

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

Y. Qiang Sun¹, Pedram Hassanzadeh¹, M. Joan Alexander², and Christopher G Kruse³

¹Rice University

²NorthWest Research Associates, CoRA Office

³NorthWest Research Associates

December 17, 2022

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 ($\Delta x = 3 \sim 4 \text{ km}$) 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.

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

4 **Y. Qiang Sun¹, Pedram Hassanzadeh¹, M. Joan Alexander², Christopher G.**
5 **Kruse²**

6 ¹Rice University, Houston, Texas, United States

7 ²NorthWest Research Associates, Boulder, Colorado, United States

8 **Key Points:**

- 9 • In a library of WRF simulations, we compare methods for estimating 3D gravity wave
10 drag force that are un- and under-resolved by GCMs.
11 • For drag associated with vertical fluxes, different methods agree on time- and zonal-
12 mean but not on instantaneous spatiotemporal patterns.
13 • Drag associated with horizontal fluxes is significant but is very sensitive to the esti-
14 mation 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 ($\Delta x = 3 \text{ km}$) 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 $O(1 \text{ km})$ to $O(100 \text{ km})$ 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 $\sim 1 \text{ km}$ 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-

65 opment, gravity wave parameterization (GWP) is now a critical component of GCMs to
 66 enable them to reproduce realistic atmospheric circulation mean, variability, and response
 67 to climate change (e.g., McFarlane, 1987; Scinocca & McFarlane, 2000; Y. Kim et al., 2003;
 68 Beres, 2004; Alexander et al., 2010; Richter et al., 2010; Lott et al., 2012; Plougonven &
 69 Zhang, 2014). For example, generating quasi-biennial oscillation (QBO) in GCMs requires
 70 skillful GWPs (e.g., Richter et al., 2020).

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

89 Recently, Machine Learning (ML) techniques have emerged as alternative tools for
 90 developing parameterizations for climate models. They have been used in parameterizing a
 91 variety of SGS processes with promising results (e.g., Schneider et al., 2017; Rasp et al., 2018;
 92 Bolton & Zanna, 2019; Maulik et al., 2019; Chattopadhyay et al., 2020; Yuval & O’Gorman,
 93 2020; Kashinath et al., 2021; Gentine et al., 2021; Guan et al., 2022). Matsuoka et al.
 94 (2020) were among the first to apply ML to GWs. Focused on the orographic GWs over the
 95 Hokkaido region of Japan, they trained a convolutional neural network to connect the large-
 96 scale tropospheric state and the small-scale GW wind fluctuations in the lower stratosphere.
 97 Recently, Amiramjadi et al. (2022) also found success in reconstructing the non-orographic
 98 GWs in the ERA5 dataset with a random forest regressor. Both of these studies identified
 99 fluctuations associated with GWs using a simple moving-box average and demonstrated the
 100 feasibility of using ML to represent GWs. However, these studies only focused on learning
 101 GWs or momentum fluxes at one level (100 *hPa*), without further calculating the GWD,
 102 which is required to develop GWPs for GCMs. A number of other studies have also shown
 103 the power of ML for GWP through emulating current GWP schemes (Chantry et al., 2021;
 104 Espinosa et al., 2022). These emulation efforts provide valuable insight on various promises
 105 and challenges of using ML for GWPs, though a number of key challenges, e.g., related to
 106 GWD extraction and lateral GW propagation, cannot be investigated through emulation (see
 107 below).

108 One key challenge for the data-driven approach is the availability of sufficient obser-
 109 vationally constrained data of GW momentum transport for training the ML algorithms.
 110 With limited availability of observations of GWs and the challenges associated with sparsity
 111 and noise, high-resolution GW-resolving model simulations must play a critical role in gen-
 112 erating the training data. A number of case studies have verified that high-resolution models
 113 are able to capture the key characteristics of observed GWs (Bramberger et al., 2020; Kruse
 114 et al., 2022). The second key challenge in the most common data-driven approach (the
 115 so-called “supervised” or “offline” learning) is the need to extract, from the high-resolution
 116 simulations, the true GWD due to the un- and under-resolved GWs; hereafter, we refer
 117 to this collectively as the *SGS drag*. This SGS GWD is what has to be added to a low-

118 resolution GCM to properly account for the un- and under-resolved GWs (note that this
 119 depends on the GCM's *effective resolution*; more discussions to follow later). In the GW
 120 modeling community, a number of methods have been used in the past to separate GWs
 121 from the large-scale flow and quantify the SGS fluxes or GWD (e.g., G. J. Shutts & Vosper,
 122 2011; Kruse & Smith, 2015; Žagar et al., 2015; Stephan et al., 2019; Matsuoka et al., 2020;
 123 Amiramjadi et al., 2022; Polichtchouk et al., 2022). Some studies pursued a simple box-
 124 averaging method (e.g., Matsuoka et al., 2020) or a cut-off low-pass filter (e.g., Polichtchouk
 125 et al., 2022). There are also more rigorous methods to separate the balanced large-scale
 126 components and the unbalanced GWs based on linear wave theory and Helmholtz decom-
 127 position (e.g., Callies et al., 2014; Žagar et al., 2015). Stephan et al. (2019) computed the
 128 resolved GW pseudo-momentum fluxes in month-long global convection-permitting simula-
 129 tions with two other methods. These methods showed overall good agreement on the general
 130 shape of the longitudinal profiles of GW fluxes, but systematic differences were found for
 131 the amplitudes of the pseudo-momentum fluxes even after averaging over the 30-day period,
 132 implying the importance of the extraction method if we were to use these high-resolution
 133 data for training the ML algorithms.

