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

Y. Qiang Sun[1], Pedram Hassanzadeh[1], M. Joan Alexander[2], Christopher G. Kruse[2]

[1]Rice University, Houston, Texas, United States
[2]NorthWest Research Associates, Boulder, Colorado, United States

Key Points:

- In a library of WRF simulations, we compare methods for estimating 3D gravity wave drag force that are un- and under-resolved by GCMs.
- 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.

Corresponding author: Y. Qiang Sun, `ys91@rice.edu`

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 (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 (WRF). 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 $O(1\ km)$ to $O(100\ km)$ 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 (SGS) drag (GWD), resulting from the propagation and breaking of these waves. Existing SGS parameterizations for GWD in general circulation models (GCMs) are all highly simplified; e.g., 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 $\sim 1\ km$ to $1000\ km$, 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; Palmer et al., 1986; Fritts & Alexander, 2003; 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 to 100 km (Fritts & Nastrom, 1992; Eyring et al., 2016; 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 GCM and other information of the sub-grid scale sources. After decades of devel-

opment, gravity wave 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., McFarlane, 1987; Scinocca & McFarlane, 2000; Y. Kim et al., 2003; Beres, 2004; Alexander et al., 2010; Richter et al., 2010; Lott et al., 2012; Plougonven & Zhang, 2014). 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 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), often 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 (Bölöni et al., 2016; Achatz et al., 2017; 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, e.g., 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 impacts. (Sigmond & Scinocca, 2010; Richter et al., 2020, 2022).

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., Schneider et al., 2017; Rasp et al., 2018; Bolton & Zanna, 2019; Maulik et al., 2019; Chattopadhyay et al., 2020; Yuval & O'Gorman, 2020; Kashinath et al., 2021; Gentine et al., 2021; Guan et al., 2022). 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 dataset 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, e.g., 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 this 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., G. J. Shutts & Vosper, 2011; Kruse & Smith, 2015; Žagar et al., 2015; Stephan et al., 2019; Matsuoka et al., 2020; Amiramjadi et al., 2022; Polichtchouk et al., 2022). Some studies pursued a simple box-averaging method (e.g., Matsuoka et al., 2020) or a cut-off low-pass 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., Sato et al., 2009; Muraschko et al., 2015; Ehard et al., 2017). 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 use data from a library of 20 convection-permitting (3 km) tropical WRF simulations to

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 i) Helmholtz decomposition, ii) Spatial filtering to compute the full SGS stress and the resulting GWD, and iii) Same as (ii) 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., Zhou et al., 2019; Zanna & Bolton, 2021; Grooms et al., 2021; Beck & Kurz, 2021; Guan 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 WRF setup and data

All data used in this study are generated using WRF, with initial conditions from reanalysis data and boundaries nudged towards 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 ($\sim$ 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 prevent 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.

For now, the library only includes the tropical regions (see the domains in Fig. 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, QBO phases, and precipitation distribution (Fig. 1c). Two of these simulations, one from the 2016 summertime all-ocean West Pacific (WP), which is in the westerly phase of QBO, and one from the 2020 wintertime land-ocean Indian Ocean (IO), 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 3600 $km\times$ 3600 km. The simulations are done at 3 km grid spacing. There are 180 vertical model levels in total. The vertical grid spacing is close to 200 m near the lower boundary and gradually increases to a maximum of 600 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 RRTM (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 (ECMWF) reanalysis data (ERA5). As we are using the same method as in Kruse et al. (2022) to nudge the simulation boundaries towards the ERA5 data (hence limiting the GW signals there), the model output data near the domain boundary ($<$ 300 km) are neglected when conducting analysis for the GWs. The analysis domains are hence 3000 $km\times$ 3000 km.

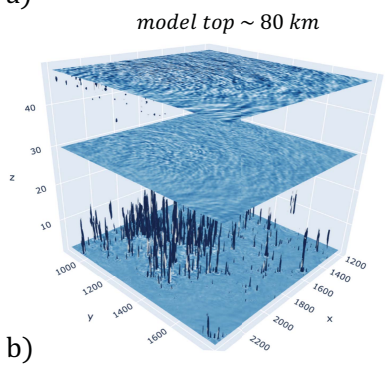
In addition to traditional prognostic variables (e.g., u, v, w, T, p, q), we also modified the WRF model to add diagnostic variables like 3D reflectivity and 3D diabatic heating, which are the key sources for the GWs in the tropics. The output frequency is every 15 minutes in order to capture the life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

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 $\overline{(\cdot)}$. For any variable $\phi(\boldsymbol{x}, t)$, spatial filtering is defined as (e.g., Sagaut, 2006; Grooms et al., 2021; Guan et al., 2022)

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

a)



c)

Region	Simulation Time	Region	Simulation Time
IO	2016/01 2017/05 2018/09 2020/01	MC	2017/05 2018/01 2020/01 2020/11
WP	2016/08 2017/11 2020/01 2020/12	AF	2017/05 2019/08 2020/01
EP	2016/04 2017/05 2020/01	SA	2017/05 2020/01

b)

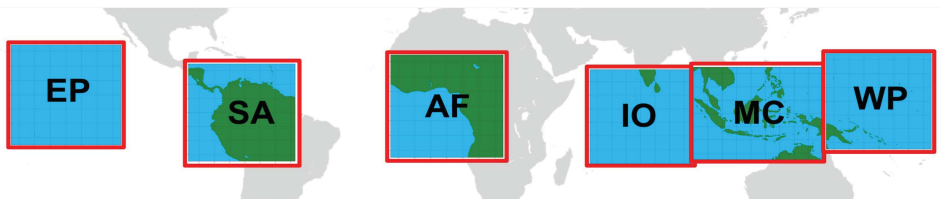


Figure 1. a) A snapshot of vertical velocity in a WRF simulation. b) Tropical regional domains where the WRF simulations are conducted. c) The list of 20 week-long WRF simulations conducted in this regional tropical channel setup, among which two representative cases are chosen (blue color represents Case 1, and magenta color represents Case 2).

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 $\phi(\boldsymbol{x}, t)$ can be separated into two components

$$\phi = \tilde{\phi} + \phi' \tag{2}$$

where $\tilde{\phi}$ contains the large scales (larger than Δ) and ϕ' containts 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., Sagaut, 2006; Grooms et al., 2021; Guan et al., 2022), we define "filtering" as an operation that only separates the scales but does not change the grid resolution (e.g., all 3 terms in Eq. (2) remain on the high-resolution grid). "Coarse-graining", defined later in this section, is the operation that changes resolution, e.g., from the WRF's high-resolution to a GCM's low-resolution grid.

Second, it should be highlighted that Eq. (2), while it appears analogous, is not the same as Reynolds decomposition in this application. This is because spatial filtering (Eq. (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 (Leonard, 1975; Clark et al., 1979; Sagaut, 2006; Alfonsi, 2009). The importance of this distinction will become clear later in the Results section, and has been already pointed out in a number of other studies, e.g., on quantifying (and even determining the sign of) momentum exchange between atmosphere and ocean at small scales (e.g., Aluie et al., 2018; Rai et al., 2021).

A major question in using Eq. (2) is the choice of filter type and size (length scale, Δ in Eq. (1)). As described below, in the 3 methods used here (and generally, in many other methods), Eq. (2) might be used to separate GWs from the large-scale flow, or to separate

Table 1. The 2D low-pass spatial filters used in this study (Eq. (1)). All filters are implemented in spectral space using their transfer function (e.g., Guan et al., 2022). Here, $\boldsymbol{r}$ and $\boldsymbol{k}$ are coordinates in physical space and spectral space, respectively, with $\boldsymbol{r} = (r_x, r_y)$, and $\boldsymbol{k} = (k_x, k_y)$. $\hat{(.)}$ is the Fourier transform, and Δ is the filter size as in Eq. (1).

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

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 even further complicated. Even in a GCM with grid spacing dx, GWs with wavelength larger than $2dx$ may not be 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 sizes, here, we use 3 filter types and two length scales $\Delta = 200\ km$ 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, i.e., twice the low-resolution model's grid spacing (Pope, 2000; Sagaut, 2006; Guan et al., 2022). $\Delta = 700\ km$ 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 use one commonly used coarse-graining strategy: we simply truncate the wavenumbers greater than the cut-off wavenumber corresponding to the GCM grid spacing (100 km in this case). 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 & Zhang (2016) to reduce problems with non-periodic boundaries.

To better illustrate the effects of these filtering and coarse-graining operations, Fig. 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 $3000\ km \times 3000\ km$ 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 ($\tilde{u}$ and $\tilde{v}$) 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 $\overline{\tilde{\phi}}$ and $\overline{\phi}$, which implies that $\tilde{\phi}' \neq 0$ with the Gaussian filtering applied here.

