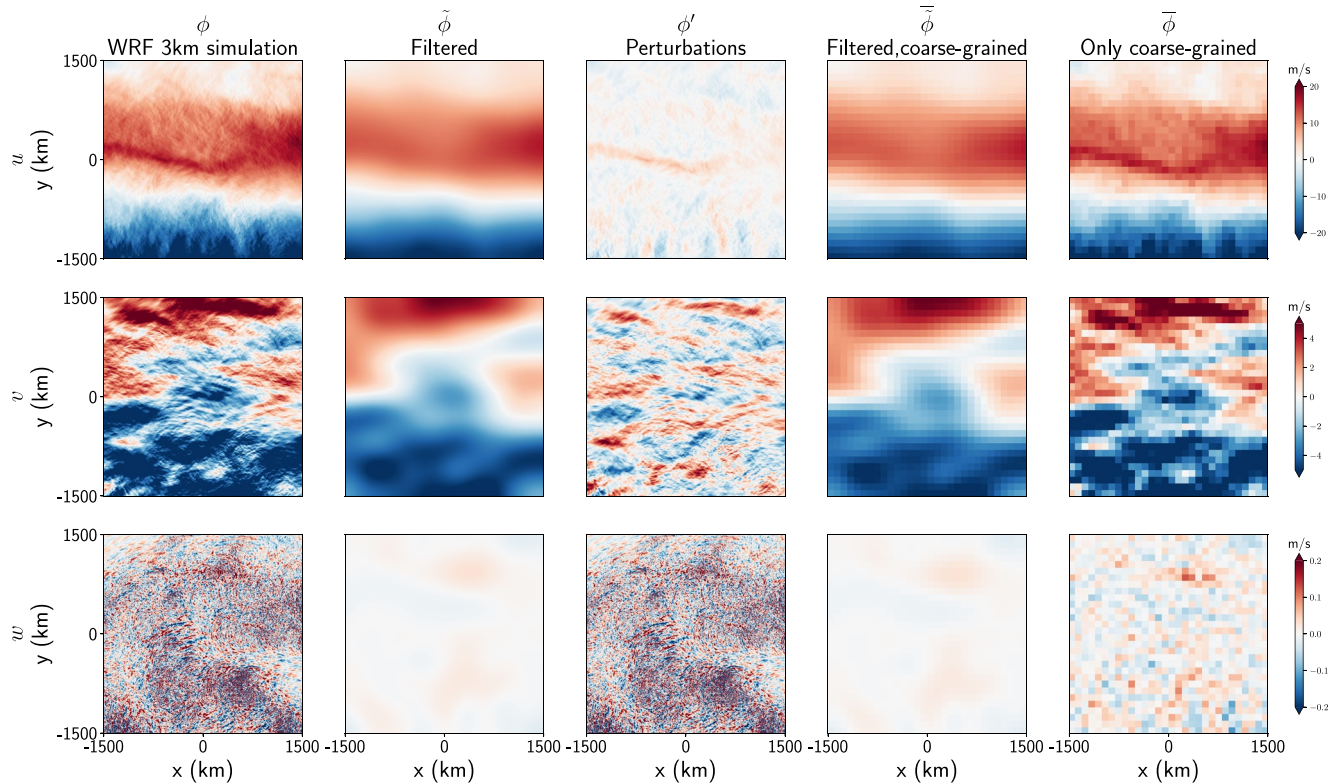


Figure 2. Examples of the effects of low-pass filtering and coarse-graining operators used in this study. The snapshots are from Case 1, on 3 August 2016, 12:00 UTC, at 30 km height. A Gaussian filter with $\Delta = 700$ km is used. The coarse-graining is done by truncating all wavenumbers greater than that corresponding to the 100-km grid. ϕ here can be either u , v , or w .

2.3. Gravity Wave Drag Extraction

The three methods used for SGS GWD extraction are:

1. *Helmholtz decomposition method*: This is a common practice in the GW literature for separating GWs from the background flow (e.g., Callies et al., 2014; Wei et al., 2022). The key idea is to divide the full flow into divergent and rotational components using Helmholtz decomposition, and then assume that the divergent component entirely consist of GWs. The drag derived using this method will be referred to as HELM_D.
2. *Un- and under-resolved SGS (UUGS) method*: This is the common practice in the LES community for computing the fluxes or drags that need to be parameterized (e.g., Germano, 1992; Leonard, 1975; Sagaut, 2006), and has been successfully used to provide training data for ML techniques for SGS modeling of a variety of geophysical flows (e.g., Guan et al., 2022; Maulik et al., 2019; Subel et al., 2023; Yuval & O’Gorman, 2020; Zanna & Bolton, 2020). The key idea here is to use spatial filtering and a rigorous mathematical derivation of the SGS terms. The drag derived using this method will be referred to as UUGS_D.
3. *Reynolds stress method*: This is an approach that has been used in both GW and LES communities (e.g., Amiramjadi et al., 2022; Clark et al., 1979; Kruse & Smith, 2015; Polichtchouk et al., 2022), and bears similarities to both Helmholtz decomposition and UUGS methods. The drag derived using this method will be referred to as REYN_D.

Next, we introduce these three methods in more details.

2.3.1. Method 1: Helmholtz Decomposition Method

Using Helmholtz decomposition to compute the divergent and rotational components of a global wind field has been well studied for decades (e.g., Chen & Wiin-Nielsen, 1976). However, for regional domains such as those of our WRF simulations, the Helmholtz decomposition is not uniquely defined, and boundary conditions must be imposed to obtain a unique solution (e.g., Lynch, 1988; Skamarock & Klemp, 2008). Therefore, how we provide the boundary conditions for the Helmholtz decomposition solver could affect the results

(e.g., Cao et al., 2014). As we are nudging our WRF simulations toward ERA5 reanalysis data, the following novel procedure is proposed to avoid the boundary-condition dependency for the decomposition of our WRF simulations:

1. First, the high-resolution WRF data are regridded to the 0.25° ERA5 grid within the WRF domain using conservative interpolation.
2. Outside the WRF domain, we fill the global 0.25° grid with ERA5 reanalysis data at the same time as WRF outputs to construct a “synthetic” global field. Linear interpolation of ERA5 reanalysis data is used if WRF outputs are at different times/levels compared to the reanalysis.
3. Helmholtz decomposition using a widely employed public function (<https://www.ncl.ucar.edu/Applications/wind.shtml>) is applied to the newly constructed “global fields” to get the global rotational and divergent wind components. No boundary condition is needed in this approach. See Figure S1 in Supporting Information S1 for an example of the global field and its rotational and divergence components.
4. The derived global rotational wind components are then linearly interpolated back to the high-resolution WRF grid. This now serves as the large-scale background for the simulated flow.
5. The divergent winds, mostly GWs, are then defined as deviations of the full flow in WRF simulations from the large-scale background we get in (4). Given that the divergent winds could contain large-scale Kelvin waves in the tropics, and that these waves and large-scale GWs could be resolved by the GCM, an additional high-pass spatial filter is applied to remove the signals that could be fully resolved by the low-resolution GCM grid. Here, we use a Gaussian high-pass filter with filter size of 700 km.

The outcome of step (5) is the GW perturbations for the horizontal winds (u'_H and v'_H). The subscript H here denotes the use of Helmholtz decomposition in the process. The vertical winds w are not involved here in the Helmholtz decomposition of horizontal winds. Given that vertical wind w is dominated by small scales and has negligible large-scale signals (Figure 2), we simply apply a high-pass filter as in (5) to the full fields to get w' .

The 3D zonal momentum flux due to SGS GWs is then defined as:

$$\mathbf{MF}_x = [MF_{xx}, MF_{yx}, MF_{zx}] \quad (4)$$

where

$$MF_{xx} = \tilde{\rho} u'_H \tilde{u}'_H \quad (5)$$

$$MF_{yx} = \tilde{\rho} u'_H \tilde{v}'_H \quad (6)$$

$$MF_{zx} = \tilde{\rho} u'_H \tilde{w}'. \quad (7)$$

Note that while the Helmholtz decomposition separate the GWs and their fluxes, for the purpose of data-driven SGS modeling, we still need to further separate the SGS (un- and under-resolved) component. Here, in step (5), this is done using spatial filtering, which is also the approach used by Kruse and Smith (2015) in their analyses of GWs.

The first two components in Equation 4 are the zonal and meridional flux of zonal momentum due to SGS GWs, respectively. They will also be referred to as lateral momentum fluxes. The last component in Equation 4 is the vertical flux of zonal momentum due to SGS GWs.

As mentioned earlier, for the purpose of training a data-driven parameterization that could be coupled to a low-resolution GCM, momentum fluxes derived in Equation 4 need to be further coarse-grained to the targeted GCM grid. We note here again that the filtering of MF components with $\Delta = 700$ km, then coarsening to the 100 km GCM grid, is a way to include phase-averaged fluxes from GWs with horizontal scales that is under-resolved by the 100 km GCM. The 3D SGS zonal momentum fluxes then become $\overline{\mathbf{MF}}_x$,

$$\overline{\mathbf{MF}}_x = [\overline{MF}_{xx}, \overline{MF}_{yx}, \overline{MF}_{zx}]. \quad (8)$$

Based on Equation 8, the zonal SGS GWD after coarse-graining ($\overline{GW D_x} = \overline{GW D_{xx}} + \overline{GW D_{yx}} + \overline{GW D_{zx}}$), which is what needed to train a data-driven GWP, can be calculated as the divergence of $\overline{MF_x}$:

$$\overline{GW D_{xx}} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF_{xx}}}{\partial x} \quad (9)$$

$$\overline{GW D_{yx}} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF_{yx}}}{\partial y} \quad (10)$$

$$\overline{GW D_{zx}} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF_{zx}}}{\partial z}. \quad (11)$$

Note that Equation 11, the vertical divergence of the vertical flux of zonal momentum due to SGS GWs, is often considered to be the dominant component in previous studies, and the only term that is conventionally represented in existing SGS parameterizations. This has been the case in the development of physics-based GWP, and in the past efforts focused on extracting SGS GWD from high-resolution simulations (e.g., Alexander et al., 2010; Matsuoka et al., 2020; Polichtchouk et al., 2022). Yet, as shown in Kruse et al. (2022), this is not always the case, and the horizontal divergence of lateral momentum fluxes (Equations 9 and 10) could also play a substantial role, as will be also shown here later in the Results section.

2.3.2. Method 2: Un- and Under-Resolved Sub-Grid Scale Method (UUGS_D)

One can quantify the missing drag in a low-resolution GCM compared to a high-resolution GCM by filtering and coarse-graining of the governing equations of the latter, following the common practice in LES (e.g., Pope, 1975; Sagaut, 2006). Details of such derivation for zonal momentum are presented in Appendix A. This analysis shows that for example, the zonal SGS GWD is

$$\begin{aligned} \overline{GW D_x} &= \overline{GW D_{xx}} + \overline{GW D_{yx}} + \overline{GW D_{zx}} \\ &= -\frac{1}{\bar{\rho}} \frac{\partial}{\partial x} \left[\bar{\rho} (\overline{u'u} - \bar{u}\bar{u}) \right] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial y} \left[\bar{\rho} (\overline{u'v} - \bar{u}\bar{v}) \right] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} \left[\bar{\rho} (\overline{u'w} - \bar{u}\bar{w}) \right]. \end{aligned} \quad (12)$$

Note that the SGS GW momentum fluxes here can be interpreted as the difference between the filtered and coarse-grained flux in high-resolution simulations and the flux a coarse-resolution GCM would give based on the filtered and coarse-grained prognostic variables (see Appendix A).