134 The third key challenge is related to the 3D propagation of GWs and the resulting
 135 3D SGS GWD. The aforementioned studies focused only on the vertical momentum fluxes
 136 of GWs, as the current operational GWP schemes ignore lateral propagation of the waves
 137 and the resulting lateral fluxes and their contribution to the total GWD. However, there
 138 is growing evidence that horizontal propagation of GWs has to be considered in GWP
 139 to produce a realistic atmospheric circulation (e.g., Sato et al., 2009; Muraschko et al.,
 140 2015; Ehard et al., 2017). Only few studies have discussed the lateral momentum fluxes
 141 in high-resolution simulations (Eckermann et al., 2015; Jiang et al., 2019). More recently,
 142 through analyzing the lateral momentum flux in the high-resolution simulations over the
 143 Drake Passage, Kruse et al. (2022) showed that the meridionally propagating mountain
 144 waves significantly enhanced the zonal drag. Additionally, their work suggested that not
 145 accounting for these meridional fluxes would result in GWD in the wrong direction at and
 146 below the polar night jet.

147 In this paper, we use data from a library of 20 convection-permitting (3 km) tropical
 148 WRF simulations to

- 149 1. Compare 3 methods that are commonly used in the GW and large-eddy simula-
 150 tion (LES) literature to quantify the SGS fluxes and drags. These methods are i)
 151 Helmholtz decomposition, ii) Spatial filtering to compute the full SGS stress and the
 152 resulting GWD, and iii) Same as (ii) but only for the Reynolds stress.
- 153 2. Quantify the contribution of both vertical and horizontal fluxes of horizontal mo-
 154 mentum to the total GWD to investigate if the latter should be included in SGS
 155 parameterizations too.

156 Item 1 is crucial because any data-driven method, ML or otherwise, is as good as the data
 157 used for the training. Note that the challenges associated with extracting the SGS terms
 158 for ML training are not limited to the GW applications, and are in fact relevant to many
 159 climate/turbulence processes and currently an active area of research (e.g., Zhou et al., 2019;
 160 Zanna & Bolton, 2021; Grooms et al., 2021; Beck & Kurz, 2021; Guan et al., 2022).

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

166 In the rest of the paper, we will first introduce the high-resolution data and the 3 meth-
 167 ods used for SGS GWD extraction. We will then compare the results for the SGS vertical

168 momentum flux and GWD, followed by similar analyses for the SGS GWD associated with
 169 lateral momentum fluxes. Concluding remarks and discussions are in the last section.

170 2 Data and methods

171 2.1 WRF setup and data

172 All data used in this study are generated using WRF, with initial conditions from re-
 173 analysis data and boundaries nudged towards reanalysis data. For the purpose of this work,
 174 the WRF model is modified according to Kruse et al. (2022) to support a deep configura-
 175 tion that runs up to 1 *Pa* (~ 80 *km*). Two key modifications of the released WRF version
 176 4.1 model are made to achieve the high model-top here. First, low-order interpolators are
 177 used to prevent the over-/under-shooting of default higher-order interpolators, preventing
 178 the intersection of analysis levels near complex topography after horizontal interpolation.
 179 Second, the default lateral relaxation is replaced with grid-point nudging confined to the
 180 lateral boundaries for the model to run stably.

181 For now, the library only includes the tropical regions (see the domains in Fig. 1). We
 182 have conducted a total of 20 simulations in 6 domains, where the dates of the week-long
 183 runs are chosen to sample the seasonal cycle, QBO phases, and precipitation distribution
 184 (Fig. 1c). Two of these simulations, one from the 2016 summertime all-ocean West Pacific
 185 (WP), which is in the westerly phase of QBO, and one from the 2020 wintertime land-ocean
 186 Indian Ocean (IO), which is in the easterly phase of QBO, are chosen as representative cases 1
 187 and 2, respectively. The first day of all simulations is treated as spin-up periods and not used
 188 in analyses. The horizontal domain size is 3600 *km* \times 3600 *km*. The simulations are done at
 189 3 *km* grid spacing. There are 180 vertical model levels in total. The vertical grid spacing is
 190 close to 200 *m* near the lower boundary and gradually increases to a maximum of 600 *m* near
 191 the model top. For these tropical simulations, we largely use the ‘‘Tropical’’ WRF physics
 192 suite (e.g., Qiao et al., 2019), but with a different surface layer scheme. The parameterization
 193 set includes the WRF Single-Moment 6-class (WSM6) microphysics scheme (Hong & Lim,
 194 2006), the Yonsei University planetary boundary layer scheme (Hong et al., 2006), the
 195 RRTM (Rapid Radiative Transfer Model) for longwave and shortwave radiation (Iacono et
 196 al., 2008; Pincus et al., 2003), and the revised surface layer scheme developed in Jimenez
 197 et al. (2012). Note that no cumulus scheme is used given the 3 *km* grid spacing and, most
 198 importantly, no GWP is used.