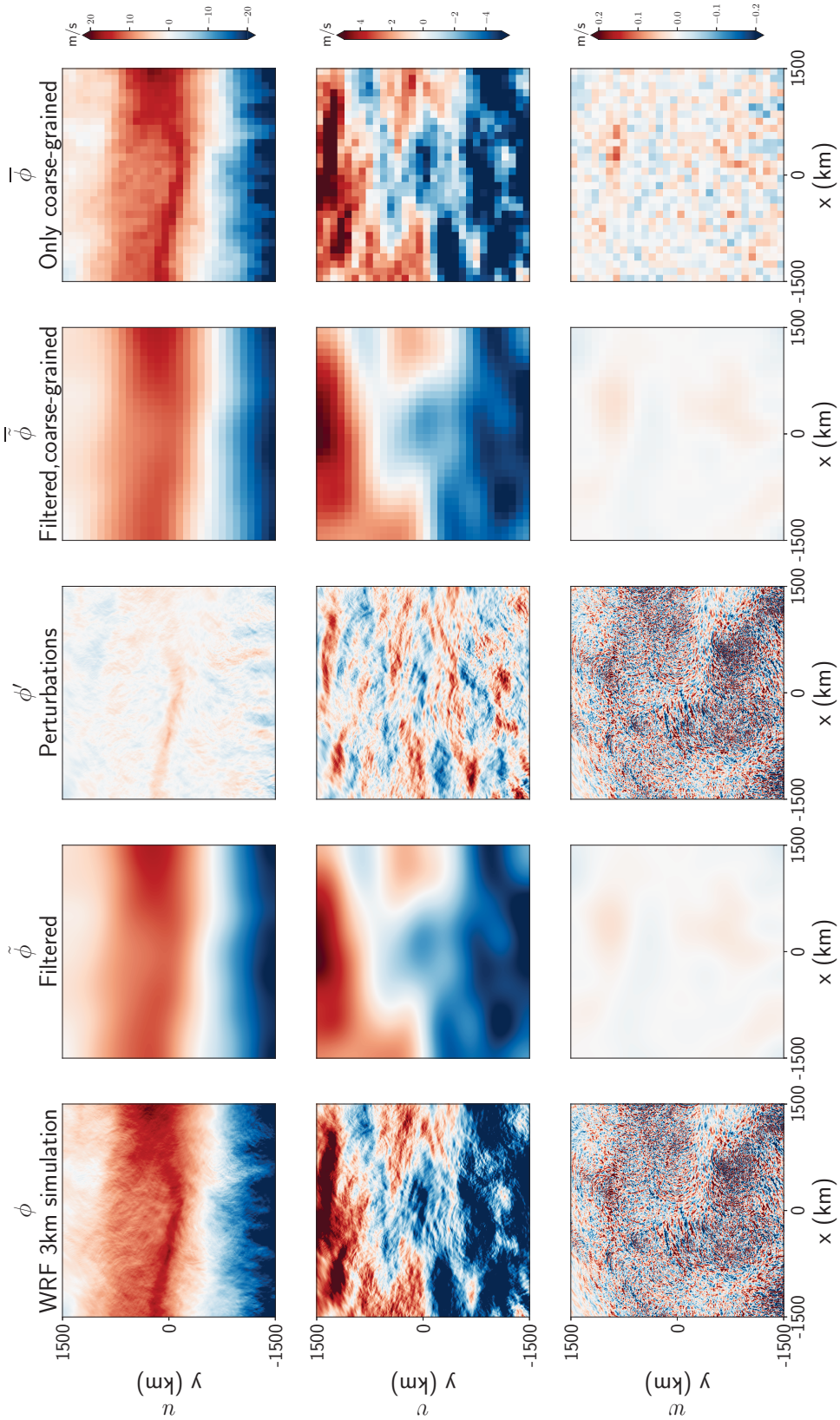


Figure 2. Examples of the effects of low-pass filtering and coarse-graining operators used in this study. The snapshots are from Case 1, on August 3rd 2016, 12:00 UTC, at 30 km height. A Gaussian filter with $\Delta = 700\ km$ is used. The coarse-graining is done by truncating all wavenumbers greater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

2.3 Gravity wave drag (GWD) 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 sub-grid scale (UUGS) method*: This is the common practice in the LES community for computing the fluxes or drags that need to be parameterized (e.g., Leonard, 1975; Germano, 1992; 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., Maulik et al., 2019; Zanna & Bolton, 2020; Yuval & O'Gorman, 2020; Guan et al., 2022; Subel et al., 2022). 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., Clark et al., 1979; Kruse & Smith, 2015; Polichtchouk et al., 2022; Amiramjadi 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 towards ERA5 reanalysis data, the following novel procedure is proposed to avoid the boundary-condition dependency for the decomposition of our WRF simulations:

a) First, the high-resolution WRF data are regridded to the 0.25° ERA5 grid within the WRF domain using conservative interpolation.
b) 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.
c) 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 Fig. S1 for an example of the global field and its rotational and divergence components.
d) 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.
e) 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 (d). 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 (e) 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 (Fig. 2), we simply apply a high-pass filter as in (e) to the full fields to get w'.

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

$$MF_x = [\, MF_{xx} \;\; MF_{yx} \;\; MF_{zx} \,] \tag{3}$$

where

$$MF_{xx} = \tilde{\rho}\widetilde{u'_H u'_H} \tag{4}$$

$$MF_{yx} = \tilde{\rho}\widetilde{u'_H v'_H} \tag{5}$$

$$MF_{zx} = \tilde{\rho}\widetilde{u'_H w'}. \tag{6}$$

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 (e), this is done using spatial filtering, which is also the approach used by Kruse & Smith (2015) in their analyses of GWs.

The first two components in Eq. (3) 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 Eq. (3) 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 Eq. (3) 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 $\overline{MF}_x$,

$$\overline{MF}_x = [\, \overline{MF}_{xx} \;\; \overline{MF}_{yx} \;\; \overline{MF}_{zx} \,]. \tag{7}$$

Based on Eq. (7), the zonal SGS GWD after coarse-graining ($\overline{GWD}_x = \overline{GWD}_{xx} + \overline{GWD}_{yx} + \overline{GWD}_{zx}$), which is what needed to train a data-driven GWP, can be calculated as the divergence of $\overline{MF}_x$:

$$\overline{GWD}_{xx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial \overline{MF}_{xx}}{\partial x} \tag{8}$$

$$\overline{GWD}_{yx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial \overline{MF}_{yx}}{\partial y} \tag{9}$$

$$\overline{GWD}_{zx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial \overline{MF}_{zx}}{\partial z}. \tag{10}$$

Note that Eq. (10), 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 lateral divergence of lateral momentum fluxes (Eqs. (8) and (9)) could also play a substantial role, as will be also shown here later in the Results section.

2.3.2 Method 2: UUGS 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

$$\begin{aligned}\overline{GWD}_x &= \overline{GWD}_{xx} + \overline{GWD}_{xy} + \overline{GWD}_{xz} \\ &= -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial x}\left[\bar{\tilde{\rho}}(\widetilde{\overline{uu}} - \bar{\tilde{u}}\,\bar{\tilde{u}})\right] - \frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial y}\left[\bar{\tilde{\rho}}(\widetilde{\overline{uv}} - \bar{\tilde{u}}\,\bar{\tilde{v}})\right] - \frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial z}\left[\bar{\tilde{\rho}}(\widetilde{\overline{uw}} - \bar{\tilde{u}}\,\bar{\tilde{w}})\right]. \quad (11)\end{aligned}$$

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 Eqs. (8)-(10), the zonal SGS GWD in Eq. (11) 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 Eq. (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 Eq. (11). For example, as shown in Eq. (A13), the Reynolds stress $\widetilde{\overline{u'w'}}$ is a part of the (but not the entire) total SGS vertical flux, $\widetilde{\overline{uw}} - \bar{\tilde{u}}\,\bar{\tilde{w}}$. The other two components (e.g., $\bar{\tilde{u}}\widetilde{\overline{w'}}$) arise because as mentioned before, in spatial filtering and coarse-graining, terms like $\widetilde{\overline{w'}}$ are not necessarily zero (e.g., Pope, 2000; Sagaut, 2006). Similar analysis can be done for $\overline{GWD}_{xx}$ and $\overline{GWD}_{xy}$, showing the appearance of Reynolds stresses $\widetilde{\overline{u'u'}}$ and $\widetilde{\overline{u'v'}}$ as well as other stresses, including $\bar{\tilde{u}}\widetilde{\overline{u'}}$ and $\bar{\tilde{v}}\widetilde{\overline{u'}}$. Different from the HELM_D method that only considers direct contributions of SGS GW perturbations to the GWD, the UUGS_D method (Eq. (11)) 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 (Kruse et al., 2016; Matsuoka et al., 2020; Amiramjadi et al., 2022); hence, the components of $\overline{GWD}_x$ can be written as

$$\overline{GWD}_{xx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial(\bar{\tilde{\rho}}\,\widetilde{\overline{u'u'}})}{\partial x} \quad (12)$$

$$\overline{GWD}_{yx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial(\bar{\tilde{\rho}}\,\widetilde{\overline{u'v'}})}{\partial y} \quad (13)$$

$$\overline{GWD}_{zx} = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial(\bar{\tilde{\rho}}\,\widetilde{\overline{u'w'}})}{\partial z}, \quad (14)$$

though often only $\overline{GWD}_{zx}$ is considered. There are two ways to interpret these equations. First, one can obtain Eqs. (12)-(14) if only the Reynolds stresses in Eq. (11) are accounted for, and the other stresses, including cross-scale interactions are ignored. Second, Eqs. (12)-(14) are the same as Eqs. (8)-(10) if the GW perturbations are identified using filtering (e.g., as $u' = u - \tilde{u}$) rather than as the divergent component of the wind field.

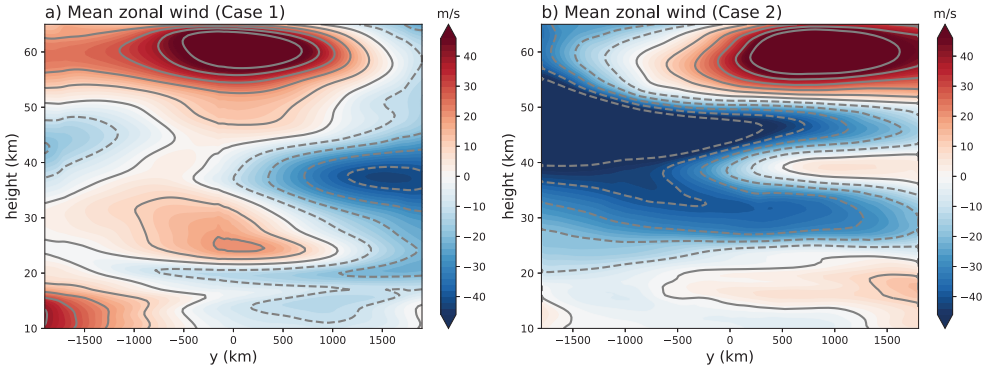


Figure 3. Zonal-mean, time-mean zonal wind u for two representative WRF simulation cases. The contour interval is 10 m, with dash lines implying zero and negative values. a) Case 1 with westerly QBO winds; b) Case 2 with easterly QBO winds .