Similar to Equations 9–11, the zonal SGS GWD in Equation 12 also has three components that are associated with SGS zonal, meridional, and vertical fluxes of zonal momentum, respectively, though here these components involve full fields rather than perturbations. However, using Equation 2 for each component of the velocity vector, we can see that a Reynolds stress is one of the three components of each term in Equation 12. For example, as shown in Equation A13, the Reynolds stress $\overline{u'w'}$ is a part of the (but not the entire) total SGS vertical flux, $\overline{u'w} - \bar{u}\bar{w}$. The other two components (e.g., $\overline{u'u}$) arise because as mentioned before, in spatial filtering and coarse-graining, terms like $\overline{u'w'}$ are not necessarily zero (e.g., Pope, 2000; Sagaut, 2006). Similar analysis can be done for $\overline{GW D_{xx}}$ and $\overline{GW D_{yx}}$, showing the appearance of Reynolds stresses $\overline{u'u'}$ and $\overline{u'v'}$ as well as other stresses, including $\overline{u'u}$ and $\overline{u'v}$. Different from the HELM_D method that only considers direct contributions of SGS GW perturbations to the GWD, the UUGS_D method (Equation 12) also includes the cross-scale interactions between the SGS GWs and the resolved large-scale flow, which is also missing in the low-resolution GCMs (see the derivation in Appendix A).

2.3.3. Method 3: Reynolds Stress Method

In this approach, the three components of $\overline{MF_x}$ are computed similar to a number of past studies (Amiranjadi et al., 2022; Kruse et al., 2016; Matsuoka et al., 2020); hence, the components of $\overline{GW D_x}$ can be written as

$$\overline{GW D_{xx}} = -\frac{1}{\bar{\rho}} \frac{\partial (\bar{\rho} \overline{u'u'})}{\partial x} \quad (13)$$

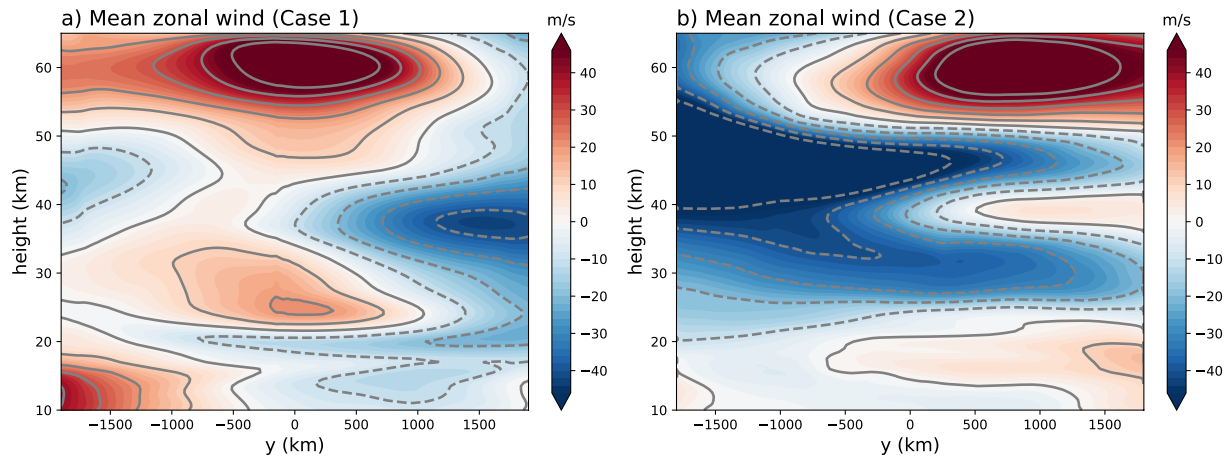


Figure 3. Zonal-mean (weather research and forecasting model [WRF] domain only), time-mean zonal wind u for two representative WRF simulation cases. The contour interval is 10 m, with dash lines implying zero and negative values. (a) Case 1 with westerly quasi-biennial oscillation (QBO) winds; (b) Case 2 with easterly QBO winds.

$$\overline{GW D_{yx}} = -\frac{1}{\bar{\rho}} \frac{\partial (\bar{\rho} \overline{u'v'})}{\partial y} \quad (14)$$

$$\overline{GW D_{zx}} = -\frac{1}{\bar{\rho}} \frac{\partial (\bar{\rho} \overline{u'w'})}{\partial z}, \quad (15)$$

though often only $\overline{GW D_{zx}}$ is considered. There are two ways to interpret these equations. First, one can obtain Equations 13–15 if only the Reynolds stresses in Equation 12 are accounted for, and the other stresses, including cross-scale interactions are ignored. Second, Equations 13–15 are the same as Equations 9–11 if the GW perturbations are identified using filtering (e.g., as $u' = u - \bar{u}$) rather than as the divergent component of the wind field.

3. Results

3.1. Zonal-Mean, Time-Mean Zonal Wind in the Weather Research and Forecasting Model Simulations

Figure 3 shows the zonal-mean zonal wind averaged over the 5-day simulation period in the two representative cases (see Figure 1c). The QBO winds are clear in both cases (different phases), with a maximum of ~ 20 m/s in the tropical stratosphere (~ 25 km). The semiannual oscillation (SAO) can also be seen near the stratopause (~ 60 km), with a much stronger wind amplitude than the QBO. The existence of westerly winds in the QBO and SAO at the Equator means they have greater angular momentum than that of the rotating Earth. This “superrotation” cannot be explained by direct thermal forcing or symmetric circulations, but must arise from the effects of wave forcing. In our following analysis, we will mainly examine the zonal SGS GWD, as both QBO and SAO are mostly zonal circulations.

3.2. Vertical Flux of Zonal Momentum Due To Sub-Grid Scale Gravity Waves

While GWs propagate both vertically and horizontally once excited, it is believed that the GWD due to the vertical fluxes are dominant and hence the single-column approximation is used in most GCMs. Here, we first examine the GWD due to SGS vertical fluxes of zonal momentum. For the representative cases, the zonal-mean, time-mean zonal SGS GWD associated with vertical fluxes is shown in Figure 4. The left column shows zonal SGS GWD calculated using Equation 11 with the HELM_D method. The zonal SGS GWD in the middle column is based on the REYN_D method (Equation 15), where the GW perturbations are derived with a low-pass Gaussian filter ($\Delta = 700$ km) in Table 1. The right column is the zonal SGS GWD calculated using the UUGS_D method (last term of Equation 12) with the same low-pass Gaussian filter.

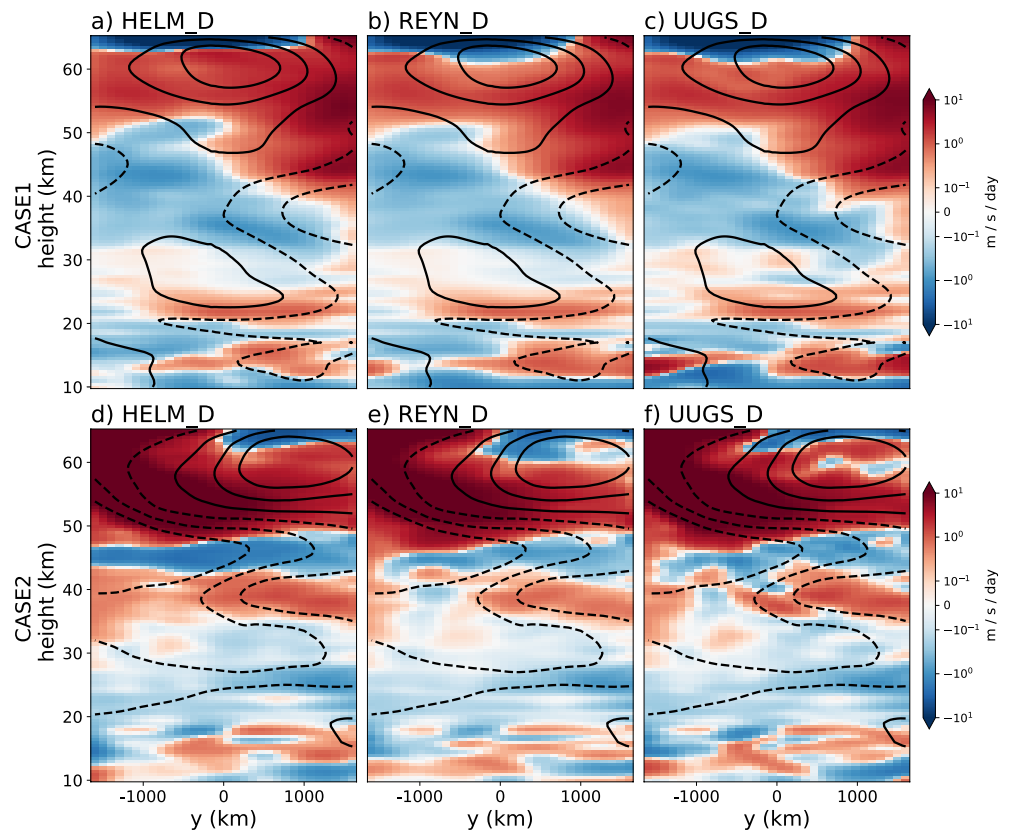


Figure 4. Zonal-mean (weather research and forecasting model domain only), time-mean zonal sub-grid scale (SGS) Gravity Wave Drag (shading) due to the un-/under-resolved (SGS) vertical fluxes derived using three methods for both representative cases (upper row: Case 1, lower row: Case 2). Left column, (a and d): Helmholtz decomposition method (HELM_D). Middle column, (b and e): Reynolds stress method (REYN_D). Right column, (c and f): Un- and under-resolved sub-grid scale method. As in Figure 3, lines show the mean zonal winds, but with a contour interval of 20 m/s.

We can see that for these zonally averaged time-mean GWD patterns, all methods give fairly consistent results. This supports the simplifications made in many previous studies that only consider the Reynolds stress term as in Equation 15 when they estimated the GWD. The agreement between HELM_D and the two Gaussian filter-based methods also shows that the mean zonal SGS GWD associated with vertical fluxes is not very sensitive to the methods used for separating the GWs and the large-scale background flow. We also notice that at the upper stratosphere, close to the SAO region, the GWD is mostly positive (negative) when the zonal wind shear is positive (negative), showing that vertically propagating SGS GWs help maintain and drive the zonal wind there (Alexander et al., 2010). Moreover, for the QBO region, the maximum drag is below the wind maximum (e.g., Case 1 in Figure 4), implying the role of SGS GWs in the downward propagation of the zonal winds.

While the mean zonal SGS GWD is the most important factor for maintaining the time-mean, zonal-mean momentum budget (hence the QBO and SAO), we need instantaneous snapshots of SGS GWD over the whole domain for developing data-driven GWP schemes. However, the picture is very different if we examine the zonal SGS GWD for each GCM column calculated based on different methods at a randomly chosen time. Figure 5 shows two horizontal snapshots in Case 1 and Case 2 for the SGS vertical fluxes of zonal momentum at 30 km (QBO region) with the same methods used in Figure 4. While the SGS vertical fluxes of zonal momentum estimated using HELM_D and the REYN_D methods might show some similarities, they significantly differ from what we find using the UUGS method. The UUGS method in general gives stronger amplitude for the GWD. Also, additional spatial variability not seen by the HELM_D and REYN_D methods can be found in the SGS GWD extracted using the UUGS method.