199 Both the initial condition and the boundary condition come from the fifth-generation
 200 European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA5).
 201 As we are using the same method as in Kruse et al. (2022) to nudge the simulation boundaries
 202 towards the ERA5 data (hence limiting the GW signals there), the model output data near
 203 the domain boundary (< 300 *km*) are neglected when conducting analysis for the GWs.
 204 The analysis domains are hence 3000 *km* \times 3000 *km*.

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

209 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 $(\bar{\cdot})$, and b) coarse-graining, denoted with $(\tilde{\cdot})$. For any variable $\phi(\mathbf{x}, t)$, spatial filtering
 is defined as (e.g., Sagaut, 2006; Grooms et al., 2021; Guan et al., 2022)

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

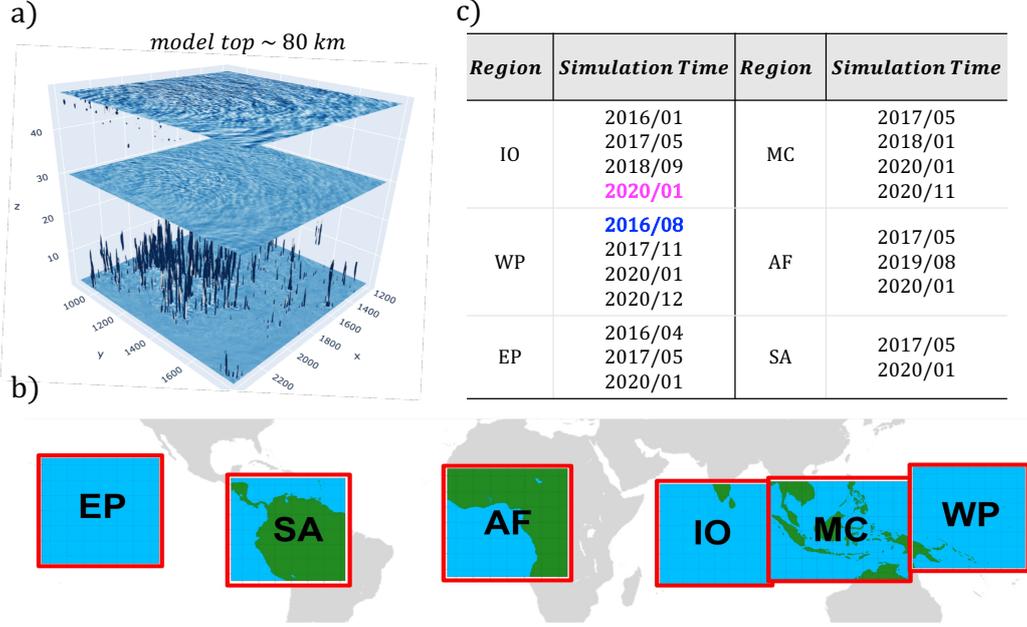


Figure 1. a) A snapshot of vertical velocity in a WRF simulation. b) Tropical regional domains where the WRF simulations are conducted. c) The list of 20 week-long WRF simulations conducted in this regional tropical channel setup, among which two representative cases are chosen (blue color represents 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 $\phi(\mathbf{x}, t)$ can be separated into two components

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

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

Two key points need to be clarified here. One is that following the convention used in recent literature (e.g., 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 $\tilde{\phi} \neq \bar{\phi}$ 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, \mathbf{r} and \mathbf{k} are coordinates in physical space and spectral space, respectively, with $\mathbf{r} = (r_x, r_y)$, and $\mathbf{k} = (k_x, k_y)$. $\hat{(\cdot)}$ is the Fourier transform, and Δ is the filter size as in Eq. (1).

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

229 the un-resolved and under-resolved GWs from the resolved GWs, or both. The choice of
 230 filter type (e.g., Gaussian, top-hat or box, sharp-spectral) can affect the extracted SGS
 231 terms, as already shown in a number of past studies including in the context of geophysical
 232 turbulence (e.g., Leonard, 1975; Zanna & Bolton, 2021; Beck & Kurz, 2021). Figure B1
 233 shows an example of the effect of filter type on the spectrum of zonal wind from our WRF
 234 simulations. Different low-pass filters (e.g., top-hat and sharp-spectral) have been used in
 235 previous studies to separate the GWs from the large-scale background (Kruse & Smith,
 236 2015; Matsuoka et al., 2020; Polichtchouk et al., 2022), though a systematic study on the
 237 effect of filter type and the potential implications for the extracted SGS terms is lacking.