3 Results

3.1 Zonal-mean, time-mean zonal wind in the WRF simulations

Figure 3 shows the zonal-mean zonal wind averaged over the 5-day simulation period in the two representative cases (see Table 1). The QBO winds are clear in both plots (different phases), with a maximum of $\sim 20\ m/s$ in the tropical stratosphere ($\sim 25\ km$). The semiannual oscillation (SAO) can also be seen near the stratopause ($\sim 60\ km$), 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 SGS GWs

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 Fig. 4. The left column shows zonal SGS GWD calculated using Eq. (10) with the HELM_D method. The zonal SGS GWD in the middle column is based on the REYN_D method (Eq. (14)), 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 Eq. (11)) with the same low-pass Gaussian filter.

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 Eq. (14) 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

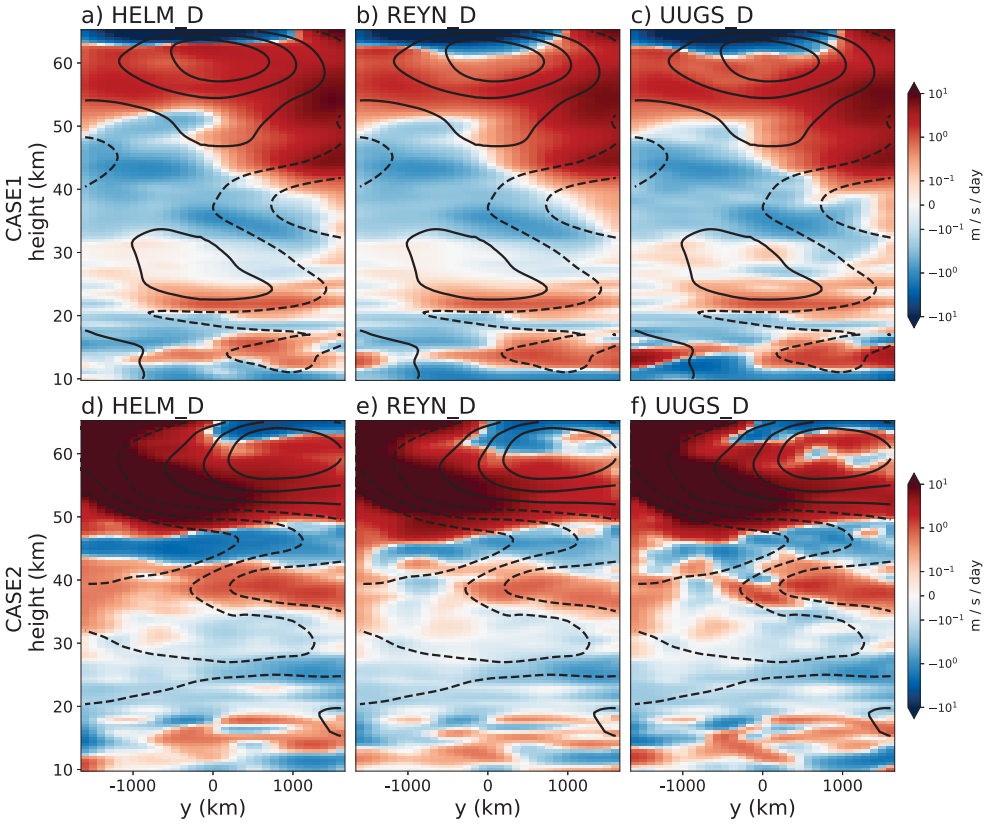


Figure 4. Zonal-mean, time-mean zonal SGS GWD (shading) due to the un-/under-resolved (SGS) vertical fluxes derived using three methods for both representative cases (upper row: Case 1 , lower row: Case 2). Left column, a & d: Helmholtz decomposition method (HELM_D). Middle column, b & e: Reynolds stress method (REYN_D). Right column, c & f: UUGS method (UUGS_D). As in Fig. 3, lines show the mean zonal winds, but with a contour interval of 20 m/s.

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 Fig. 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 Fig. 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

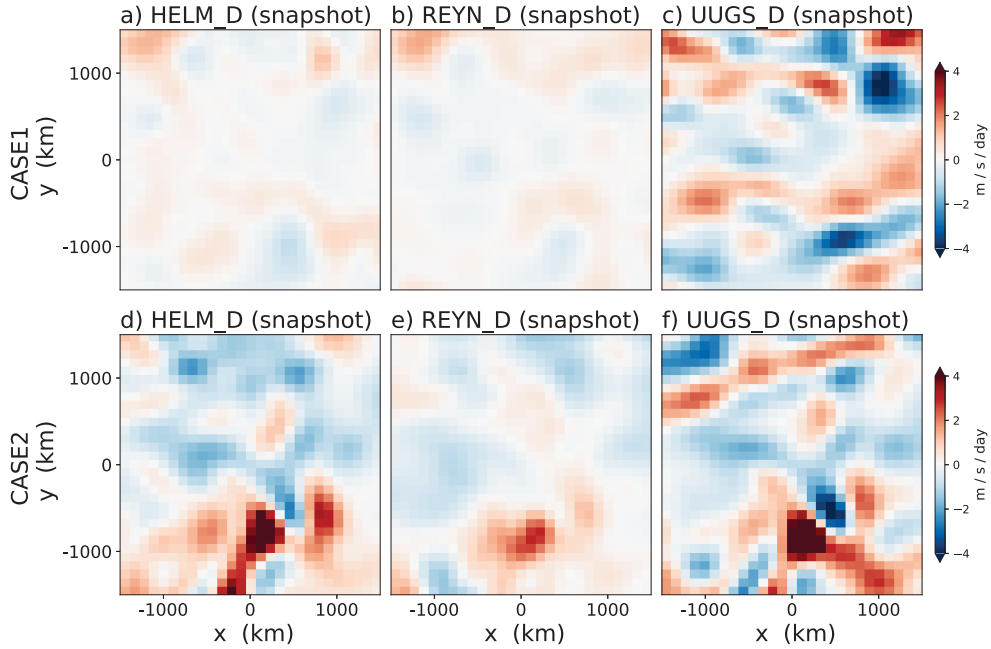


Figure 5. Snapshots of zonal SGS GWD due to vertical fluxes at 30 km height calculated using three methods. Upper row: Case 1. Lower row: Case 2. Left column, a & d: Helmholtz decomposition method $\frac{1}{\bar{\tilde{\rho}}}\frac{\partial\overline{\tilde{\rho}\widetilde{u'_H w'}}}{\partial z}$. Middle column, b & e: Reynolds stress method $\frac{1}{\bar{\tilde{\rho}}}\frac{\partial\overline{\tilde{\rho}\widetilde{u'w'}}}{\partial z}$. Right column, c & f: UUGS method $-\frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial z}\left[\bar{\tilde{\rho}}(\overline{\tilde{u}\tilde{w}}-\bar{\tilde{u}}\,\bar{\tilde{w}})\right]$.

differences among them. While the PDFs we obtain using the HELM_D and REYN_D methods are fairly similar, the PDFs from the UUGS method often have a clearly 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., Eq. (14)), 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 (Fig. 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 is the same as Fig. 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).

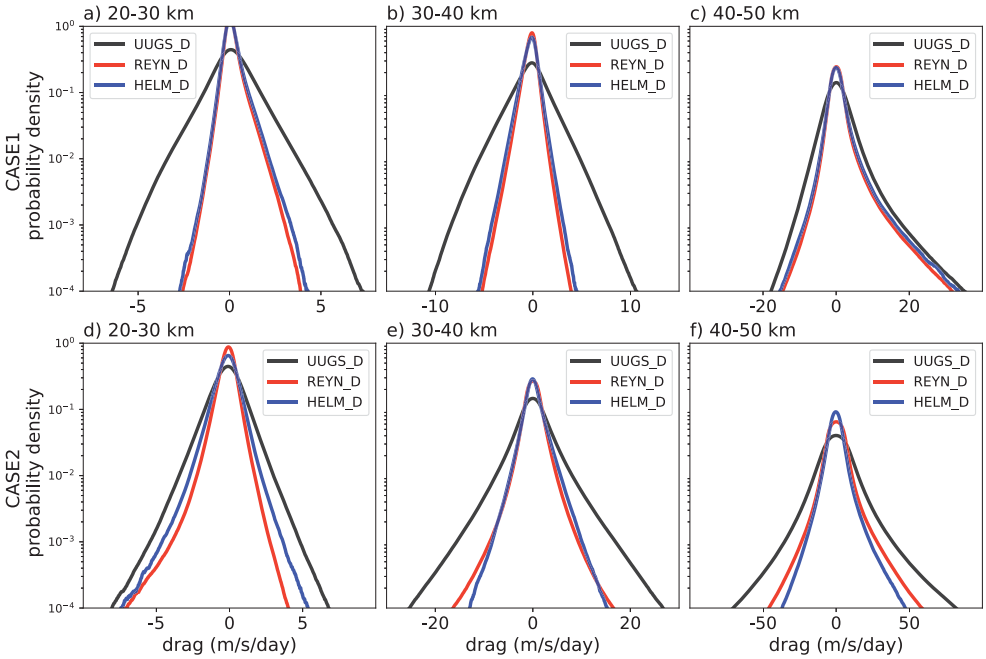


Figure 6. Probability density function (PDF) of zonal SGS GWD due to vertical fluxes calculated using three methods at different heights. Upper row: Case 1. Lower row: Case 2. Left column, a & d: 20-30 km. Middle column, b & e: 30-40 km. Right column, c & f: 40-50 km. Note the differences in the x-axes.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