Figure 6 shows the probability density functions (PDFs) for the zonal SGS GWD associated with vertical fluxes using these three methods, as another way of presenting the differences among them. While the PDFs

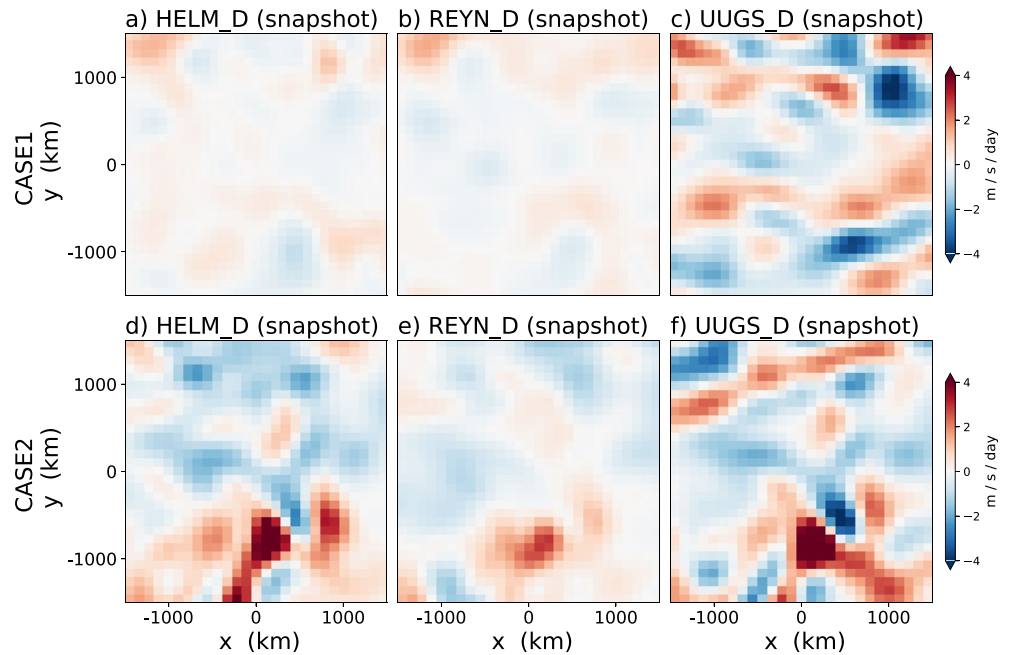


Figure 5. Snapshots of zonal sub-grid scale Gravity Wave Drag due to vertical fluxes at 30 km height calculated using three methods. Upper row: Case 1. Lower row: Case 2. Left column, (a and d): Helmholtz decomposition method $-\frac{1}{\bar{\rho}} \frac{\partial \bar{\rho}' u' w'}{\partial z}$. Middle column, (b and e): Reynolds stress method $-\frac{1}{\bar{\rho}} \frac{\partial \bar{\rho}' u' w'}{\partial z}$. Right column, (c and f): Un- and under-resolved sub-grid scale method $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial z} \left[\bar{\rho} (\bar{u} \bar{w} - \bar{u} \bar{w}) \right]$.

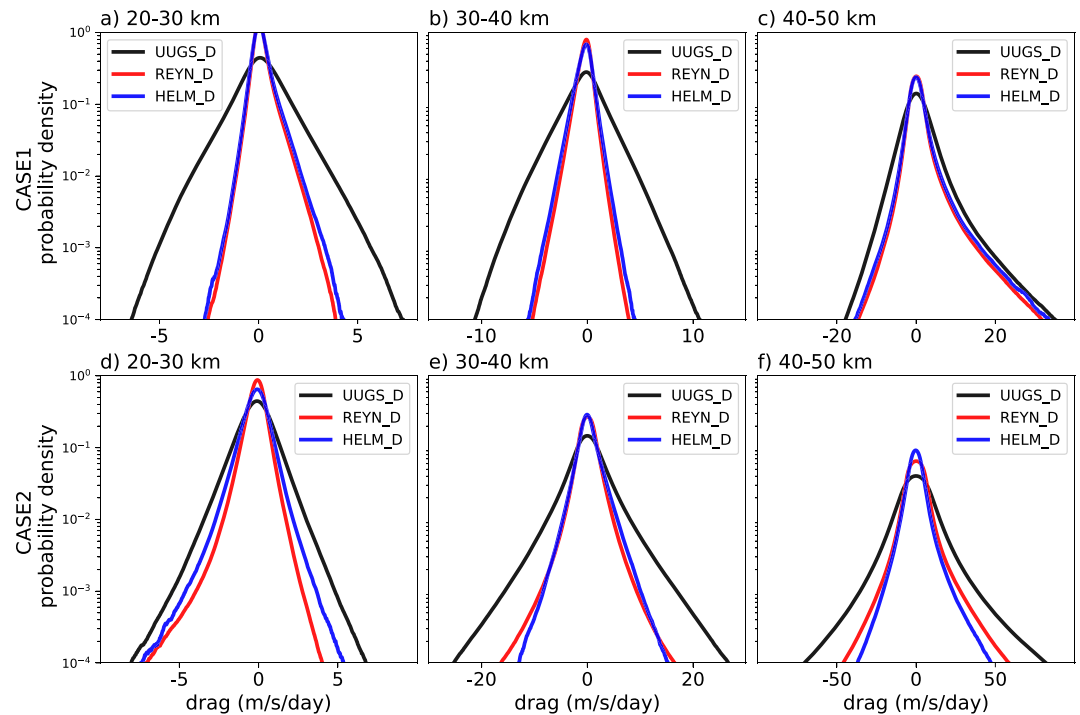


Figure 6. Probability density function (PDF) of zonal sub-grid scale (SGS) Gravity Wave Drag (GWD) due to vertical fluxes calculated using three methods at different heights. Upper row: Case 1. Lower row: Case 2. Left column, (a and d): 20–30 km. Middle column, (b and e): 30–40 km. Right column, (c and f): 40–50 km. Note the differences in the x-axes. See Figure S2 in Supporting Information S1 for PDFs of the meridional SGS GWD due to vertical fluxes. See Figure S3 in Supporting Information S1 for the PDFs of data from all 20 cases combined.

we obtain using the HELM_D and REYN_D methods are fairly similar, the PDFs from the UUGS method often have a wider distribution. For both cases studied here, we find that the GWD from the UUGS method in general has higher variability, both temporal and spatial, compared with the other methods, that are based on the Reynolds stress term alone (e.g., Equation 15), which is one of the three components of the total stress estimated in the UUGS method (see Appendix A). The degree of differences in PDFs depend on the case and height. For example, the differences are smaller at 40–50 km for Case 1, possibly due to the weak zonal winds there (Figure 4).

So far, we have discussed the two representative cases. The same conclusions are reached if we examine the other cases, or all cases together. Figure S3 in Supporting Information S1 is the same as Figure 6, but with data from all 20 cases combined.

To sum up, for the zonal SGS GWD due to vertical fluxes, the 3 methods studied here provide fairly consistent time-mean, zonal-mean results. However, to develop data-driven GWP schemes, we need snapshots of GWD at specific time and locations. For such snapshots, the GWD extracted using the UUGS_D method has additional spatial and temporal variability, compared to the GWD from the other two methods that are based on the Reynolds stress alone. One reason for this difference is that the UUGS_D method accounts for more components of the stress that represent the interactions between the missing GWs and large-scale background, which are mostly ignored in the HELM_D and REYN_D methods. Whether this additional variability would be efficiently learned using the ML algorithm and help improve the performance of the targeted GCM should be carefully investigated in future studies (see Section 4 for further discussions).

3.3. Horizontal Flux of Zonal Momentum Due To Sub-Grid Scale Gravity Waves

In addition to the SGS vertical fluxes, the SGS horizontal momentum fluxes associated with GWs could also lead to zonal SGS GWD (see Equations 12–14). However, these horizontal fluxes have received much less attention in previous studies and are totally neglected in most GCMs' GWP schemes with the single-column approximation. In recent years, ignoring the lateral propagation of GWs has been recognized as a key weakness of state-of-the-art GWP schemes. Yet, quantitative studies on the importance of SGS horizontal fluxes have been limited to a few case studies (e.g., Kruse et al., 2022; G. J. Shutts & Vosper, 2011). With all 3 methods introduced in Section 2, we can also calculate the SGS horizontal fluxes of zonal momentum to quantify and gain insight into the role of lateral propagation of SGS GWs in these high-resolution simulations.

To illustrate the importance of SGS lateral fluxes, we first examine the time-mean, zonal-mean effects of adding divergence of the horizontal fluxes of zonal momentum in the calculation of zonal SGS GWD. Figure 7 shows GWD calculated using only SGS vertical fluxes (last term of Equation 12) versus the total GWD calculated using the entire Equation 12 and their differences, that is, the contribution from the horizontal fluxes. We see that the zonal SGS GWD associated with the vertical flux, which is largely due to vertical propagation of GWs, dominates the results. This is consistent with the previous understanding that most of the GWs propagate upward, which is also the basis for the single-column approximation. However, in some critical regions, the role of lateral fluxes is more evident. For example, the amplitude of the GWD due to lateral momentum fluxes is comparable to the GWD due to vertical fluxes near the QBO region (e.g., at 30 km level in Case 1, 35 km in Case 2). As another example, in Case 2, at levels below the SAO (~50 km), it is clear that the lateral momentum fluxes dominate the GWD there, even leading to a change of direction of the total zonal SGS GWD. The strong shear of the background zonal wind in that region likely enhanced the lateral propagation of the GWs (e.g., Dunkerton, 1984; Kruse et al., 2022; Sato et al., 2009).

Examining the PDFs of zonal SGS GWD, which highlights its variability, further shows the importance of the SGS horizontal fluxes. Similar to Figure 6, Figure 8 shows, separately, the PDFs of the GWD associated with SGS zonal fluxes, SGS meridional fluxes, and SGS vertical fluxes. We find that the amplitudes of GWD from these 3 components are fairly close, and there is no evidence of one component dominating over the other two everywhere. To reconcile this with the zonal-mean, time-mean results (Figure 7), we point out that the *mean* GWD associated SGS horizontal fluxes suffers more from cancellations due to opposite lateral propagation directions of GWs, whereas most vertically propagating GWs go upward. However, we emphasize again that any GWP scheme would need to feed instantaneous GWD to the GCMs; therefore, to develop a data-driven GWP scheme, the instantaneous patterns of GWD have to be derived from the high-resolution data.

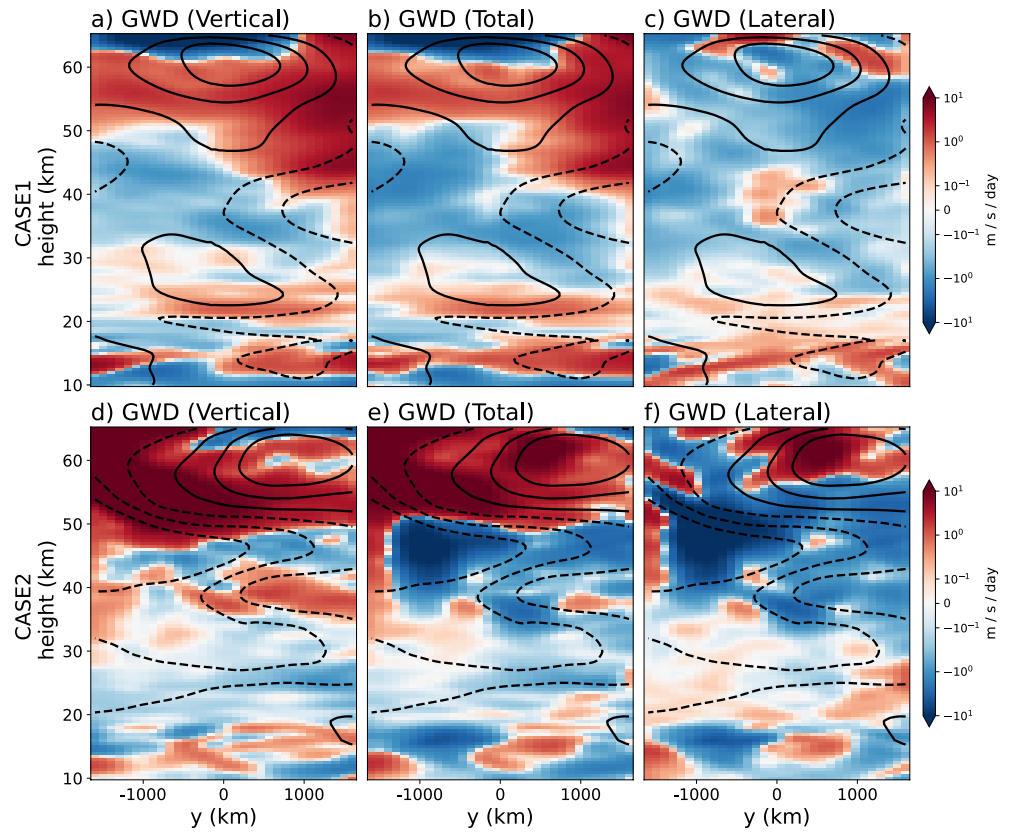


Figure 7. Zonal-mean (weather research and forecasting model domain only), time-mean zonal sub-grid scale Gravity Wave Drag due to only the vertical fluxes, total fluxes, and the lateral fluxes calculated using the UUGS_D method (Equation 12). Left column: Only the vertical component $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\bar{u}\bar{w} - \bar{u}\bar{w})]$. Middle column: All terms $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial x} [\bar{\rho}(\bar{u}\bar{u} - \bar{u}\bar{u})] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial y} [\bar{\rho}(\bar{u}\bar{v} - \bar{u}\bar{v})] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\bar{u}\bar{w} - \bar{u}\bar{w})]$. Right column: Only the first two terms, that is, only the horizontal fluxes. As in Figure 3, lines show the mean zonal winds, but with a contour interval of 20 m/s.