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

250 To systematically quantify the effects of filter type and sizes, here, we use 3 filter
 251 types and two length scales $\Delta = 200$ km and 700 km to help with understanding the scale-
 252 awareness when building a data-driven GWP in the future (Table 1). Note that these choices
 253 of Δ are motivated by assuming that the low-resolution GCM has grid spacing of 100 km
 254 ($\sim 1^\circ$ resolution). $\Delta = 200$ km is based on the common choice for Δ in the LES literature,
 255 i.e., twice the low-resolution model’s grid spacing (Pope, 2000; Sagaut, 2006; Guan et al.,
 256 2022). $\Delta = 700$ km is based on the effective-resolution study of Skamarock (2004); this is
 257 the filter size used for the presented results, unless indicated otherwise.

258 Once resolved fluxes are quantified point-wise on the original grid, the effective fluxes
 259 within a hypothetical GCM grid cell must be computed on a coarse GCM grid. As mentioned
 260 before, we refer to this operation as coarse-graining. Admittedly, this terminology has
 261 not been uniformly adopted in the literature, though it has been recommended by several

262 recent studies (Grooms et al., 2021; Guan et al., 2022). Also, note that in some studies
 263 filtering and coarse-graining are done via just one operation, rather than two separate ones
 264 (e.g., Brenowitz & Bretherton, 2018; Yuval & O’Gorman, 2020). With all these issues in
 265 mind, here, we use one commonly used coarse-graining strategy: we simply truncate the
 266 wavenumbers greater than the cut-off wavenumber corresponding to the GCM grid spacing
 267 (100 km in this case). Note that in this study, for computational efficiency, both filtering and
 268 coarse-graining are done in the spectral (Fourier) space, and mirrored tiles are added around
 269 the original domain following Sun & Zhang (2016) to reduce problems with non-periodic
 270 boundaries.

271 To better illustrate the effects of these filtering and coarse-graining operations, Fig. 2
 272 shows examples of the high-resolution WRF snapshots, and filtered (Gaussian with $\Delta =$
 273 700 km) and coarse-grained 3D velocity fields at 30 km height. The full u, v, w in the
 274 3000 km \times 3000 km domain are shown in the left column. After the filtering operation,
 275 the velocity fields are separated into the large-scale (second column) and the perturbation
 276 (third column) components. We also apply coarse-graining operators to these fields (fourth
 277 and fifth columns) to transfer them to a 30 \times 30 grid, similar to that of a GCM with a
 278 grid spacing of 100 km. From this plot, we notice systematic differences between horizontal
 279 winds and vertical winds. For the horizontal winds u and v , the large-scale background (\tilde{u}
 280 and \tilde{v}) are much larger in amplitudes than the small-scale perturbations u' and v' , whereas
 281 for the vertical velocity, the large-scale background is almost negligible, with all the signal
 282 at small scales w' . Moreover, notice that there can be significant differences between $\tilde{\phi}$ and
 283 $\bar{\phi}$, which implies that $\tilde{\phi}' \neq 0$ with the Gaussian filtering applied here.