3.3 Horizontal flux of zonal momentum due to SGS GWs

In addition to the SGS vertical fluxes, the SGS horizontal momentum fluxes associated with GWs could also lead to zonal SGS GWD (see Eqs. (11)-(13)). 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., G. J. Shutts & Vosper, 2011; Kruse et al., 2022). 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 Eq. (11)) vs. the total GWD calculated using the entire Eq. (11) and their differences, i.e., 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

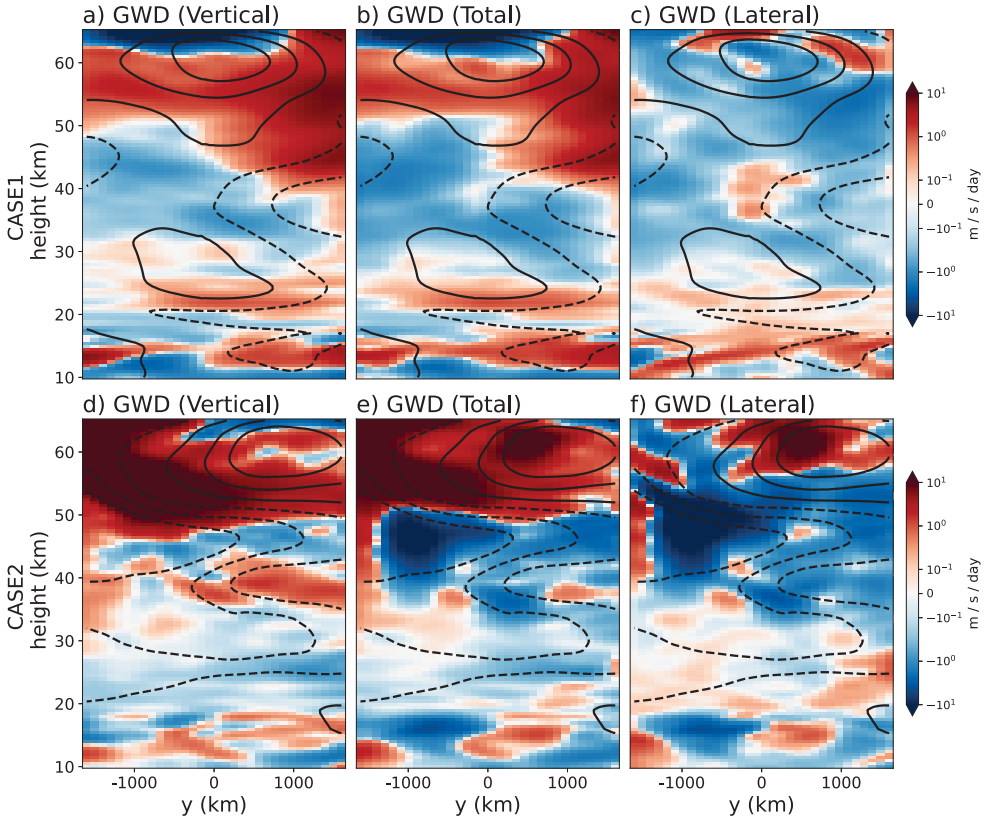


Figure 7. Zonal-mean, time-mean zonal SGS GWD due to only the vertical fluxes, total fluxes, and the lateral fluxes calculated using the UUGS_D method (Eq. (11)). Left column: Only the vertical component $-\frac{1}{\bar{\bar{\rho}}}\frac{\partial}{\partial z}\left[\bar{\bar{\rho}}(\overline{\widetilde{uw}}-\bar{\bar{u}}\,\bar{\bar{w}})\right]$. Middle column: All terms $-\frac{1}{\bar{\bar{\rho}}}\frac{\partial}{\partial x}\left[\bar{\bar{\rho}}(\overline{\widetilde{uu}}-\bar{\bar{u}}\,\bar{\bar{u}})\right]-\frac{1}{\bar{\bar{\rho}}}\frac{\partial}{\partial y}\left[\bar{\bar{\rho}}(\overline{\widetilde{uv}}-\bar{\bar{u}}\,\bar{\bar{v}})\right]-\frac{1}{\bar{\bar{\rho}}}\frac{\partial}{\partial z}\left[\bar{\bar{\rho}}(\overline{\widetilde{uw}}-\bar{\bar{u}}\,\bar{\bar{w}})\right]$. Right column: Only the first two terms, i.e., only the horizontal fluxes. As in Fig. 3, lines show the mean zonal winds, but with a contour interval of 20 m/s.

another example, in Case 2, at levels below the SAO ($\sim 50\ km$), 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.

Examining the PDFs of zonal SGS GWD, which highlights its variability, further shows the importance of the SGS horizontal fluxes. Similar to Fig. 6, Fig. 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 (Fig. 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.

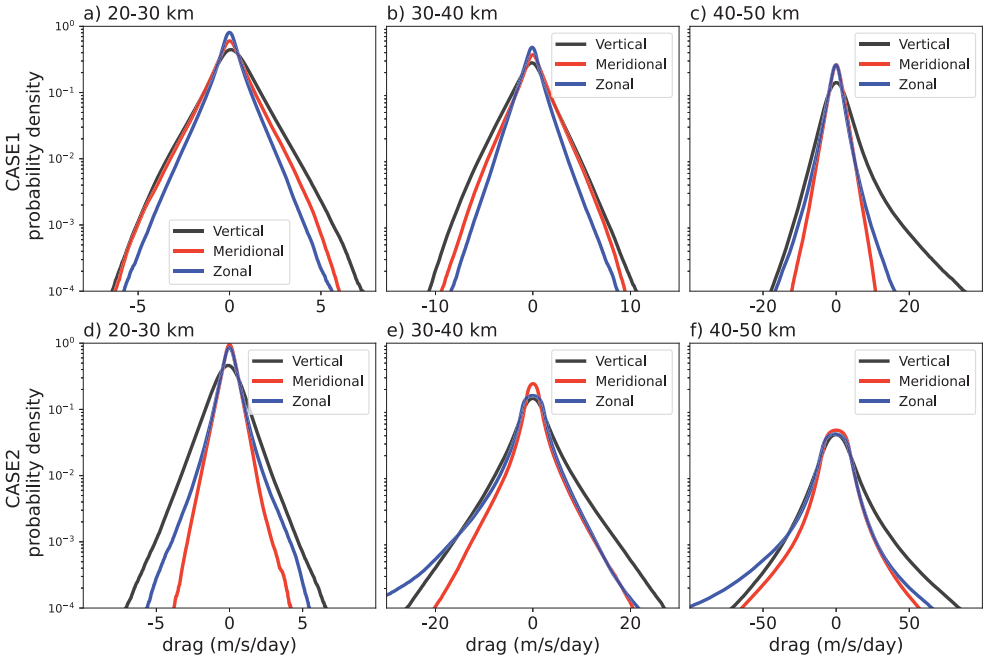


Figure 8. Probability density function (PDF) of zonal SGS GWD due to SGS zonal (blue), meridional (red), and vertical (black) momentum fluxes for both cases. Upper row: Case 1; Lower panel: Case 2. Left column, a & d: 20-30 km. Middle column, b & e: 30-40 km. Right column, c & f: 40-50 km. Note the difference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

So far, we have used the UUGS_D method to calculate the GWD due to SGS horizontal fluxes (Figs. 7-8). 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 Eq. (11)), respectively. Different from Fig. 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 Fig. 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 Fig. 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 Figs. 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 (Fig. 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

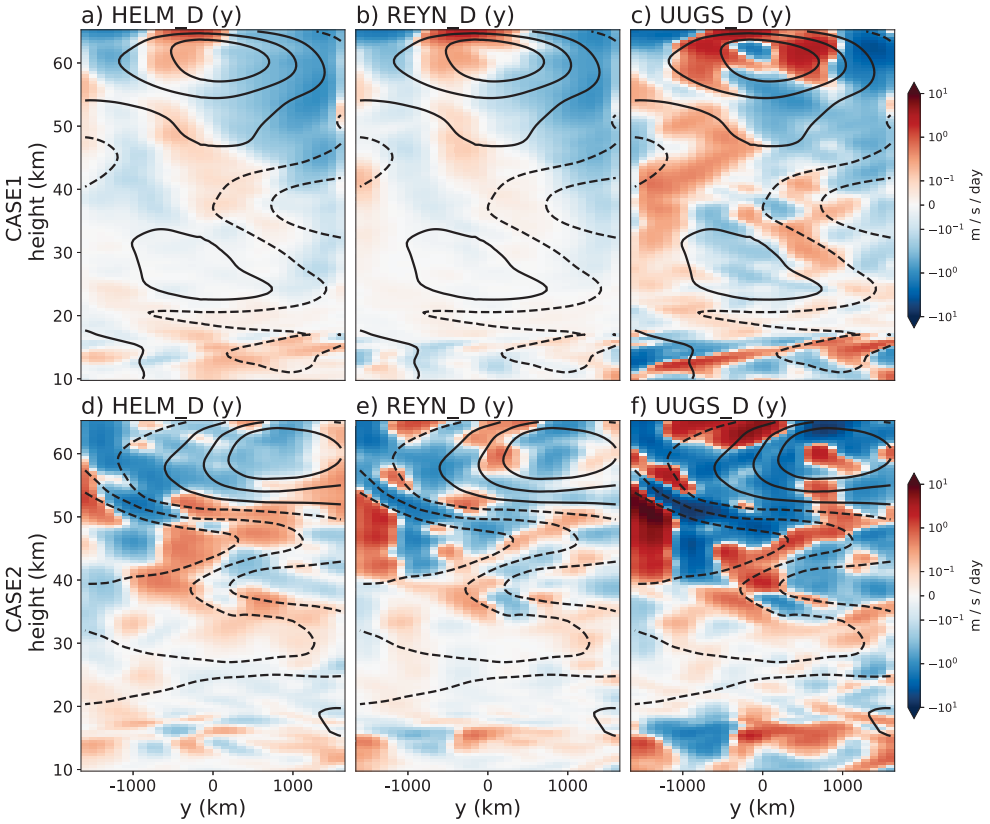


Figure 9. Zonal-mean, time-mean zonal SGS GWD, similar to Fig. 4, but for the GWD due to the un-/under-resolved (SGS) horizontal fluxes. Only the meridional direction is shown here; examining the zonal direction shows a similar story.