So far, we have used the UUGS_D method to calculate the GWD due to SGS horizontal fluxes (Figures 7 and 8). Compared to the vertical fluxes shown earlier, calculations of SGS GWD due to horizontal momentum fluxes have a much stronger sensitivity to the choice of the method. Figure 9 shows the time-mean, zonal mean SGS GWD associated with the meridional fluxes of the zonal momentum, calculated using HELM_D, REYN_D, and UUGS_D (the second term in Equation 12), respectively. Different from Figure 4, the results here strongly depend on the method, even after averaging over time (simulation period) and space (zonal direction). This suggests that if we want to include the lateral propagation of GWs in the data-driven GWP schemes, then we must carefully examine the GWD extraction methodology. The PDFs in Figure 10 show the same story. The SGS GWD induced by the lateral fluxes are much larger if calculated using the UUGS method compared to the other two (note the logarithmic color bar). It is clear that drag due to Reynolds stress is not the dominant term anymore when we consider GWD due to the SGS lateral fluxes. One explanation for this is that there are fundamental differences between the scales and amplitudes of the horizontal winds (u, v) and the vertical winds (w), as already shown in Figure 2. The vertical velocity is dominated by small-scale features with negligible signal at the resolved scales in GCMs, which results in weak interactions between the resolved scales and the small scales. On the contrary, the horizontal winds are dominated by winds at the resolved scales, which means much stronger interactions between the resolved scales and the unresolved scales, and hence the large differences between UUGS_D and REYN_D.

Moreover, while Figures 9 and 10 suggest similarities between the SGS GWD associated with the SGS horizontal fluxes calculated using the HELM_D and the REYN_D methods (Figures 9a and 9b), substantial differences can exist even between the SGS GWD patterns extracted using these two methods. Figure 11 shows the correlation between instantaneous GWD calculated using the HELM_D and REYN_D method. For the SGS GWD due to vertical fluxes, as already discussed, there is a good match between these two methods. However, for SGS GWD

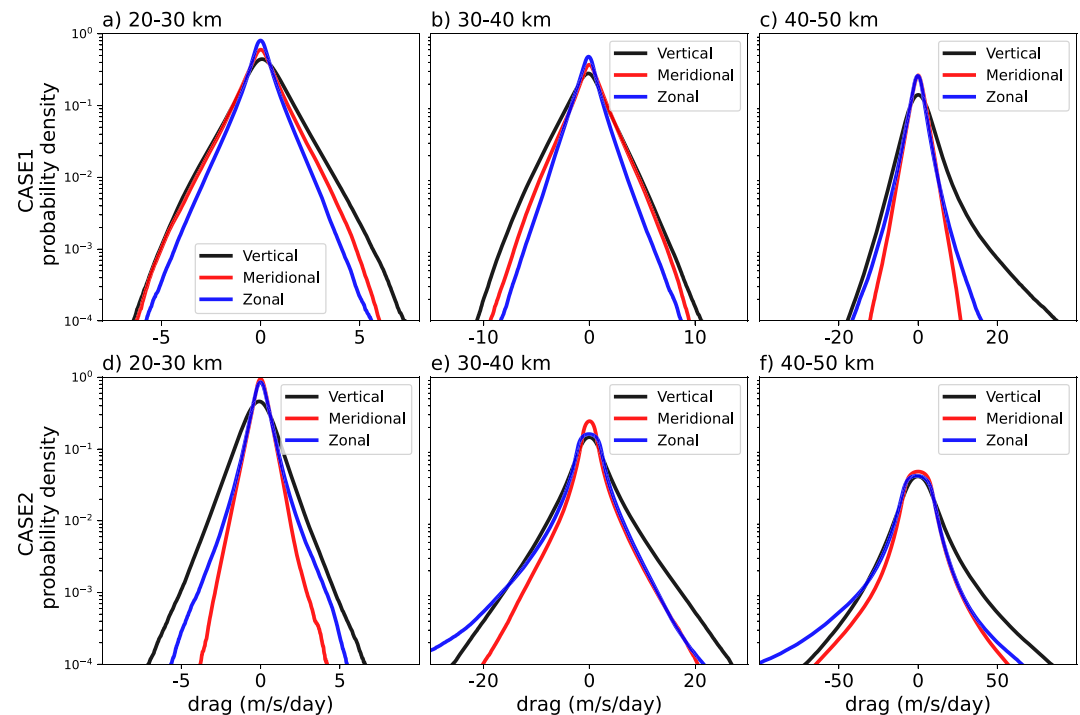


Figure 8. Probability density function (PDF) of zonal sub-grid scale (SGS) Gravity Wave Drag due to SGS zonal (blue), meridional (red), and vertical (black) momentum fluxes for both cases. Upper row: Case 1; Lower panel: Case 2. Left column, (a and d): 20–30 km. Middle column, (b and e): 30–40 km. Right column, (c and f): 40–50 km. Note the difference in the x-axes.

due to the horizontal fluxes, the correlation is fairly weak, even though mathematically similar expressions are used for GWD in both methods. These results, again, show the high sensitivity of the lateral momentum fluxes and the resulting GWD to the details of the extraction method.

So far, we have discussed SGS horizontal fluxes in the two representative cases. Again, we reach the same conclusions if other cases, or all cases together, are examined. Figure S5 in Supporting Information S1 is the same as Figure 10, but with data from all 20 cases combined.

3.4. Sensitivity to the Filter Type/Size and the General Circulation Model Resolution

Until now, we have presented all the analyses using the Gaussian filter and $\Delta = 700$ km. Here, we explore the effects of using a smaller filter size ($\Delta = 200$ km) and two other filter types: top-hat (box) and the sharp-spectral. The kernels and transfer functions of these 3 low-pass filters are listed in Table 1). It should be noted that a few novel filters have been recently developed (e.g., Aluie et al., 2018; Grooms et al., 2021) to handle complex model grids such as the non-uniform ones (see the footnote in Appendix A). However, with the uniform 3 km grid spacing in our WRF simulations, these 3 commonly used filters serve the purpose of this study.

As mentioned earlier, Figure B1 shows the power spectrum of the zonal winds before and after these low-pass filters are applied. This figure demonstrates the overall similarities between the outcome of the Gaussian and top-hat filters, at least up to the filtering scale, and major differences with the outcome of the sharp-spectral filter. Figure B2 shows snapshots of the SGS vertical momentum flux (Reynolds stress and total stress) extracted using these 3 filters and $\Delta = 700$ km. Again, we see that the Gaussian and top-hat filters overall yield fairly similar results. The outcomes of the sharp-spectral filter on the other hand, show differences in both amplitude and pattern, though the degree of difference is more pronounced for the Reynolds stress.

All the results shown so far are with filter size $\Delta = 700$ km, coarse-grained to the GCM resolution of 100 km. However, this choice of 700 km is rather subjective, as there is no well-defined physical scale separation for GWs. Moreover, the appropriate filtering scale depends on the capability of a given GCM to resolve the GWs

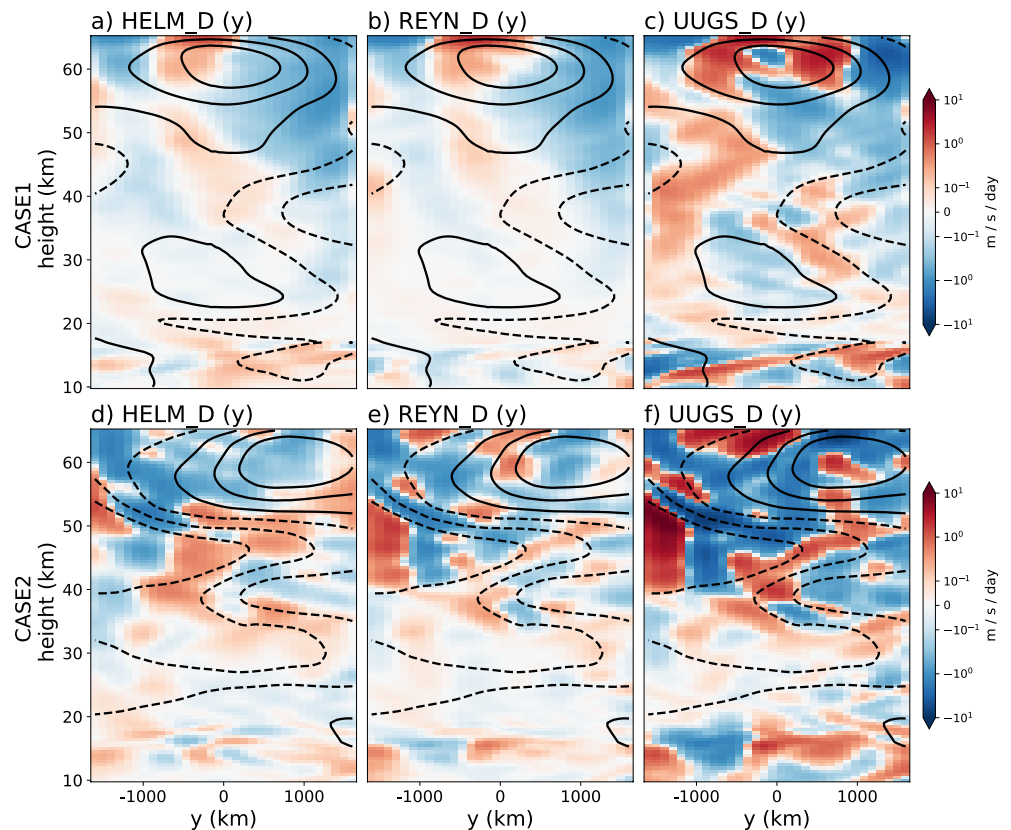


Figure 9. Zonal-mean (weather research and forecasting model domain only), time-mean zonal sub-grid scale (SGS) Gravity Wave Drag (GWD), similar to Figure 4, but for the GWD due to the un/under-resolved (SGS) horizontal fluxes. Only the meridional direction is shown here; examining the zonal direction shows a similar story.

larger than the GCM's grid spacing, that is, it depends on “effective resolution” of the GCM, which in turn depends on the GCM's numerical schemes and choices of grid-scale filters, like hyperdiffusion (e.g., Klaver et al., 2020). The $\Delta = 700$ km used here is based on studies showing that the effective resolution of WRF for GWs is 7 times the grid spacing (Skamarock, 2004). Admittedly, Δ should be chosen based on the effective resolution of the target GCM, not that of the GW-resolving model. To examine the sensitivity of the results to this choice, below we also present analysis with filtering scale that is twice the GCM grid spacing (i.e., $\Delta = 200$ km), which is based on the lower limit and is also the typically used value in the LES literature (Guan et al., 2022; Pope, 2000; Sagaut, 2006). Furthermore, with the increase in computing power, some GCMs now have grid spacing of 0.5° or even smaller. Therefore, below, we also show results with for a GCM with the grid spacing of 30 km.