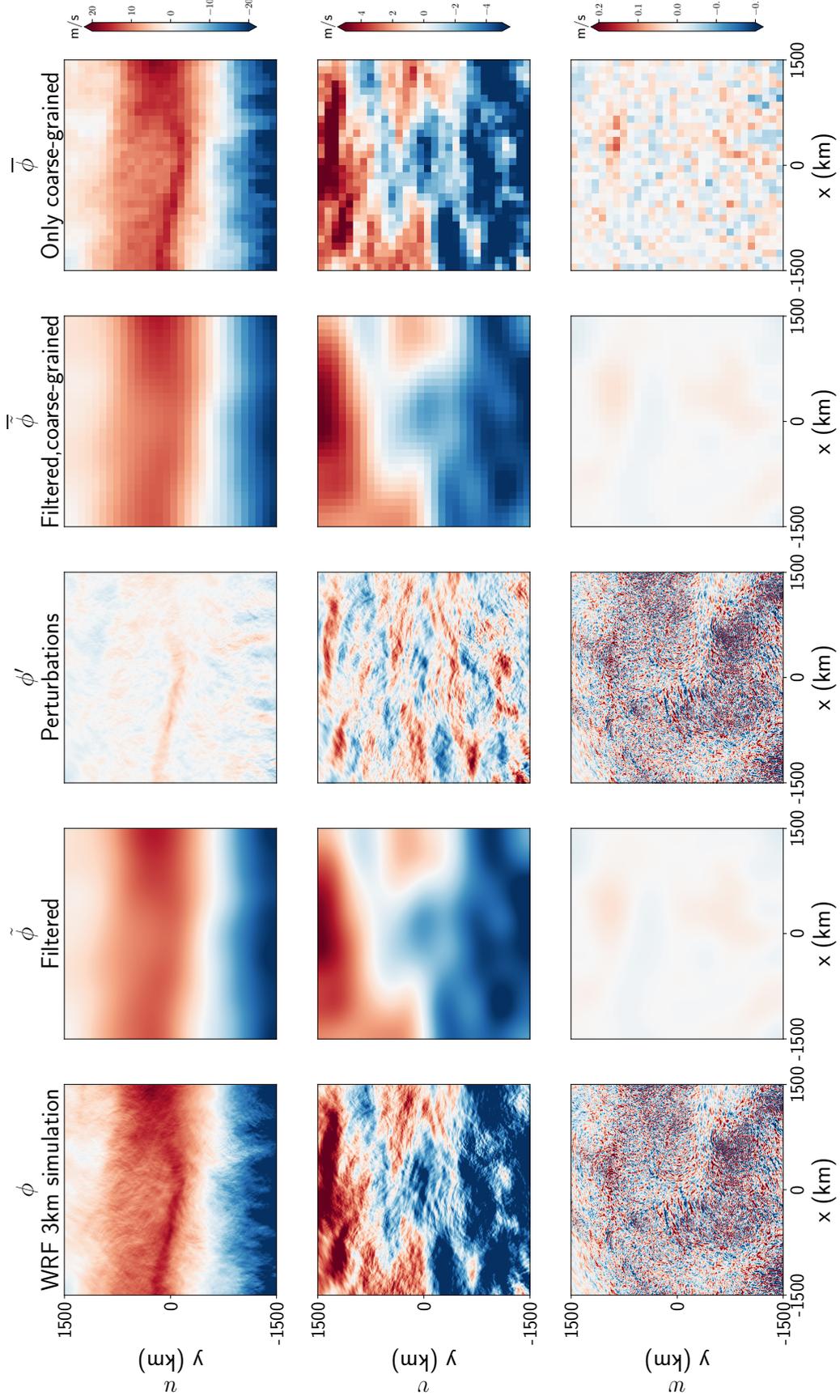


Figure 2. Examples of the effects of low-pass filtering and coarse-graining operators used in this study. The snapshots are from Case 1, on August 3rd 2016, 12:00 UTC, at 30 km height. A Gaussian filter with $\Delta = 700$ km is used. The coarse-graining is done by truncating all wavenumbers greater 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.nc1.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

333 waves and large-scale GWs could be resolved by the GCM, an additional high-pass
 334 spatial filter is applied to remove the signals that could be fully resolved by the low-
 335 resolution GCM grid. Here, we use a Gaussian high-pass filter with filter size of
 336 700 km.

337 The outcome of step (e) is the GW perturbations for the horizontal winds (u'_H and
 338 v'_H). The subscript H here denotes the use of Helmholtz decomposition in the process. The
 339 vertical winds w are not involved here in the Helmholtz decomposition of horizontal winds.
 340 Given that vertical wind w is dominated by small scales and has negligible large-scale signals
 341 (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:

$$MF_x = [MF_{xx} \quad MF_{yx} \quad MF_{zx}] \quad (3)$$

where

$$MF_{xx} = \tilde{\rho} \widetilde{u'_H u'_H} \quad (4)$$

$$MF_{yx} = \tilde{\rho} \widetilde{u'_H v'_H} \quad (5)$$

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

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

346 The first two components in Eq. (3) are the zonal and meridional flux of zonal mo-
 347 mentum due to SGS GWs, respectively. They will also be referred to as lateral momentum
 348 fluxes. The last component in Eq. (3) is the vertical flux of zonal momentum due to SGS
 349 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 ,

$$\overline{MF}_x = [\overline{MF}_{xx} \quad \overline{MF}_{yx} \quad \overline{MF}_{zx}]. \quad (7)$$

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 :

$$\overline{GWD}_{xx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{xx}}{\partial x} \quad (8)$$

$$\overline{GWD}_{yx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{yx}}{\partial y} \quad (9)$$

$$\overline{GWD}_{zx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{zx}}{\partial z}. \quad (10)$$

350 Note that Eq. (10), the vertical divergence of the vertical flux of zonal momentum due to
 351 SGS GWs, is often considered to be the dominant component in previous studies, and the
 352 only term that is conventionally represented in existing SGS parameterizations. This has
 353 been the case in the development of physics-based GWP, and in the past efforts focused
 354 on extracting SGS GWD from high-resolution simulations (e.g., Alexander et al., 2010;
 355 Matsuoka et al., 2020; Polichtchouk et al., 2022). Yet, as shown in Kruse et al. (2022), this
 356 is not always the case, and the lateral divergence of lateral momentum fluxes (Eqs. (8) and
 357 (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

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

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 $\overline{u'w'}$ is a part of the (but not the entire) total SGS vertical flux, $\overline{u\tilde{w}} - \tilde{u}\tilde{w}$. The other two components (e.g., $\overline{u\tilde{v}'}$) arise because as mentioned before, in spatial filtering and coarse-graining, terms like \tilde{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 $\overline{u'u'}$ and $\overline{u'v'}$ as well as other stresses, including $\overline{u\tilde{u}'}$ and $\overline{v\tilde{u}'}$. 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

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

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

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

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.