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 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 is the same as Fig. 10, but with data from all 20 cases combined.

3.4 Sensitivity to the filter type/size and the GCM 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.

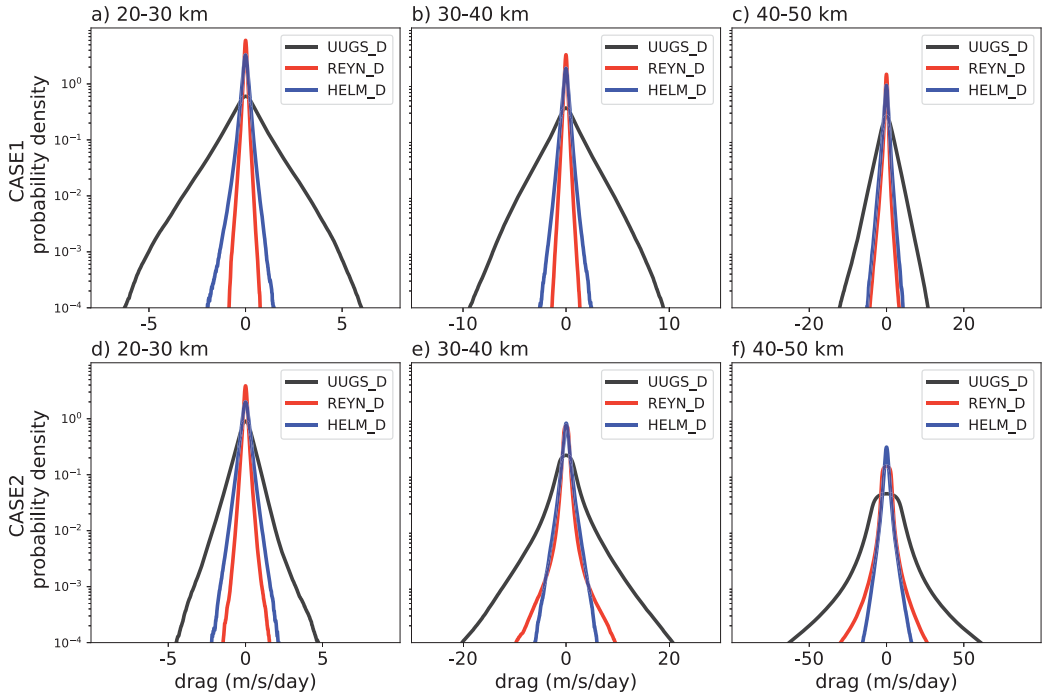


Figure 10. Same as Fig. 6, but for the un- and under-resolved zonal GWD due to SGS meridional momentum fluxes. Note the difference in the x-axes. See Fig. S4 for the PDFs of zonal GWD due to the SGS zonal momentum fluxes. See Fig. S5 for the PDFs of data from all 20 cases combined.

As mentioned earlier, Fig. 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 larger than the GCM's grid spacing, i.e., 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 LES literature (Pope, 2000; Sagaut, 2006; Guan et al., 2022). 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.

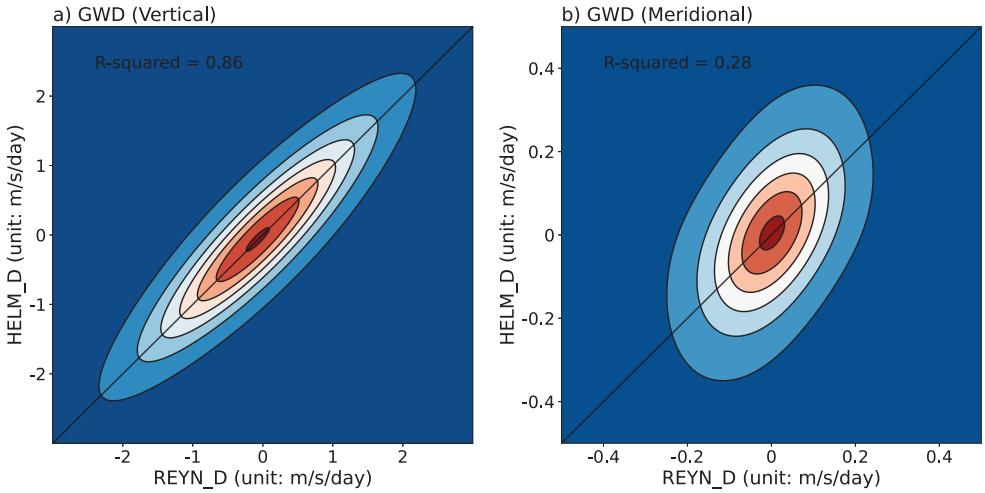


Figure 11. Joint PDFs between the SGS GWD calculated using the HELM_D and REYN_D methods for Case 1. a) GWD due to SGS vertical fluxes and b) GWD due to SGS meridional fluxes. The R-squared value is shown to measure the correlation between these two methods.

Figure 12 shows PDFs of the 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) and $\Delta = 200$ 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, the zonal SGS GWD may not be necessarily reduced, as less averaging of the momentum flux is also applied with a smaller Δ. As a result, in both cases, for the zonal SGS GWD associated with vertical fluxes (solid lines), we find larger values when the filter length scale is reduced (compare the tails of the PDFs in the left and middle columns). The zonal SGS GWD associated with horizontal fluxes may become larger or smaller depending on the case.

Moreover, Fig. 12 also shows that the SGS GWD is not reduced with a smaller GCM grid spacing, and in fact, might become even larger in some cases due to the effects of increased gradient(see panel f), which suggest the need of 3D GWP even in a high-resolution GCM.

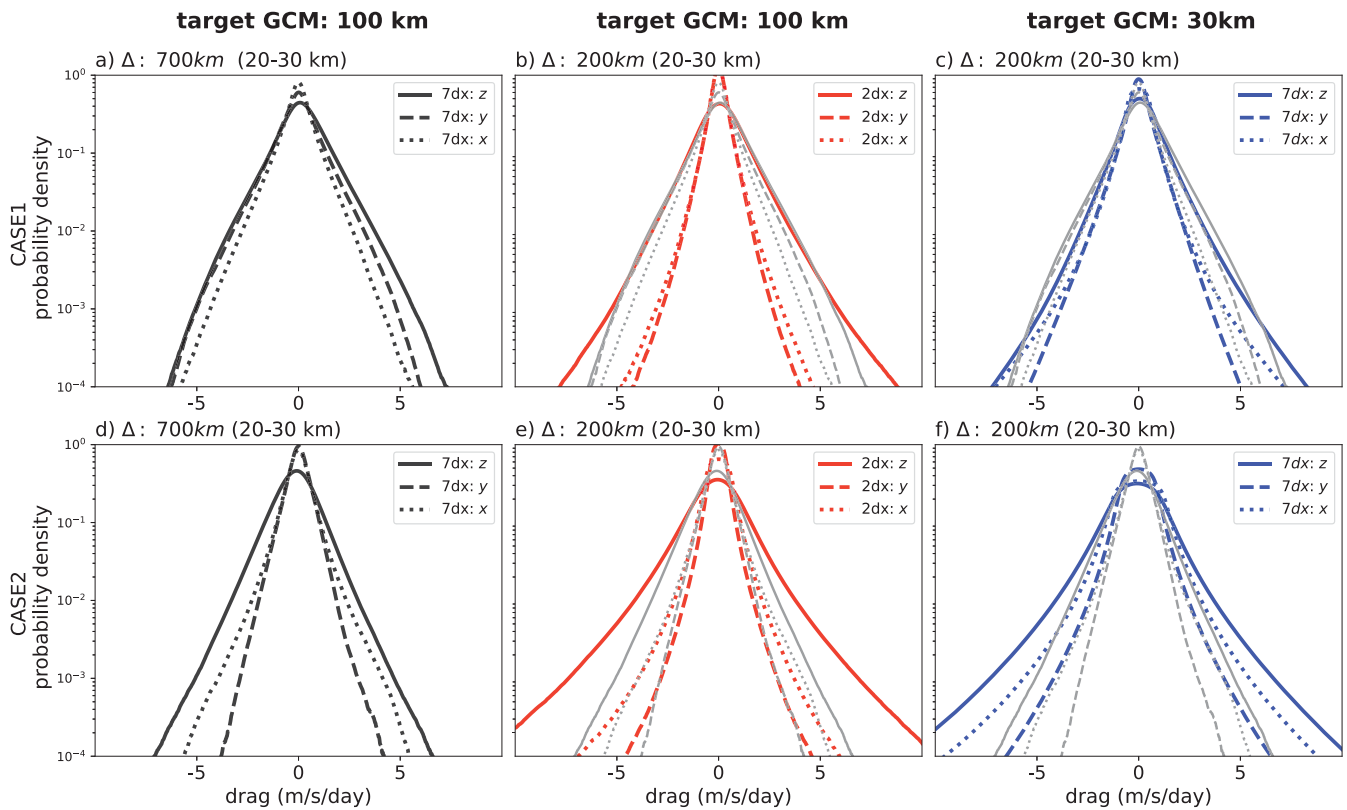


Figure 12. The PDFs of GWD at $20-30$ km heights associated with the SGS zonal (dotted), meridional (dashed), and vertical (solid) fluxes extracted using the UUGS_D method (with Gaussian filter) for 3 different sets of filter size Δ and GCM grid spacing. Left column: $\Delta = 700$ km, 100 km grid spacing. Middle columns: $\Delta = 200$ km, 100 km grid spacing. Right column: $\Delta = 200$ km, 30 km grid spacing. 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.