Figure 12 shows PDFs of local instantaneous values of GWD from the SGS zonal, meridional, and vertical fluxes for 3 sets of choices: GCM grid spacing of 100 km and $\Delta = 700$ km (left column), $\Delta = 200$ km (middle column), and GCM grid spacing of 30 km and $\Delta = 200$ km (right column). The second choice is meant to show the influences of effective resolution change while the third choice is meant to show what happens with higher-resolution GCMs. Although one might expect smaller SGS GW wind perturbations with reduced Δ , SGS GWD is not always reduced, as less averaging of the momentum flux is also applied with a smaller Δ . As a result, in both cases, for the SGS GWD associated with vertical fluxes (solid lines), we find slightly SGS GWD amplitude at the tails of the PDF when the filter length scale is reduced (compare the tails of the PDFs in the left and middle columns).

Moreover, when we reduce the filter size and the GCM grid spacing together, Figures 12c and 12f also show that the SGS GWD at the tails of the PDF are not reduced, which suggests the need of 3D GWP even in a high-resolution GCM. This result is also not sensitive to the choice of the filtering method. Figure S6 in Supporting Information S1 shows the same plot with a different approach (REYN_D). Still, we can see that the instantaneous SGS GWD at the tails of the PDFs are not reduced with 30 km GCM grid spacing.

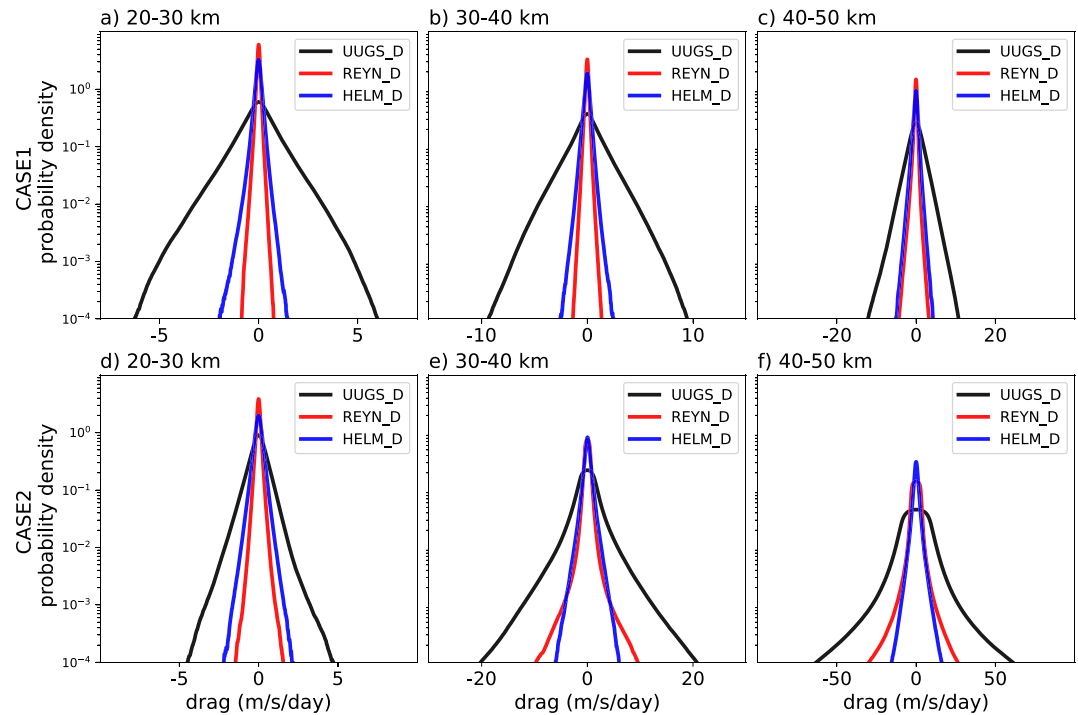


Figure 10. Same as Figure 6, but for the un- and under-resolved zonal Gravity Wave Drag (GWD) due to sub-grid scale (SGS) meridional momentum fluxes. Note the difference in the x -axes. See Figure S4 in Supporting Information S1 for the probability density functions (PDFs) of zonal GWD due to the SGS zonal momentum fluxes. See Figure S5 in Supporting Information S1 for the PDFs of data from all 20 cases combined.

4. Summary and Discussion

The use of ML algorithms for developing data-driven SGS parameterization of GWs has gained attention in recent years, given the increasing availability of data from observations and high-resolution GW-resolving simulations, and a few successful case study and emulation attempts (e.g., Amiramjadi et al., 2022; Chantry et al., 2021; Espinosa et al., 2022; Matsuoka et al., 2020). Powerful ML techniques have recently emerged for data-driven weather/climate modeling, suggesting that concerns such as stability, lack of physical constraints, learning in the small-data regime, and interpretability could be addressed, at least to some degree, in the near future (e.g.,

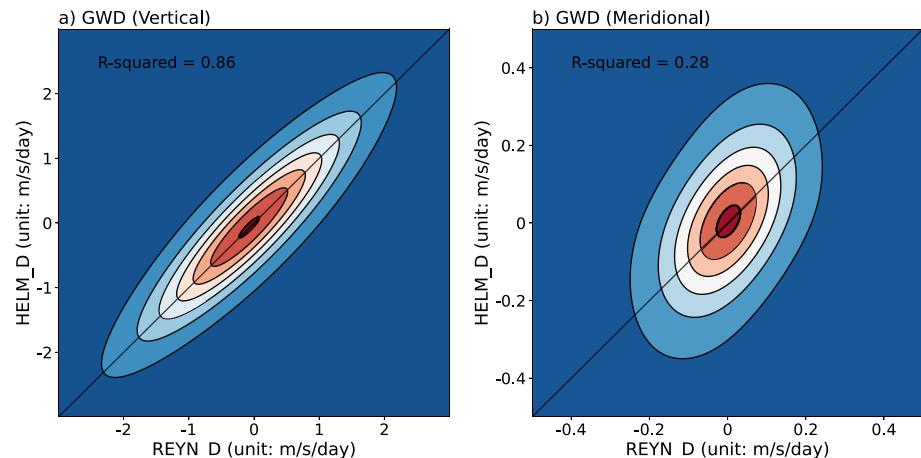


Figure 11. Joint probability density functions between the sub-grid scale (SGS) Gravity Wave Drag (GWD) calculated using the HELM_D and REYN_D methods for Case 1. (a) GWD due to SGS vertical fluxes and (b) GWD due to SGS meridional fluxes. The R -squared value is shown to measure the correlation between these two methods.

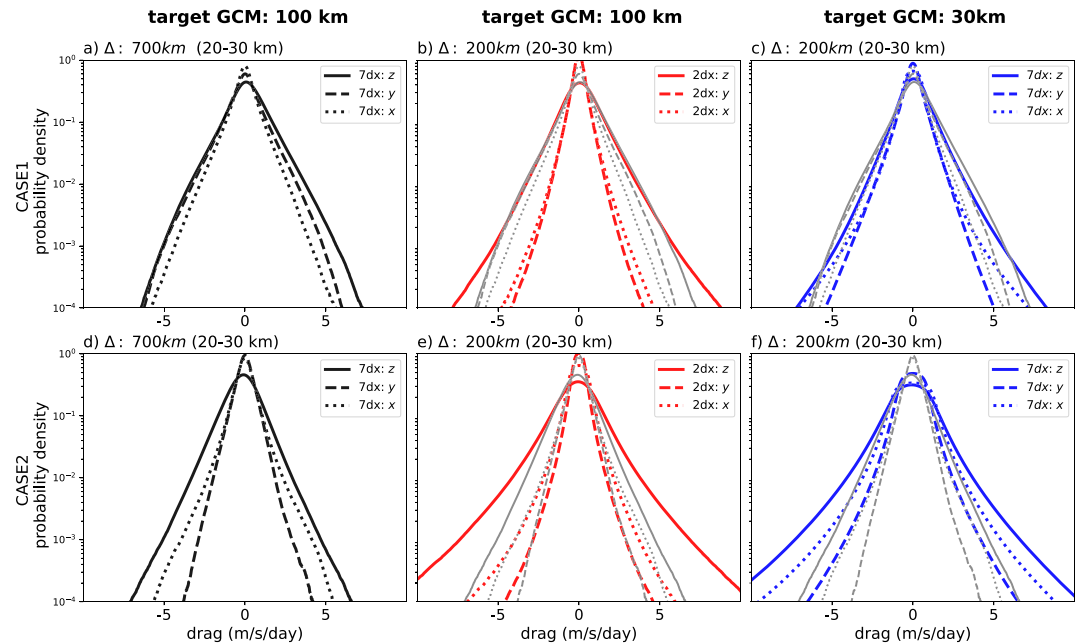


Figure 12. The probability density functions (PDFs) of Gravity Wave Drag at 20–30 km heights associated with the sub-grid scale zonal (dotted), meridional (dashed), and vertical (solid) fluxes extracted using the UUGS_D method (with Gaussian filter) for 3 different sets of filter size Δ and general circulation model (GCM) grid spacing. Left column: $\Delta = 700$ km, with targeted GCM grid spacing of 100 km. Middle columns: $\Delta = 200$ km, with targeted GCM grid spacing of 100 km. Right column: $\Delta = 200$ km, with targeted GCM grid spacing of 30 km. Top row: Case 1. Bottom row: Case 2. The thin gray lines in the middle and right columns show the PDFs from the left column to facilitate comparison.

Beucler et al., 2021; Dunbar et al., 2021; Guan et al., 2023; Mamalakis et al., 2022; Pathak et al., 2022; Subel et al., 2023). However, the best ML algorithm is just as good as the data used in the training. As a result, a major remaining challenge in developing data-driven GWP schemes (and in general, any data-driven parameterizations) that has not received much attention is extracting the SGS GWD from high-resolution simulations. This GWD is what needs to be learned in terms of the resolved flow during training.

As the first step in addressing this challenge, in this study, we have generated a library of 20 tropical convection-permitting WRF simulations and systematically compared the sensitivity of the extracted under- and un-resolved (SGS) 3D GWD to the choices of methods and parameters. Three methods from the GW and LES literature have been examined (HELM_D, UUGS_D, and REYN_D). The key conclusions obtained from these comparisons are:

1. For GWD due to SGS vertical momentum fluxes, all three methods give consistent time-mean, zonal-mean results. Yet, if we consider snapshots at different times and locations, the GWD from the UUGS_D method has additional spatial and temporal variability compared to the GWD in other methods. This additional variability is partially due to the fact that the UUGS_D method includes cross-scale interactions between the SGS GWs and the large-scale background flow resolved by a GCM. Given that a GWP needs to provide patterns of GWD at each time step of the GCM, correctly representing the variability of the GWD in the training data set could be essential. It is unknown yet whether this will improve the performance of the targeted GCMs in terms of conventional metrics (e.g., QBO statistics). It is possible that additional variability may provide some of the same benefits as stochastic parameterization in ensemble weather and climate prediction (Lott et al., 2012; Palmer et al., 2005; G. Shutts, 2005).
2. There is a growing number of studies showing that the lateral propagation of GWs plays a significant role in the resolved flow's momentum budget, and could even reverse the direction of GWD for certain regions and cases (e.g., Kruse et al., 2022). Our comprehensive analysis of these lateral effects supports this conclusion. The SGS GWD associated with lateral momentum fluxes has comparable amplitudes to the SGS GWD associated with vertical momentum fluxes. This is true not only when the spatiotemporal variability is considered, but also in the time-mean, zonal-mean GWD. Our findings strongly suggest the need for including the effects

of SGS horizontal fluxes in the GWP schemes. However, there are practical implementation challenges for a truly 3D scheme in GCMs (Y. H. Kim et al., 2021). Therefore, further tests, both offline and online (coupled), are needed to see if 3D GWP schemes improve the circulation variability in GCMs. That said, there is existing evidence for SGS modeling of other physical processes that would benefit from including neighboring columns, providing further incentive for considering horizontally non-local parameterizations (e.g., Fatkullin & Vanden-Eijnden, 2004; Guan et al., 2022; Wang et al., 2022).