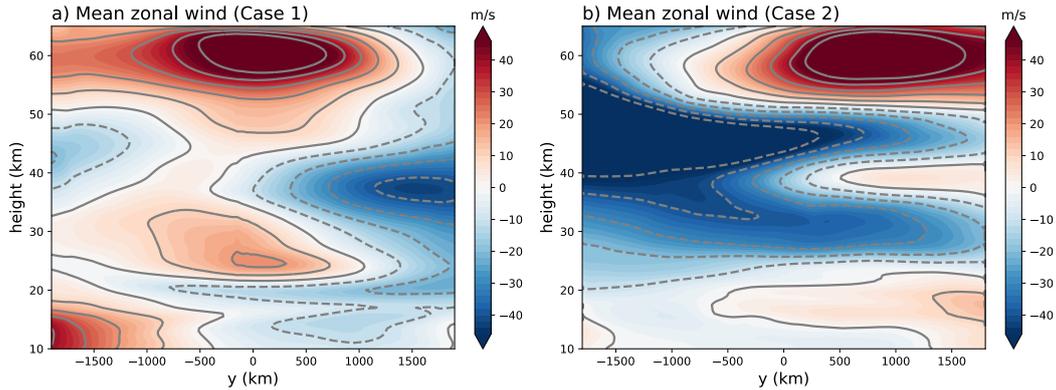


Figure 3. Zonal-mean, time-mean zonal wind u for two representative WRF simulation cases. The contour interval is 10 m , with dash lines implying zero and negative values. a) Case 1 with westerly QBO winds; b) Case 2 with easterly QBO winds .

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 $\sim 20\text{ m/s}$ in the tropical stratosphere ($\sim 25\text{ km}$). The semiannual oscillation (SAO) can also be seen near the stratopause ($\sim 60\text{ km}$), 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 ($\Delta = 700\text{ km}$) 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

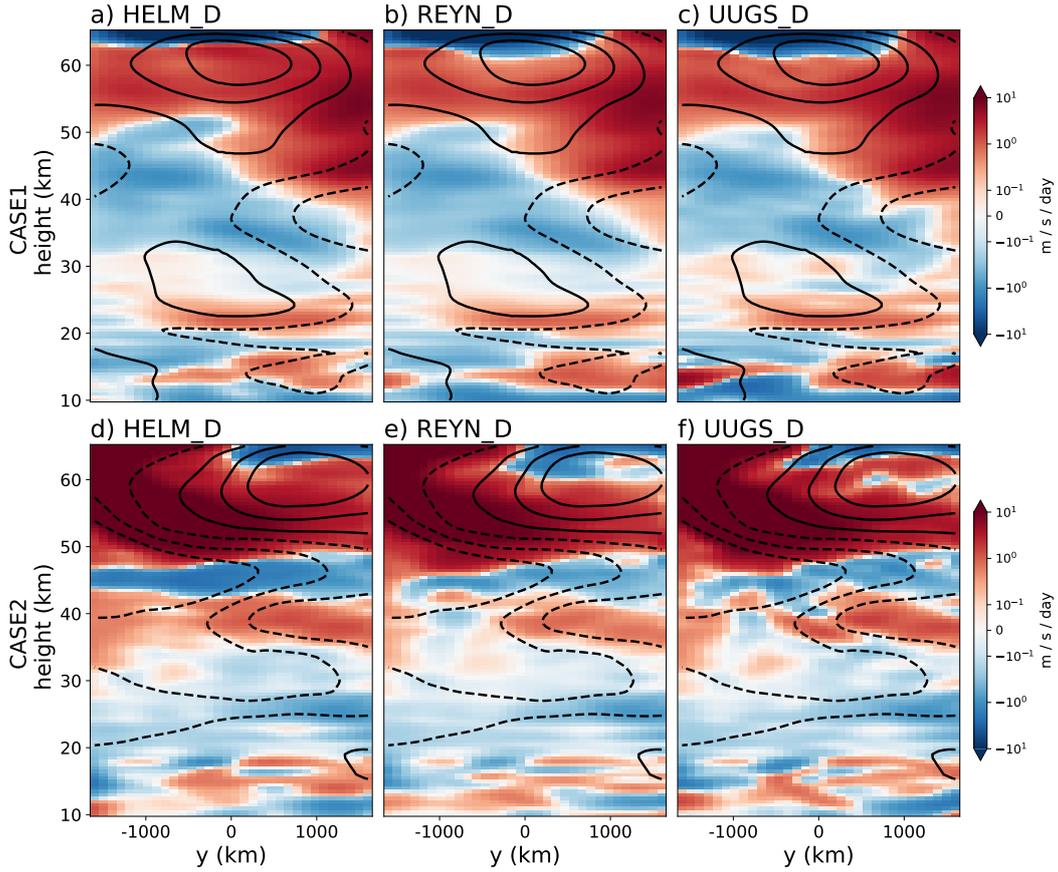


Figure 4. Zonal-mean, time-mean zonal SGS GWD (shading) due to the un-/under-resolved (SGS) vertical fluxes derived using three methods for both representative cases (upper row: Case 1, lower row: Case 2). Left column, a & d: Helmholtz decomposition method (HELM_D). Middle column, b & e: Reynolds stress 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 .

422 GWs help maintain and drive the zonal wind there (Alexander et al., 2010). Moreover, for
 423 the QBO region, the maximum drag is below the wind maximum (e.g., Case 1 in Fig. 4),
 424 implying the role of SGS GWs in the downward propagation of the zonal winds.