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., Matsuoka et al., 2020; Chantry et al., 2021; Espinosa et al., 2022; Amiramjadi et al., 2022). 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., Beucler et al., 2021; Dunbar et al., 2021; Guan et al., 2023; Mamalakis et al., 2022; Subel et al., 2022; Pathak et al., 2022). 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 dataset 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 (G. Shutts, 2005; Palmer et al., 2005; Lott et al., 2012).
2. There are 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 support 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., Wang et al., 2022).
3. Adding to the compelxity, 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 (Fig. 12). Interestingly, within the explored GCM grid spacing ($30\ km - 100\ km$), 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, e.g., 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, e.g., 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, i.e., 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 datasets 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. 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 (Jeevanjee, 2017; Bramberger et al., 2020). We have conducted limited simulations with 1 km 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. 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., 2022).

Appendix A UUGS Drag Extraction

To better illustrate the GWD paramterization 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:

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u u)}{\partial x} + \frac{\partial(\rho u v)}{\partial y} + \frac{\partial(\rho u w)}{\partial z} = -\frac{\partial p}{\partial x} + \rho f v + \rho F_x, \tag{A1}$$

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

The problem of the parameterization of GWs and/or other sub-grid scale physical processes arises because 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:

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

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

$$\phi = \tilde{\phi} + \phi'. \tag{A3}$$

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 $\overline{()}$, is required to transfer this large-scale part to the GCM grid, so that,

$$\phi^G \approx \overline{\tilde{\phi}} \tag{A4}$$

Applying Eq. (A4) to Eq. (A2) yields

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

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 Eq. (A1), and assume that the operations are commutative [1], which means, e.g.,

$$\overline{\frac{\partial \widetilde{\phi}}{\partial x}} = \frac{\partial \bar{\tilde{\phi}}}{\partial x}. \tag{A6}$$

[1] This assumptions is valid in our study. The three filters used here commute with spatial derivatives if applied on a uniform grid, which is the case for the WRF's horizontal grid. No filtering or coarse-graining

Then we get

$$\frac{\partial(\overline{\widetilde{\rho u}})}{\partial t} + \frac{\partial(\overline{\widetilde{\rho u u}})}{\partial x} + \frac{\partial(\overline{\widetilde{\rho u v}})}{\partial y} + \frac{\partial(\overline{\widetilde{\rho u w}})}{\partial z} = -\frac{\partial \bar{\tilde{p}}}{\partial x} + \overline{\widetilde{\rho f v}} + \overline{\widetilde{\rho F_x}}. \tag{A7}$$

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

$$\rho(x, y, z, t) \approx \tilde{\rho}(z). \tag{A8}$$

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

$$\overline{\widetilde{\rho\phi}} \approx \overline{\tilde{\bar{\rho}}\tilde{\phi}} = \bar{\tilde{\rho}}\bar{\tilde{\phi}} \tag{A9}$$

Subtracting Eq. (A7) from Eq. (A5), and applying Eq. (A9), we finally get:

$$\begin{aligned} \mathbf{X_x^G} &= \frac{\partial(\bar{\tilde{\rho}}\bar{\tilde{u}}\bar{\tilde{u}})}{\partial x} + \frac{\partial(\bar{\tilde{\rho}}\bar{\tilde{u}}\,\bar{\tilde{v}})}{\partial y} + \frac{\partial(\bar{\tilde{\rho}}\bar{\tilde{u}}\bar{\tilde{w}})}{\partial z} - \left[\frac{\partial(\overline{\widetilde{\rho u u}})}{\partial x} + \frac{\partial(\overline{\widetilde{\rho u v}})}{\partial y} + \frac{\partial(\overline{\widetilde{\rho u w}})}{\partial z}\right] \\ &= -\frac{\partial}{\partial x}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uu}} - \bar{\tilde{u}}\bar{\tilde{u}})\right] - \frac{\partial}{\partial y}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uv}} - \bar{\tilde{u}}\,\bar{\tilde{v}})\right] - \frac{\partial}{\partial z}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uw}} - \bar{\tilde{u}}\bar{\tilde{w}})\right]. \end{aligned} \tag{A10}$$

Similarly, for the meridional momentum equation, we obtain

$$\mathbf{X_y^G} = -\frac{\partial}{\partial x}\left[\bar{\tilde{\rho}}(\overline{\widetilde{vu}} - \bar{\tilde{v}}\bar{\tilde{u}})\right] - \frac{\partial}{\partial y}\left[\bar{\tilde{\rho}}(\overline{\widetilde{vv}} - \bar{\tilde{v}}\,\bar{\tilde{v}})\right] - \frac{\partial}{\partial z}\left[\bar{\tilde{\rho}}(\overline{\widetilde{vw}} - \bar{\tilde{v}}\bar{\tilde{w}})\right] \tag{A11}$$

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

$$\overline{GWD}_x = -\frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial x}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uu}} - \bar{\tilde{u}}\,\bar{\tilde{u}})\right] - \frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial y}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uv}} - \bar{\tilde{u}}\,\bar{\tilde{v}})\right] - \frac{1}{\bar{\tilde{\rho}}}\frac{\partial}{\partial z}\left[\bar{\tilde{\rho}}(\overline{\widetilde{uw}} - \bar{\tilde{u}}\,\bar{\tilde{w}})\right], \tag{A12}$$

which is the equation used in the main text.

The terms in brackets on the right-hand side of Eq. (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 (Leonard, 1975; Germano, 1986; Sagaut, 2006), e.g.,

$$\begin{aligned} &\overline{\widetilde{uw}} - \bar{\tilde{u}}\bar{\tilde{w}} \\ &= \overline{(\widetilde{u} + u')(\widetilde{w} + w')} - \overline{(\widetilde{\widetilde{u} + u'})}\,\overline{(\widetilde{\widetilde{w} + w'})} \\ &= \underbrace{\left(\overline{\widetilde{\tilde{u}\tilde{w}}} - \bar{\tilde{\tilde{u}}}\,\bar{\tilde{\tilde{w}}}\right)}_{Leonard\ \ stress} + \underbrace{\left(\overline{\widetilde{\tilde{u}w'}} + \overline{\widetilde{u'\tilde{w}}} - \bar{\tilde{\tilde{w}}}'\bar{\tilde{\tilde{u}}} - \bar{\tilde{\tilde{u}}}'\bar{\tilde{\tilde{w}}}\right)}_{cross\ stress} + \underbrace{\left(\overline{\widetilde{u'w'}} - \bar{\tilde{\tilde{w}}}'\bar{\tilde{\tilde{w}}}'\right)}_{Reynolds\ stress} \end{aligned} \tag{A13}$$

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., Leonard, 1975; Galmarini et al., 2000). 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.

is done in the vertical direction (where WRF's grid is non-uniform). Note that on non-uniform grids, such as GCMs' grids, special treatments are needed; see, e.g., Grooms et al. (2021).

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 tiles 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.

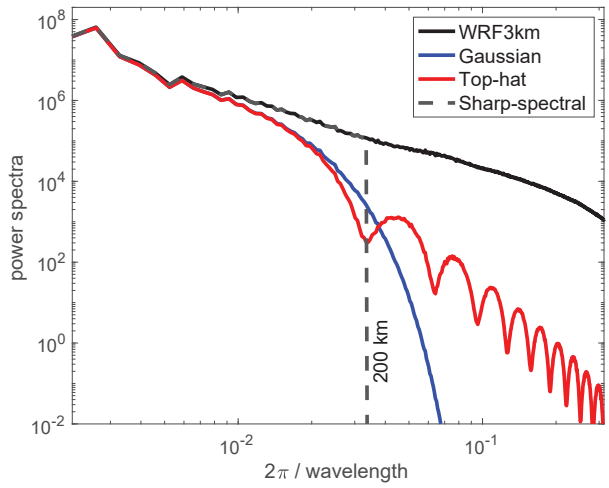


Figure B1. Example of the power spectrum of the zonal wind from the 3-km WRF simulations before and after the low-pass filtering using different kernels (with length scale $\Delta = 200\,km$). The black line shows the spectrum before filtering, while the blue, red, and green lines show the the spectrum after applying the Gaussian, top-hat (box), and sharp-spectral filters. Note that the dashed green line coincides with the black line for scales smaller 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).