3. Adding to the complexity, we have found that the GWD due to SGS lateral momentum fluxes could be sensitive to the methods used to extract them. Even the time-mean, zonal-mean GWD could be very different when different methods are used. The instantaneous GWD amplitudes from the UUGS_D method could be much larger than those from the REYN_D method due to the strong cross-scale interactions between SGS GWs and the horizontal background flow that could be resolved by the GCMs. This suggests that to include the SGS GWD due to the horizontal fluxes in the data-driven GWP schemes, further research needs to be done on the extraction methodology.
4. The sensitivity of extracted SGS GWD to the length scale (Δ) of the filtering operation and the horizontal resolution of the GCM are also studied. Our results suggest that both the effective resolution of a GCM and its grid spacing have significant influences on the calculated SGS GWD (Figure 12). Interestingly, within the explored GCM grid spacing (30–100 km), the amplitude of extracted SGS GWD does not decrease as the GCM's horizontal grid spacing is reduced, suggesting the need for GWP schemes in the foreseeable future even as the GCM resolutions are increased. Given the sensitivity of the results to the filter size, the grid spacing of the GCM and its “effective” resolution might be used as inputs to design scale-aware data-driven GWPs schemes. Also note that here we have only examined the effects of the GCMs' horizontal resolution. The vertical resolution of GCMs has a major impact on how well the GWs are resolved and the resulting GWD (Skamarock et al., 2019). This issue needs to be fully investigated in future work.

All these findings point to the next two steps needed in developing data-driven GWP schemes. One step is to further work on developing theoretical and mathematical frameworks to separate the GWs from the background flows, and quantifying the under- and un-resolved fluxes for a given GCM. The others step is to use the extracted GWD from this library using different methods and choices, train ML algorithms such as deep neural networks, couple them to GCMs such as WACCM, and investigate the large-scale circulation variability, for example, of the QBO. With proper metrics of the large-scale variability (e.g., period and amplitude of QBO), we could potentially gain insight into which method and choice of filter type/size lead to a GWP scheme that produces the most realistic circulation, compared to observations. However, there could be several practical challenges in doing this. First, it may not be easy to isolate the performance of the GWP scheme from biases in the GCMs' large-scale circulation and other parameterizations, for example, that of moist convection, which is the source of convective generated GWs. That said, some of these biases, such as the latter one, could be corrected for the purpose of this analysis.

Second, the traditional single-column approach uses inputs (resolved flow) and outputs (GWD) only from the same GCM column and does not require any cross-column communication, which works well with the GCMs' parallelization. However, accounting for the non-local effects, that is, inputs from neighboring columns and possibly memory (history), can require cross-processor communications, which come with a large computational overhead. Recently, there has been observational evidence showing that the majority of GW momentum fluxes are typically found to be at distances closer than 400 km from convection sources (Corcos et al., 2021). This is encouraging as it suggests that a small stencil of neighboring columns (which could be computationally affordable) might be enough to account for the non-local effects and lateral SGS momentum fluxes. It should be noted that the outputs of the WRF simulations are saved such that information about convection and history is available for such future investigations.

We also highlight that given the sensitivity of the SGS GWD to the filter size (Δ), the scale-awareness of the data-driven GWP scheme is critical. One potential approach is to create SGS GWD data sets for different filter sizes and GCM grid spacing and combine them all together in a training set, with the filter length scale and the GCM grid spacing serving as the inputs to the ML algorithm too.

Finally, we aim to further validate and expand the library. The GWs in the simulations can be sensitive to model physics (e.g., various schemes), numerical methods, and model dissipation and resolution. Because of this, there is a large inter-model spread in “resolved” GW fluxes and drags from convection-permitting simulations (Polichtchouk et al., 2022; Stephan et al., 2019). Observational validation is required to resolve this inter-model

spread and obtain high-fidelity data for the data-driven GWP schemes. All WRF simulations used in this study have a grid spacing of 3 km. While 3 km is enough to resolve most of the GW spectra, it is not adequate to entirely resolve convection, which is the key source of the GWs in the tropics (Bramberger et al., 2020; Jeevanjee, 2017). We have conducted limited simulations with 1 and 3 km grid spacing at periods when observations from super-pressure balloons are available. We will use these simulations to examine the effect of resolution and validate the GW fluxes in the library. Also, currently our WRF simulations are limited to the tropics. In the future, we aim to extend the library to the mid-latitude and even polar regions (and include more cases in the data set if it is not adequate for the training). We also plan to conduct pseudo-global warming experiments to examine the response of the GWD to climate change and to expand the library to include training sets for transfer learning, such that generalizable data-driven GWP schemes could be developed (Guan et al., 2022; Subel et al., 2023).

Appendix A: UUGS Drag Extraction

To better illustrate the GWD parameterization problem, here we use the zonal momentum equation as an example. The flux form of the zonal momentum equation in the atmosphere, without any approximation, can be written as follows in Cartesian coordinates:

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho uu)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z} = -\frac{\partial p}{\partial x} + \rho f v + \rho F_x, \quad (\text{A1})$$

where (u, v, w) is the 3D wind fields; p is pressure; ρ is density; f is the Coriolis parameter; and F_x is the friction and/or numerical diffusion term.

The problem of the parameterization of gravity waves and/or other sub-grid scale (SGS) physical processes arises because general circulation models (GCMs) have only a limited horizontal resolution (typically with a grid spacing on the order of 100 km). Therefore, they can only resolve the large-scale part of each physical variable. Let's use ϕ^G to denote the variable ϕ in the GCM, then the zonal momentum equation in the GCM would be:

$$\frac{\partial(\rho^G u^G)}{\partial t} + \frac{\partial(\rho^G u^G u^G)}{\partial x} + \frac{\partial(\rho^G u^G v^G)}{\partial y} + \frac{\partial(\rho^G u^G w^G)}{\partial z} = -\frac{\partial p^G}{\partial x} + \rho^G f v^G + \rho^G F_x^G + \mathbf{X}_x^G \quad (\text{A2})$$

where $\mathbf{X}_x^G$ is SGS zonal drag in the GCM due to its limited resolution. The problem is then to find $\mathbf{X}_x^G$ from high-resolution simulation data generated, for example, by WRF.

As introduced in the main text, We use $\tilde{(\cdot)}$ to represent the spatial filtering process, which largely removes/reduces signals that have horizontal scales smaller than some specific value (Δ in Equation 1). With this definition, all variables can be partitioned into the large-scale background and the perturbation parts:

$$\phi = \tilde{\phi} + \phi'. \quad (\text{A3})$$

Note that ϕ , $\tilde{\phi}$, and ϕ' have the same resolution. In a GCM, only the large-scale part $\tilde{\phi}$ can be captured. An additional coarse-graining process, denoted as $\bar{(\cdot)}$, is required to transfer this large-scale part to the GCM grid, so that,

$$\phi^G \approx \bar{\tilde{\phi}} \quad (\text{A4})$$

Applying Equation A4 to Equation A2 yields

$$\frac{\partial(\bar{\tilde{\rho}} \bar{\tilde{u}})}{\partial t} + \frac{\partial(\bar{\tilde{\rho}} \bar{\tilde{u}} \bar{\tilde{u}})}{\partial x} + \frac{\partial(\bar{\tilde{\rho}} \bar{\tilde{u}} \bar{\tilde{v}})}{\partial y} + \frac{\partial(\bar{\tilde{\rho}} \bar{\tilde{u}} \bar{\tilde{w}})}{\partial z} = -\frac{\partial \bar{p}}{\partial x} + \bar{\tilde{\rho}} f \bar{\tilde{v}} + \bar{\tilde{\rho}} \bar{\tilde{F}}_x + \mathbf{X}_x^G. \quad (\text{A5})$$

To get an expression for $\mathbf{X}_x^G$, we can apply both spatial filtering and coarse-graining operators to each term in the original Equation A1, and assume that the operations are commutative (valid for fixed resolution grids) which means, for example,

$$\frac{\partial \tilde{\phi}}{\partial x} = \frac{\partial \bar{\tilde{\phi}}}{\partial x}. \quad (\text{A6})$$

Then we get

$$\frac{\partial(\bar{\rho}u)}{\partial t} + \frac{\partial(\bar{\rho}uu)}{\partial x} + \frac{\partial(\bar{\rho}uv)}{\partial y} + \frac{\partial(\bar{\rho}uw)}{\partial z} = -\frac{\partial\bar{p}}{\partial x} + \bar{\rho}f_v + \bar{\rho}F_x. \quad (A7)$$

Next, we assume that density perturbations are negligible, $\bar{\rho} \gg \rho'$, and that only the vertical gradient of ρ is non-negligible, so that ρ is a function of height only:

$$\rho(x, y, z, t) \approx \bar{\rho}(z). \quad (A8)$$

Then, for any variable ϕ , given that both spatial filtering and coarse-graining operate only on the horizontal levels, we obtain

$$\bar{\rho}\bar{\phi} \approx \bar{\rho}\bar{\phi} = \bar{\rho}\bar{\phi} \quad (A9)$$

Subtracting Equation A7 from Equation A5, and applying Equation A9, we finally get:

$$\begin{aligned} \mathbf{X}_x^G &= \frac{\partial(\bar{\rho}\bar{u}\bar{u})}{\partial x} + \frac{\partial(\bar{\rho}\bar{u}\bar{v})}{\partial y} + \frac{\partial(\bar{\rho}\bar{u}\bar{w})}{\partial z} - \left[\frac{\partial(\bar{\rho}uu)}{\partial x} + \frac{\partial(\bar{\rho}uv)}{\partial y} + \frac{\partial(\bar{\rho}uw)}{\partial z} \right] \\ &= -\frac{\partial}{\partial x} [\bar{\rho}(\bar{u}\bar{u} - \bar{u}\bar{u})] - \frac{\partial}{\partial y} [\bar{\rho}(\bar{u}\bar{v} - \bar{u}\bar{v})] - \frac{\partial}{\partial z} [\bar{\rho}(\bar{u}\bar{w} - \bar{u}\bar{w})]. \end{aligned} \quad (A10)$$

Similarly, for the meridional momentum equation, we obtain

$$\mathbf{X}_y^G = -\frac{\partial}{\partial x} [\bar{\rho}(\bar{v}\bar{u} - \bar{v}\bar{u})] - \frac{\partial}{\partial y} [\bar{\rho}(\bar{v}\bar{v} - \bar{v}\bar{v})] - \frac{\partial}{\partial z} [\bar{\rho}(\bar{v}\bar{w} - \bar{v}\bar{w})] \quad (A11)$$

Note Equation A10 is in the density-weighted form as in Equation A2. If we were to consider the drag forces terms directly, then the density factor shall be removed, leading to

$$\overline{GWD_x} = -\frac{1}{\bar{\rho}} \frac{\partial}{\partial x} [\bar{\rho}(\bar{u}\bar{u} - \bar{u}\bar{u})] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial y} [\bar{\rho}(\bar{u}\bar{v} - \bar{u}\bar{v})] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\bar{u}\bar{w} - \bar{u}\bar{w})], \quad (A12)$$

which is the equation used in the main text.

The terms in brackets on the right-hand side of Equation A10 are differences between the filtered and coarse-grained flux and the flux calculated based on the filtered, coarse-grained prognostic variables. We will refer to these terms as the total SGS fluxes. They can be further decomposed (Germano, 1986; Leonard, 1975; Sagaut, 2006), for example,

$$\begin{aligned} \overline{uw} - \bar{u}\bar{w} &= \overline{(\tilde{u} + u')(\tilde{w} + w')} - \overline{(\tilde{u} + u')(\tilde{w} + w')} \\ &= \underbrace{\overline{(\tilde{u}\tilde{w} - \tilde{u}\tilde{w})}}_{\text{Leonard stress}} + \underbrace{\overline{(\tilde{u}w' + u'\tilde{w} - \tilde{w}'\tilde{u} - u'\tilde{w})}}_{\text{cross stress}} + \underbrace{\overline{(u'w' - u'w')}}_{\text{Reynolds stress}} \end{aligned} \quad (A13)$$

We see that the Reynolds stress is one of the three components of the total SGS flux. The total SGS flux accounts for interactions among all scales, including scales resolved by the GCMs with the un- and under-resolved scales. The importance of Leonard term and cross term has long been shown in studies of turbulent flows (e.g., Galmarini et al., 2000; Leonard, 1975). We also note that the Reynolds term here based on spatial filtering is different with the traditionally temporal-based Reynolds average in which the flow is decomposed into a mean and fluctuating components. As pointed in Aluie et al. (2018), the time-mean flow is not synonymous with large-scale flow, nor does a temporal fluctuation directly correspond to a characteristic length scale.