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

437 Figure 6 shows the probability density functions (PDFs) for the zonal SGS GWD as-
 438 sociated with vertical fluxes using these three methods, as another way of presenting the

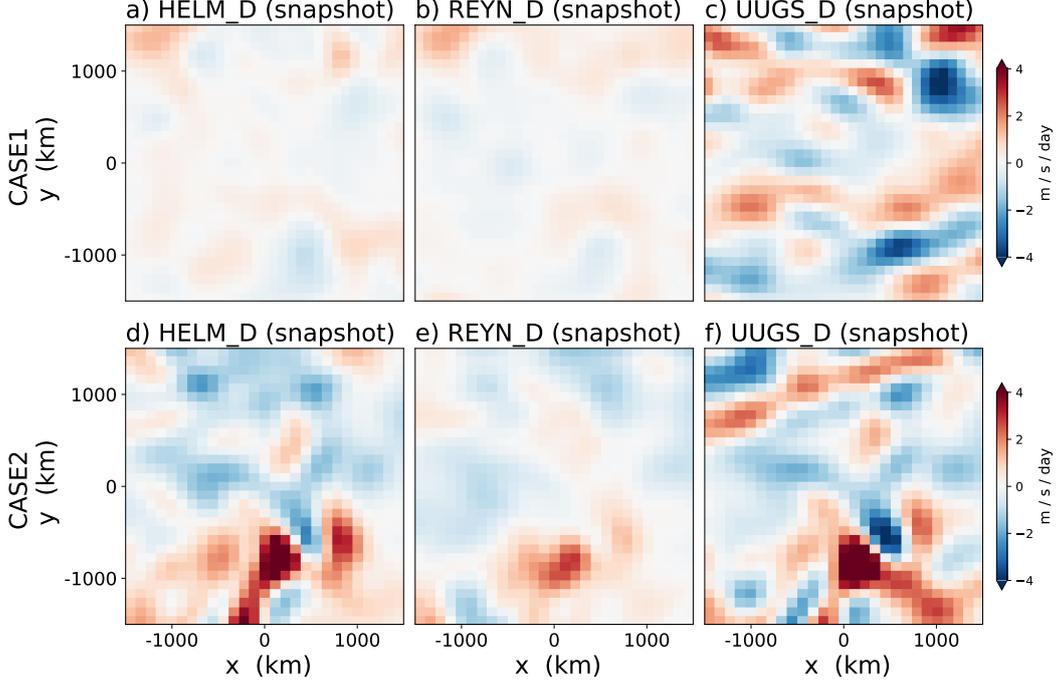


Figure 5. Snapshots of zonal SGS GWD due to vertical fluxes at 30 km height calculated using three methods. Upper row: Case 1. Lower row: Case 2. Left column, a & d: Helmholtz decomposition method $\frac{1}{\bar{\rho}} \frac{\partial \bar{\rho}'_H w'}{\partial z}$. Middle column, b & e: Reynolds stress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

439 differences among them. While the PDFs we obtain using the HELM_D and REYN_D
 440 methods are fairly similar, the PDFs from the UUGS method often have a clearly wider
 441 distribution. For both cases studied here, we find that the GWD from the UUGS method
 442 in general has higher variability, both temporal and spatial, compared with the other meth-
 443 ods, that are based on the Reynolds stress term alone (e.g., Eq. (14)), which is one of the
 444 three components of the total stress estimated in the UUGS method (see Appendix A). The
 445 degree of differences in PDFs depend on the case and height. For example, the differences
 446 are smaller at 40-50 km for Case 1, possibly due to the weak zonal winds there (Fig. 4).