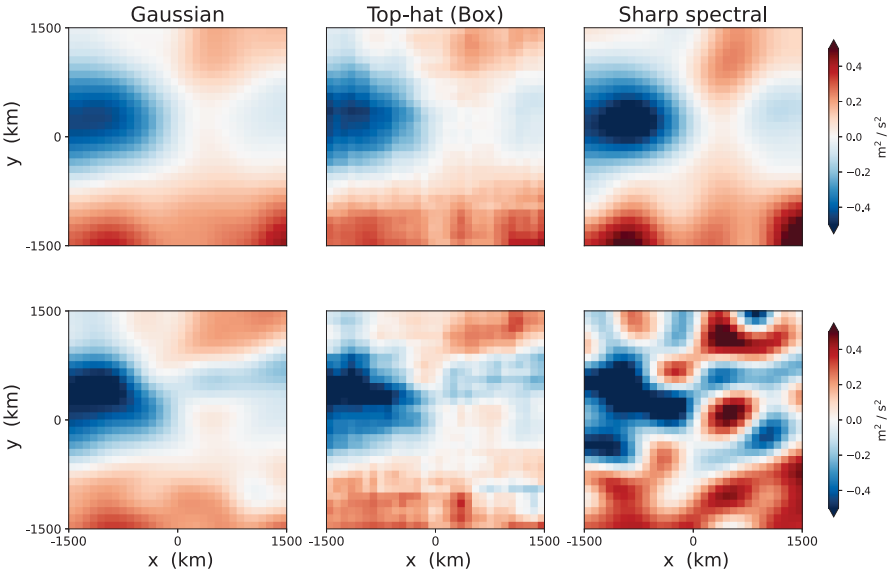


Figure B2. Snapshots of the SGS vertical momentum fluxes, calculated using the 3 different filters for Case 1 (at 40 km height). Top row: the Reynolds stress, $\overline{\widetilde{u'w'}}$. Bottom: the total SGS stress, $\overline{\widetilde{uw}} - \overline{\tilde{u}}\,\overline{\tilde{w}}$). The filter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

Open Research Section

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 data and code for all the analysis in the main text is available at `https://doi.org/10.5281/zenodo.7439397`

Acknowledgments

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*. doi: 10.1002/qj.2926

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., ... Watanabe, S. (2010). Recent developments in gravity-wave effects in climatemodels and the global distribution of gravity-wavemomentum flux from observations and models. *Quarterly Journal of the Royal Meteorological Society*, *136*. doi: 10.1002/qj.637

Alfonsi, G. (2009). Reynolds-averaged Navier–Stokes equations for turbulence modeling. *Applied Mechanics Reviews*, *62*(4).

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*. doi: 10.1175/JPO-D-17-0100.1

Amiramjadi, 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*. doi: 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.

Beres, J. H. (2004). Gravity wave generation by a three-dimensional thermal forcing. *Journal of Atmospheric Sciences*, *61*(14), 1805-1815. doi: 10.1175/1520-0469(2004)061⟨1805:GWGBAT⟩2.0.CO;2

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302.

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.

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*. doi: 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.

Bretherton, F. P. (1969). Momentum transport by gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *95*(404), 213-243. doi: 10.1002/qj.49709540402

Bölöni, 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*. doi: 10.1175/JAS-D-20-0065.1

Bölöni, 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*. doi: 10.1175/JAS-D-16-0069.1

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*. doi: 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*. doi: 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. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002477` (e2021MS002477 2021MS002477) doi: https://doi.org/10.1029/2021MS002477

Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven superparameterization using deep learning: Experimentation with multiscale Lorenz 96 systems and transfer learning. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002084.

Chen, T.-C., & Wiin-Nielsen, A. C. (1976). On the kinetic energy of the divergent and nondivergent flow in the atmosphere. *Tellus*, *28*. doi: 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.

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*. doi: 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.

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. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/72/6/jas-d-14-0147.1.xml` doi: 10.1175/JAS-D-14-0147.1

Ehard, B., Kaifler, B., Dörnbrack, A., Preusse, P., Eckermann, S. D., Bramberger, M., ... Rapp, M. (2017). Horizontal propagation of large-amplitude mountain waves into the polar night jet. *Journal of Geophysical Research*, *122*. doi: 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 CO_2. , *49*(8), e98174. doi: 10.1029/2022GL098174

Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., ... Williams, D. N. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, *7*(4), 813-830. doi: 10.5194/esd-7-813-2016

Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Review of Geophysics*, *41*(1). doi: 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 Atmospheric Sciences*, *49*(2), 111-127. doi: 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*, 2968-2976. doi: 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.

Germano, M. (1992). Turbulence: the filtering approach. *Journal of Fluid Mechanics*, *238*, 325–336.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., ... Randel, W. J. (2019). The whole atmosphere community climate model version 6 (waccm6). *Journal of Geophysical Research: Atmospheres*, *124*, 12380-12403. doi: 10.1029/2019JD030943

Grooms, I., Loose, N., Abernathey, R., Steinberg, J., Bachman, S. D., Marques, G., ... Yankovsky, E. (2021). Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002552.

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.

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. doi: 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*.

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. doi: 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 Atmospheres*, *113*. doi: 10.1029/2008JD009944

Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, *9*. doi: 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. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/76/6/jas-d-18-0290.1.xml` doi: 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*. doi: 10.1175/MWR-D-11-00056.1

Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., ... others (2021). Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200093.

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*, 65-98. Retrieved from `https://doi.org/10.3137/ao.410105` doi: 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*. doi: 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*. doi: 10.1002/asl.952

Kruse, C. G., Alexander, M. J., Hoffmann, L., Niekerk, A. V., Polichtchouk, I., Bacmeister, J. T., ... Stein, O. (2022). Observed and modeled mountain waves from the surface to

the mesosphere near the drake passage. *Journal of the Atmospheric Sciences*, *79*. doi: 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*. doi: 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*. doi: 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*. doi: 10.1029/2012GL051001

Lynch, P. (1988). Deducing the wind from vorticity and divergence. *Monthly Weather Review*, *116*. doi: 10.1175/1520-0493(1988)116⟨0086:DTWFVA⟩2.0.CO;2

Mamalakis, 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.

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. , *47*(19), e89436. doi: 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.

McFarlane, N. A. (1987). The Effect of Orographically Excited Gravity Wave Drag on the General Circulation of the Lower Stratosphere and Troposphere. *Journal of Atmospheric Sciences*, *44*(14), 1775-1800. doi: 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-kramer-brillouin theory for the simulation of the weakly nonlinear dynamics of gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *141*. doi: 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*. doi: 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*. doi: 10.1002/qj.49711247406

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., ... others (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: Atmospheres*, *108*. doi: 10.1029/2002jd003322

Plougonven, R., & Zhang, F. (2014). Internal gravity waves from atmospheric jets and fronts. *Reviews of Geophysics*, *52*. doi: 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*. doi: 10.1002/qj.4202

Pope, S. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, *72*(2), 331–340.

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*. doi: 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.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020, 4). Progress in simulating the quasi-biennial oscillation in cmip models. *Journal of Geophysical Research: Atmospheres*, *125*. doi: 10.1029/2019JD032362

Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., ... Yukimoto, S. (2022, 4). Response of the quasi-biennial oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, *148*, 1490-1518. doi: 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*. doi: 10.1175/2009JAS3112.1

Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Springer Science & 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*. doi: 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.

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*. doi: 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.

Shutts, G. J., & Vosper, S. B. (2011). Stratospheric gravity waves revealed in nwp model forecasts. *Quarterly Journal of the Royal Meteorological Society*, *137*. doi: 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.

Skamarock, W. C. (2004). Evaluating mesoscale nwp models using kinetic energy spectra. *Monthly Weather Review*, *132*. doi: 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*. doi: 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*. doi: 10.1175/MWR-D-19-0043.1

Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., ... Ziemen, F. (2022, 4). Atmospheric energy spectra in global kilometre-scale models. *Tellus A: Dynamic Meteorology and Oceanography*, *74*, 280-299. doi: 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*. doi: 10.1175/JAS-D-19-0040.1

Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2022). Explaining the physics of transfer learning in data-driven turbulence modeling. *arXiv preprint arXiv:2206.03198*.

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. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/`

`73/3/jas-d-15-0142.1.xml` doi: 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*. doi: 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*. doi: 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*. doi: 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.

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376.

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 & Fluids*, *195*, 104319.

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*. doi: 10.5194/gmd-8-1169-2015

Supporting Information for "Quantifying 3D Gravity Wave Drag in a Library of Tropical Convection-permitting Simulations for Data-driven Parameterizations"

Y. Qiang Sun[1], Pedram Hassanzadeh[1], M. Joan Alexander[2], and Christopher G. Kruse[2]

[1]Rice University, Houston, Texas, United States

[2]NorthWest Research Associates, Boulder, Colorado, United States

Contents of this file

1. Figures S1 to S5

Additional Supporting Information (Files uploaded separately)

1. Captions for Movies S1

Introduction

Additional analysis with more cases are added here in the supporting information. In the main text, the results are focused on the zonal GWD for brevity. Results for the meridional GWD is provided here. Also, when discussing the lateral fluxes in the main paper, SGS meridional fluxes are shown; here, we also show the SGS zonal fluxes.

Movie S1. At 3km, WRF is capable of resolving most of the GWs spectrum. Here we provide a movie showing w of the WRF-simulated GWs and the coarse-grained results. The movie will be uploaded separately according to AGU's submission system.

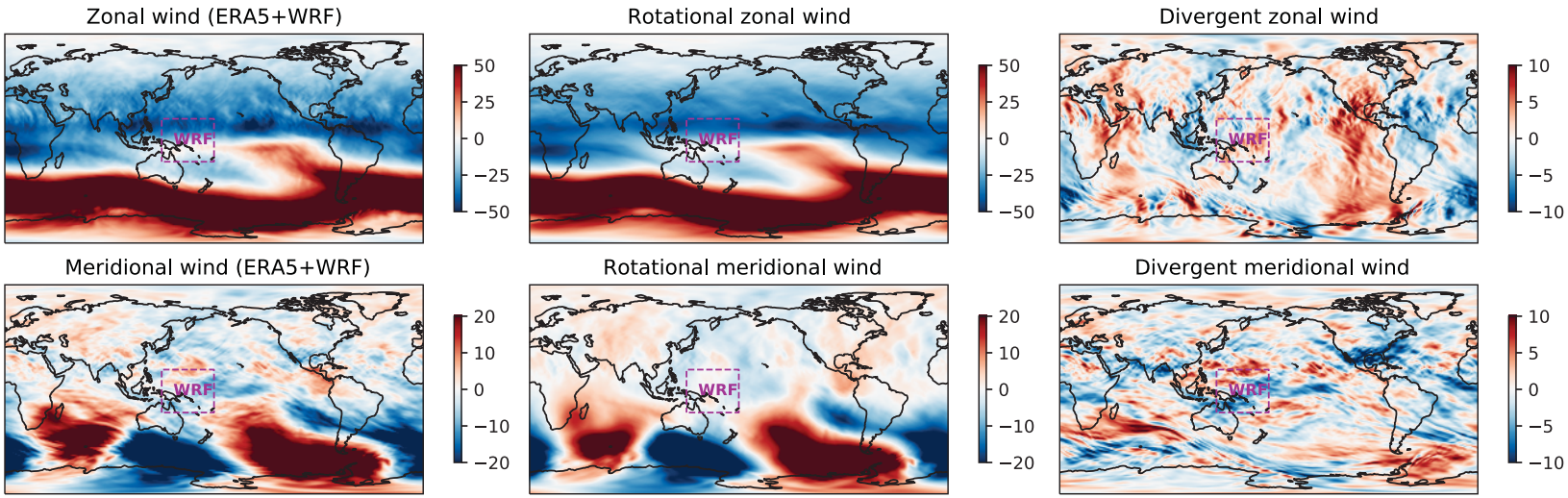


Figure S1. Snapshots from the Helmholtz decomposition approach used in this study. The magenta boxes show the WRF domain. In the left column, data within this magenta box is provided by the WRF simulation; outside this box, the interpolated ERA5 data are used. The results of the Helmholtz decomposition are shown in the middle column (rotational component) and the right column (divergent component). Note the different color scales in the panels (the units are m/s). The code for this decomposition can be found in https://www.ncl.ucar.edu/Applications/wind.shtml