Appendix B: The Effects of Filter Type

Figures B1 and B2 show the spectra and snapshots of zonal wind and SGS vertical flux when the 3 different filters are applied with the same Δ . Note that because our WRF regional domain is not periodic, we have used mirrored

ties to reduce the boundary effects. Still, Fourier-based filters such as the sharp-spectral filter might suffer from Gibbs oscillations and give non-physical results.

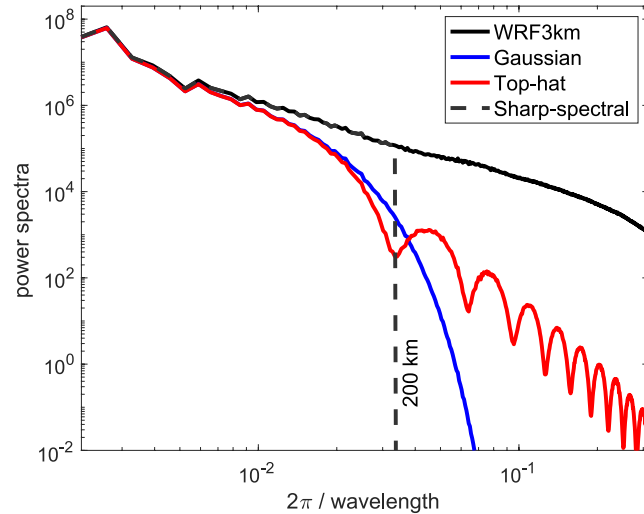


Figure B1. Example of the power spectrum of the zonal wind from the 3-km weather research and forecasting model simulations before and after the low-pass filtering using different kernels (with length scale $\Delta = 200$ km). The black line shows the spectrum before filtering, while the blue, red, and green lines show the spectrum after applying the Gaussian, top-hat (box), and sharp-spectral filters. Note that the dashed green line coincides with the black line for scales longer than 200 km. The oscillations in the red line are the well-known ringing effects of the top-hat filter (e.g., Pope, 1975; Zhou et al., 2019).

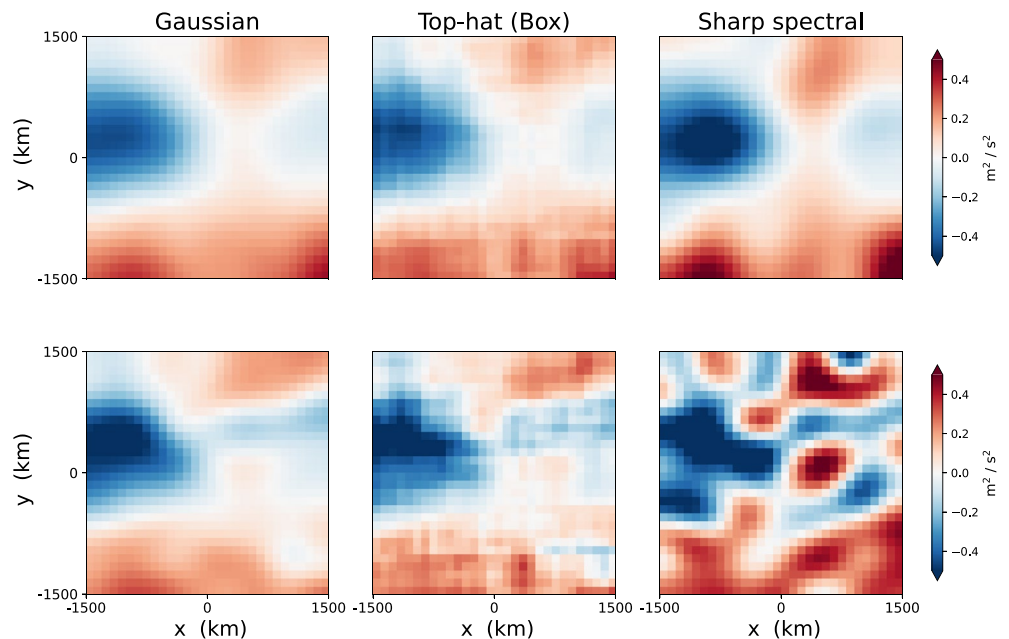


Figure B2. Snapshots of the sub-grid scale (SGS) vertical momentum fluxes, calculated using the 3 different filters for Case 1 (at 40 km height). Top row: the Reynolds stress, $u'w'$. Bottom: the total SGS stress, $u\bar{w} - \bar{u}\bar{w}$. The filter size is $\Delta = 700$ km.

Data Availability Statement

The ERA5 data can be downloaded from <https://cds.climate.copernicus.eu/>. The WRF model is available here https://www2.mmm.ucar.edu/wrf/users/download/get_source.html. The coarse-grained data and code for all the analysis in the main text is available at <https://doi.org/10.5281/zenodo.7439397>. The raw data of full WRF simulations will be made available on demand.

Acknowledgments

We thank Inna Polichtchouk and two anonymous reviewers for helpful comments and suggestions. We are grateful to Ed Gerber, Yifei Guan, Hamid Pahlavan, Aditi Sheshadri, and Claudia Stephans for insightful discussions. This work was supported by grants from the NSF OAC CSSI program (2005123 and 2004512), and by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. Computational resources were provided by NSF XSEDE (allocation ATM170020) and NCAR's CISL (allocation URIC0009).

References

- Achatz, U., Ribstein, B., Senf, F., & Klein, R. (2017). The interaction between synoptic-scale balanced flow and a finite-amplitude mesoscale wave field throughout all atmospheric layers: Weak and moderately strong stratification. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 342–361. <https://doi.org/10.1002/qj.2926>
- 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>
- Alfonsi, G. (2009). Reynolds-averaged Navier–Stokes equations for turbulence modeling. *Applied Mechanics Reviews*, 62(4), 040802. <https://doi.org/10.1115/1.3124648>
- Aluie, H., Hecht, M., & Vallis, G. K. (2018). Mapping the energy cascade in the North Atlantic Ocean: The coarse-graining approach. *Journal of Physical Oceanography*, 48(2), 225–244. <https://doi.org/10.1175/JPO-D-17-0100.1>
- Amiranjadi, M., Plougonven, R., Mohebalhojeh, A. R., & Mirzaei, M. (2022). Using machine learning to estimate non-orographic gravity wave characteristics at source levels. *Journal of the Atmospheric Sciences*, 80(2), 419–440. <https://doi.org/10.1175/JAS-D-22-0021.1>
- Beck, A., & Kurz, M. (2021). A perspective on machine learning methods in turbulence modeling. *GAMM-Mitteilungen*, 44(1), e202100002. <https://doi.org/10.1002/gamm.202100002>
- Beres, J. H. (2004). Gravity wave generation by a three-dimensional thermal forcing. *Journal of the Atmospheric Sciences*, 61(14), 1805–1815. [https://doi.org/10.1175/1520-0469\(2004\)061<1805:GWGBAT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1805:GWGBAT>2.0.CO;2)
- Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentile, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, 126(9), 098302. <https://doi.org/10.1103/physrevlett.126.098302>
- Böhlöni, G., Kim, Y. H., Borchert, S., & Achatz, U. (2021). Toward transient subgrid-scale gravity wave representation in atmospheric models. Part I: Propagation model including nondissipative wave mean-flow interactions. *Journal of the Atmospheric Sciences*, 78(4), 1317–1338. <https://doi.org/10.1175/JAS-D-20-0065.1>
- Böhlöni, G., Ribstein, B., Muraschko, J., Sgoff, C., Wei, J., & Achatz, U. (2016). The interaction between atmospheric gravity waves and large-scale flows: An efficient description beyond the nonacceleration paradigm. *Journal of the Atmospheric Sciences*, 73(12), 4833–4852. <https://doi.org/10.1175/JAS-D-16-0069.1>
- Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, 11(1), 376–399. <https://doi.org/10.1029/2018ms001472>
- Bramberger, M., Alexander, M. J., & Grimsdell, A. W. (2020). Realistic simulation of tropical atmospheric gravity waves using radar-observed precipitation rate and echo top height. *Journal of Advances in Modeling Earth Systems*, 12(8), e2019MS001949. <https://doi.org/10.1029/2019MS001949>
- Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45(12), 6289–6298. <https://doi.org/10.1029/2018gl078510>
- Bretherton, F. P. (1969). Momentum transport by gravity waves. *Quarterly Journal of the Royal Meteorological Society*, 95(404), 213–243. <https://doi.org/10.1002/qj.49709540402>
- Callies, J., Ferrari, R., & Bühler, O. (2014). Transition from geostrophic turbulence to inertia-gravity waves in the atmospheric energy spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, 111(48), 17033–17038. <https://doi.org/10.1073/pnas.1410772111>
- Cao, J., Ran, L., & Li, N. (2014). An application of the Helmholtz theorem in extracting the externally induced deformation field from the total wind field in a limited domain. *Monthly Weather Review*, 142(5), 2060–2066. <https://doi.org/10.1175/MWR-D-13-00311.1>
- 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>
- Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven super-parameterization using deep learning: Experimentation with multiscale Lorenz 96 systems and transfer learning. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002084. <https://doi.org/10.1029/2020ms002084>
- Chen, T.-C., & Wiin-Nielsen, A. C. (1976). On the kinetic energy of the divergent and nondivergent flow in the atmosphere. *Tellus*, 28(6), 486. <https://doi.org/10.3402/tellusa.v28i6.11317>
- Clark, R. A., Ferziger, J. H., & Reynolds, W. C. (1979). Evaluation of subgrid-scale models using an accurately simulated turbulent flow. *Journal of Fluid Mechanics*, 91(1), 1–16. <https://doi.org/10.1017/s002211207900001x>
- Corcos, M., Hertzog, A., Plougonven, R., & Podglajen, A. (2021). Observation of gravity waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical Research: Atmospheres*, 126(15), e2021JD035165. <https://doi.org/10.1029/2021JD035165>
- Dunbar, O. R., Garbuno-Inigo, A., Schneider, T., & Stuart, A. M. (2021). Calibration and uncertainty quantification of convective parameters in an idealized GCM. *Journal of Advances in Modeling Earth Systems*, 13(9), e2020MS002454. <https://doi.org/10.1029/2020ms002454>
- Dunkerton, T. J. (1984). Inertia-gravity waves in the stratosphere. *Journal of the Atmospheric Sciences*, 41(23), 3396–3404. [https://doi.org/10.1175/1520-0469\(1984\)041<3396:iwits>2.0.co;2](https://doi.org/10.1175/1520-0469(1984)041<3396:iwits>2.0.co;2)
- Eckermann, S. D., Ma, J., & Broutman, D. (2015). Effects of horizontal geometrical spreading on the parameterization of orographic gravity wave drag. Part I: Numerical transform solutions. *Journal of the Atmospheric Sciences*, 72(6), 2330–2347. <https://doi.org/10.1175/JAS-D-14-0147.1>
- Ehard, B., Kaifler, B., Dörnbrack, A., Preusse, P., Eckermann, S. D., Bramberger, M., et al. (2017). Horizontal propagation of large-amplitude mountain waves into the polar night jet. *Journal of Geophysical Research: Atmospheres*, 122(3), 1423–1436. <https://doi.org/10.1002/2016JD025621>
- 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*, 49(8), e98174. <https://doi.org/10.1029/2022GL098174>

- Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., et al. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, 7(4), 813–830. <https://doi.org/10.5194/esd-7-813-2016>
- Fatkuullin, I., & Vanden-Eijnden, E. (2004). A computational strategy for multiscale systems with applications to Lorenz 96 model. *Journal of Computational Physics*, 200(2), 605–638. <https://doi.org/10.1016/j.jcp.2004.04.013>
- Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Review of Geophysics*, 41(1), 1003. <https://doi.org/10.1029/2001RG000106>
- Fritts, D. C., & Nastrom, G. D. (1992). Sources of mesoscale variability of gravity waves. Part II: Frontal, convective, and jet stream excitation. *Journal of the Atmospheric Sciences*, 49(2), 111–127. [https://doi.org/10.1175/1520-0469\(1992\)049<0111:SOMVOG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0111:SOMVOG>2.0.CO;2)
- Galmarini, S., Michelutti, F., & Thunis, P. (2000). Estimating the contribution of Leonard and cross terms to the subfilter scale from atmospheric measurements. *Journal of the Atmospheric Sciences*, 57(17), 2968–2976. [https://doi.org/10.1175/1520-0469\(2000\)057<2968:ETCOLA>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<2968:ETCOLA>2.0.CO;2)
- Gentine, P., Eyring, V., & Beucler, T. (2021). Deep learning for the parametrization of subgrid processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 307–314.
- Germano, M. (1986). A proposal for a redefinition of the turbulent stresses in the filtered Navier-Stokes equations. *The Physics of Fluids*, 29(7), 2323–2324. <https://doi.org/10.1063/1.865568>
- Germano, M. (1992). Turbulence: The filtering approach. *Journal of Fluid Mechanics*, 238, 325–336. <https://doi.org/10.1017/s0022112092001733>
- 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>
- Grooms, I., Loose, N., Abernathy, R., Steinberg, J., Bachman, S. D., Marques, G., et al. (2021). Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, 13(9), e2021MS002552. <https://doi.org/10.1029/2021ms002552>
- Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022). Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *Journal of Computational Physics*, 458, 111090. <https://doi.org/10.1016/j.jcp.2022.111090>
- Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2023). Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate les. *Physica D: Nonlinear Phenomena*, 443, 133568. <https://doi.org/10.1016/j.physd.2022.133568>
- Hong, S.-Y., & Lim, J.-O. J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, 42, 129–151.
- Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9), 2318–2341. <https://doi.org/10.1175/MWR3199.1>
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, 113(D13), D13103. <https://doi.org/10.1029/2008JD009944>
- Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, 9(6), 2304–2316. <https://doi.org/10.1002/2017MS001059>
- Jiang, Q., Doyle, J. D., Eckermann, S. D., & Williams, B. P. (2019). Stratospheric trailing gravity waves from New Zealand. *Journal of the Atmospheric Sciences*, 76(6), 1565–1586. <https://doi.org/10.1175/JAS-D-18-0290.1>
- Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-Bustamante, E. (2012). A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, 140(3), 898–918. <https://doi.org/10.1175/MWR-D-11-00056.1>
- Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmailzadeh, S., et al. (2021). Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200093. <https://doi.org/10.1098/rsta.2020.0093>
- 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>
- Kim, Y. H., Bölöni, G., Borchert, S., Chun, H. Y., & Achatz, U. (2021). Toward transient subgrid-scale gravity wave representation in atmospheric models. Part II: Wave intermittency simulated with convective sources. *Journal of the Atmospheric Sciences*, 78(4), 1339–1357. <https://doi.org/10.1175/JAS-D-20-0066.1>
- Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, 21(4), e952. <https://doi.org/10.1002/asl.952>
- Kruse, C. G., Alexander, M. J., Hoffmann, L., Niekerk, A. V., Polichtchouk, I., Bacmeister, J. T., et al. (2022). Observed and modeled mountain waves from the surface to the mesosphere near the drake passage. *Journal of the Atmospheric Sciences*, 79(4), 909–932. <https://doi.org/10.1175/JAS-D-21-0252.1>
- Kruse, C. G., & Smith, R. B. (2015). Gravity wave diagnostics and characteristics in mesoscale fields. *Journal of the Atmospheric Sciences*, 72(11), 4372–4392. <https://doi.org/10.1175/JAS-D-15-0079.1>
- Kruse, C. G., Smith, R. B., & Eckermann, S. D. (2016). The midlatitude lower-stratospheric mountain wave “valve layer”. *Journal of the Atmospheric Sciences*, 73(12), 5081–5100. <https://doi.org/10.1175/jas-d-16-0173.1>
- Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier.
- Lott, F., Guez, L., & Maury, P. (2012). A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere. *Geophysical Research Letters*, 39(6), L06807. <https://doi.org/10.1029/2012GL051001>
- Lynch, P. (1988). Deducing the wind from vorticity and divergence. *Monthly Weather Review*, 116(1), 86–93. [https://doi.org/10.1175/1520-0493\(1988\)116<0086:DTWFVA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<0086:DTWFVA>2.0.CO;2)
- Mamalakos, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, 1(4), e220012. <https://doi.org/10.1175/aies-d-22-0012.1>
- Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020). Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. *Geophysical Research Letters*, 47(19), e89436. <https://doi.org/10.1029/2020GL089436>
- Maulik, R., San, O., Rasheed, A., & Vedula, P. (2019). Subgrid modelling for two-dimensional turbulence using neural networks. *Journal of Fluid Mechanics*, 858, 122–144. <https://doi.org/10.1017/jfm.2018.770>
- McFarlane, N. A. (1987). The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *Journal of the Atmospheric Sciences*, 44(14), 1775–1800. [https://doi.org/10.1175/1520-0469\(1987\)044<1775:TEOOEG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1987)044<1775:TEOOEG>2.0.CO;2)
- Muraschko, J., Fruman, M. D., Achatz, U., Hickel, S., & Toledo, Y. (2015). On the application of Wentzel-Kramers-Brillouin theory for the simulation of the weakly nonlinear dynamics of gravity waves. *Quarterly Journal of the Royal Meteorological Society*, 141(688), 676–697. <https://doi.org/10.1002/qj.2381>

- Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., & Leutbecher, M. (2005). Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, 33(1), 163–193. <https://doi.org/10.1146/annurev.earth.33.092203.122552>
- 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>
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214.
- Pincus, R., Barker, H. W., & Morcrette, J. J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research*, 108(D13), 4376. <https://doi.org/10.1029/2002jd003322>
- Plougonven, R., & Zhang, F. (2014). Internal gravity waves from atmospheric jets and fronts. *Reviews of Geophysics*, 52(1), 33–76. <https://doi.org/10.1002/2012RG000419>
- Polichtchouk, I., Wedi, N., & Kim, Y. H. (2022). Resolved gravity waves in the tropical stratosphere: Impact of horizontal resolution and deep convection parametrization. *Quarterly Journal of the Royal Meteorological Society*, 148(742), 233–251. <https://doi.org/10.1002/qj.4202>
- Pope, S. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, 72(2), 331–340. <https://doi.org/10.1017/s0022112075003382>
- Pope, S. (2000). *Turbulent flows*. Cambridge university press.
- Qiao, W., Song, J., He, H., & Li, F. (2019). Application of different wind field models and wave boundary layer model to typhoon waves numerical simulation in wavewatch III model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 71(1), 1657552. <https://doi.org/10.1080/16000870.2019.1657552>
- Rai, S., Hecht, M., Maltrud, M., & Aluie, H. (2021). Scale of oceanic eddy killing by wind from global satellite observations. *Science Advances*, 7(28), eabf4920. <https://doi.org/10.1126/sciadv.abf4920>
- Rasp, S., Pritchard, M. S., & Gentile, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, 115(39), 9684–9689. <https://doi.org/10.1073/pnas.1810286115>
- Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the quasi-biennial oscillation in CMIP models. *Journal of Geophysical Research: Atmospheres*, 125(8), e2019JD032362. <https://doi.org/10.1029/2019JD032362>
- Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., et al. (2022). Response of the quasi-biennial oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, 148(744), 1490–1518. <https://doi.org/10.1002/qj.3749>
- Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, 67(1), 136–156. <https://doi.org/10.1175/2009JAS3112.1>
- Sagaut, P. (2006). *Large eddy simulation for incompressible flows: An introduction*. Springer Science and Business Media.
- Sato, K., Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., & Takahashi, M. (2009). On the origins of mesospheric gravity waves. *Geophysical Research Letters*, 36(19), L19801. <https://doi.org/10.1029/2009GL039908>
- Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24), 12–396. <https://doi.org/10.1002/2017gl076101>
- Scinocca, J. F., & McFarlane, N. A. (2000). The parametrization of drag induced by stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological Society*, 126(568), 2353–2393. <https://doi.org/10.1002/qj.49712656802>
- Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(612), 3079–3102. <https://doi.org/10.1256/qj.04.106>
- Shutts, G. J., & Vosper, S. B. (2011). Stratospheric gravity waves revealed in NWP model forecasts. *Quarterly Journal of the Royal Meteorological Society*, 137(655), 303–317. <https://doi.org/10.1002/qj.763>
- Sigmond, M., & Scinocca, J. F. (2010). The influence of the basic state on the northern hemisphere circulation response to climate change. *Journal of Climate*, 23(6), 1434–1446. <https://doi.org/10.1175/2009jcli3167.1>
- Skamarock, W. C. (2004). Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132(12), 3019–3032. <https://doi.org/10.1175/MWR2830.1>
- Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, 227(7), 3465–3485. <https://doi.org/10.1016/j.jcp.2007.01.037>
- Skamarock, W. C., Snyder, C., Klemp, J. B., & Park, S. H. (2019). Vertical resolution requirements in atmospheric simulation. *Monthly Weather Review*, 147(7), 2641–2656. <https://doi.org/10.1175/MWR-D-19-0043.1>
- Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., et al. (2022). Atmospheric energy spectra in global kilometre-scale models. *Tellus A: Dynamic Meteorology and Oceanography*, 74, 280–299. <https://doi.org/10.16993/tellusa.26>
- Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H. (2019). Intercomparison of gravity waves in global convection-permitting models. *Journal of the Atmospheric Sciences*, 76(9), 2739–2759. <https://doi.org/10.1175/JAS-D-19-0040.1>
- Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2023). Explaining the physics of transfer learning in data-driven turbulence modeling. *PNAS nexus*, 2(3), pgad015. <https://doi.org/10.1093/pnasnexus/pgad015>
- Sun, Y. Q., & Zhang, F. (2016). Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *Journal of the Atmospheric Sciences*, 73(3), 1419–1438. <https://doi.org/10.1175/JAS-D-15-0142.1>
- Wang, P., Yuval, J., & O’Gorman, P. A. (2022). Non-local parameterization of atmospheric subgrid processes with neural networks. arXiv preprint arXiv:2201.00417. <https://doi.org/10.48550/ARXIV.2201.00417>
- Wei, J., Bölöni, G., & Achatz, U. (2019). Efficient modeling of the interaction of mesoscale gravity waves with unbalanced large-scale flows: Pseudomomentum-flux convergence versus direct approach. *Journal of the Atmospheric Sciences*, 76(9), 2715–2738. <https://doi.org/10.1175/JAS-D-18-0337.1>
- Wei, J., Zhang, F., Richter, J. H., Alexander, M. J., & Sun, Y. Q. (2022). Global distributions of tropospheric and stratospheric gravity wave momentum fluxes resolved by the 9-km ECMWF experiments. *Journal of the Atmospheric Sciences*, 79(10), 2621–2644. <https://doi.org/10.1175/JAS-D-21-0173.1>
- Yuval, J., & O’Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-17142-3>
- Žagar, N., Kasahara, A., Terasaki, K., Tribbia, J., & Tanaka, H. (2015). Normal-mode function representation of global 3-D data sets: Open-access software for the atmospheric research community. *Geoscientific Model Development*, 8(4), 1169–1195. <https://doi.org/10.5194/gmd-8-1169-2015>

- Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47(17), e2020GL088376. <https://doi.org/10.1029/2020gl088376>
- Zanna, L., & Bolton, T. (2021). Deep learning of unresolved turbulent ocean processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 298–306.
- Zhou, Z., He, G., Wang, S., & Jin, G. (2019). Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Computers and Fluids*, 195, 104319. <https://doi.org/10.1016/j.compfluid.2019.104319>