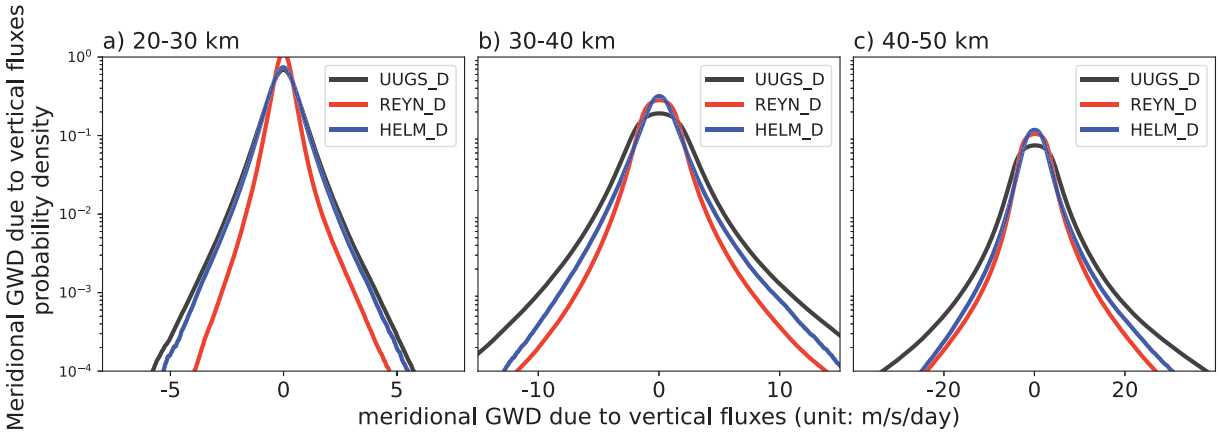


Figure S2. Sames as Fig. 6, but for the meridional GWD due to vertical fluxes (last term of Eq. (A11)). Data from Cases 1 and 2 have been combined in making these PDFs.

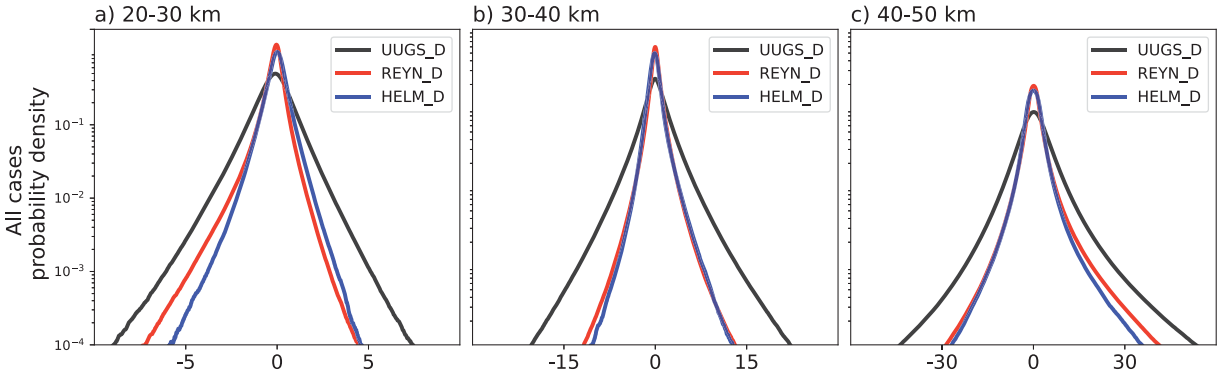


Figure S3. Same as Fig. 6, but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computinonal cost of the algorithm.

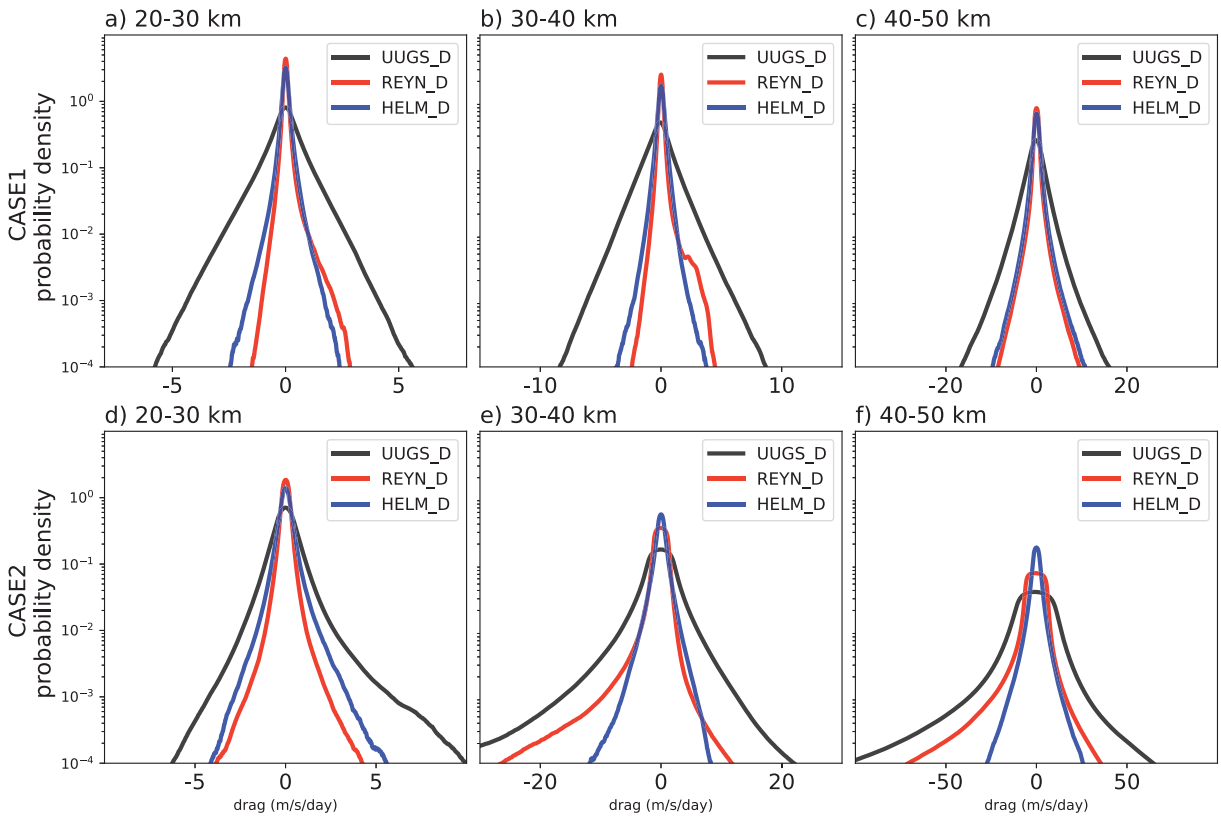


Figure S4. Same as Fig. 10, but for the zonal fluxes.

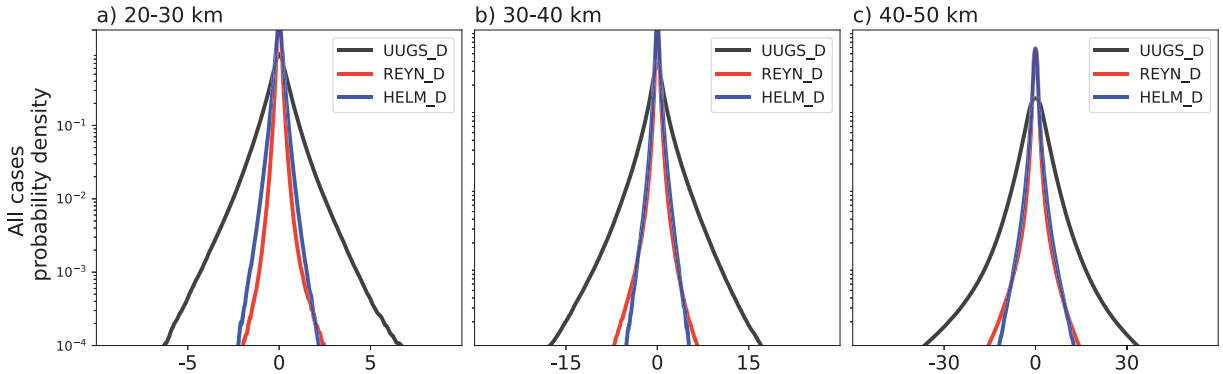


Figure S5. Same as Fig. 10 but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computinonal cost of the algorithm.