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

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

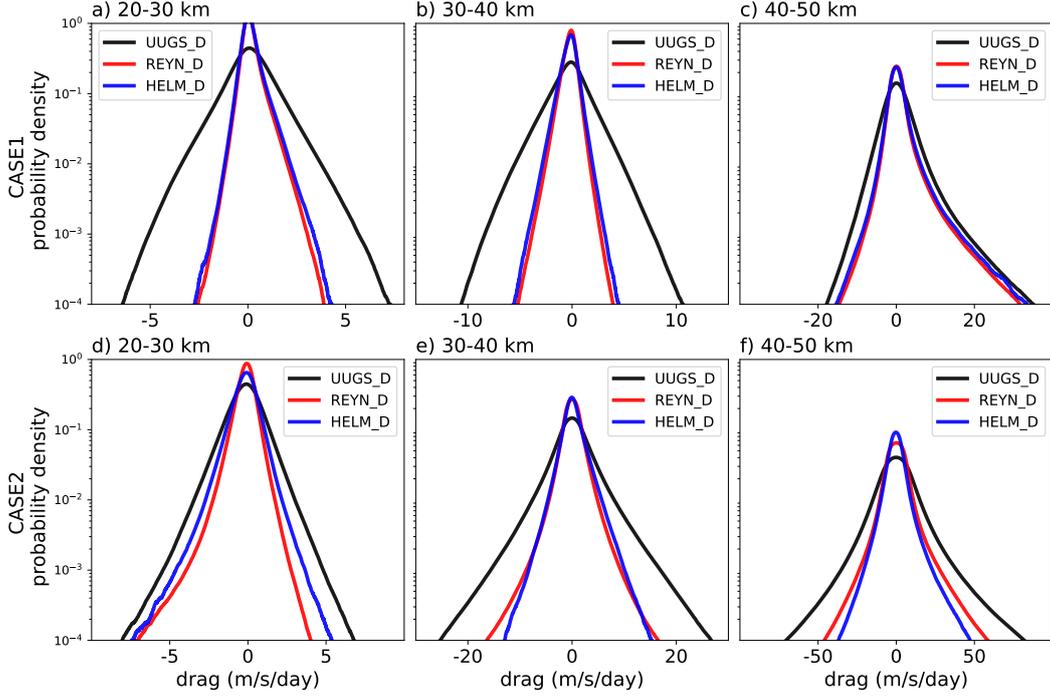


Figure 6. Probability density function (PDF) of zonal SGS GWD due to vertical fluxes calculated using three methods at different heights. Upper row: Case 1. Lower row: Case 2. Left column, a & d: 20-30 *km*. Middle column, b & e: 30-40 *km*. Right column, c & f: 40-50 *km*. Note the differences in the *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.

461

3.3 Horizontal flux of zonal momentum due to SGS GWs

462

463

464

465

466

467

468

469

470

471

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.

472

473

474

475

476

477

478

479

480

481

482

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

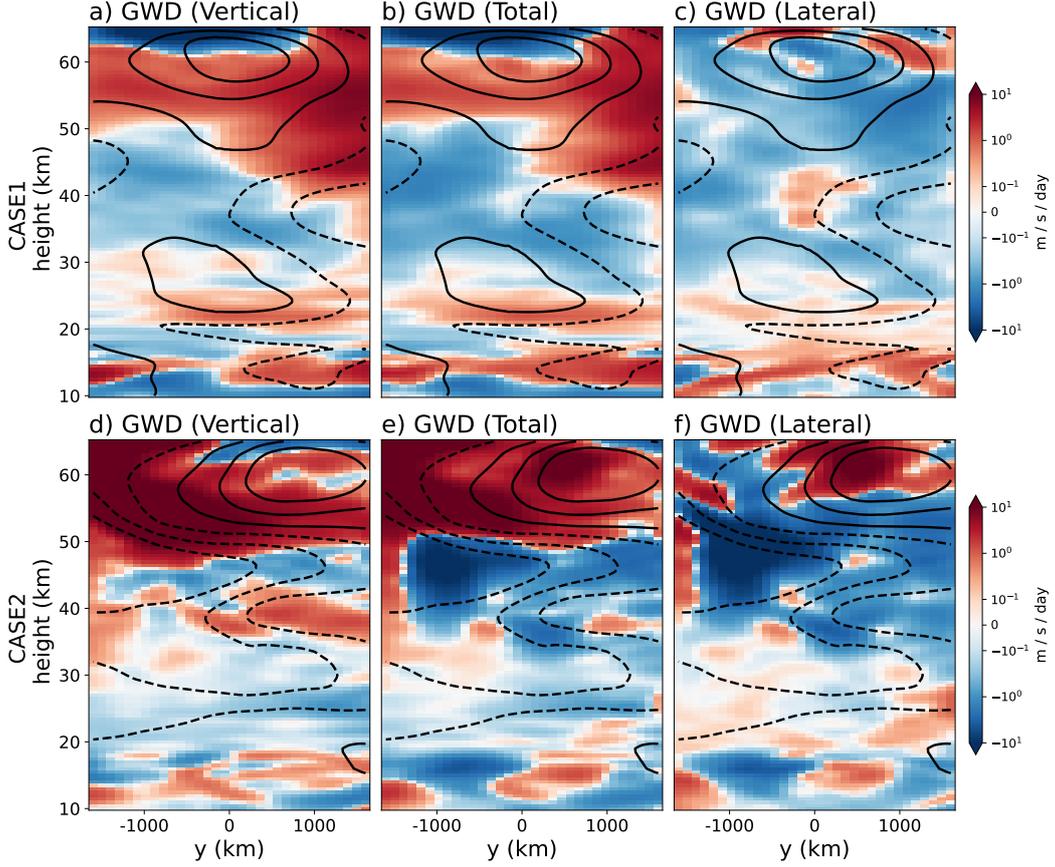


Figure 7. Zonal-mean, time-mean zonal SGS GWD due to only the vertical fluxes, total fluxes, and the lateral fluxes calculated using the UUGS_D method (Eq. (11)). Left column: Only the vertical component $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}]$. Middle column: All terms $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial x} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial y} [\bar{\rho}(\overline{uv}) - \bar{u}\bar{v}] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}]$. 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.

483 another example, in Case 2, at levels below the SAO (~ 50 km), it is clear that the lateral
 484 lateral momentum fluxes dominate the GWD there, even leading to a change of direction of the
 485 total zonal SGS GWD.

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

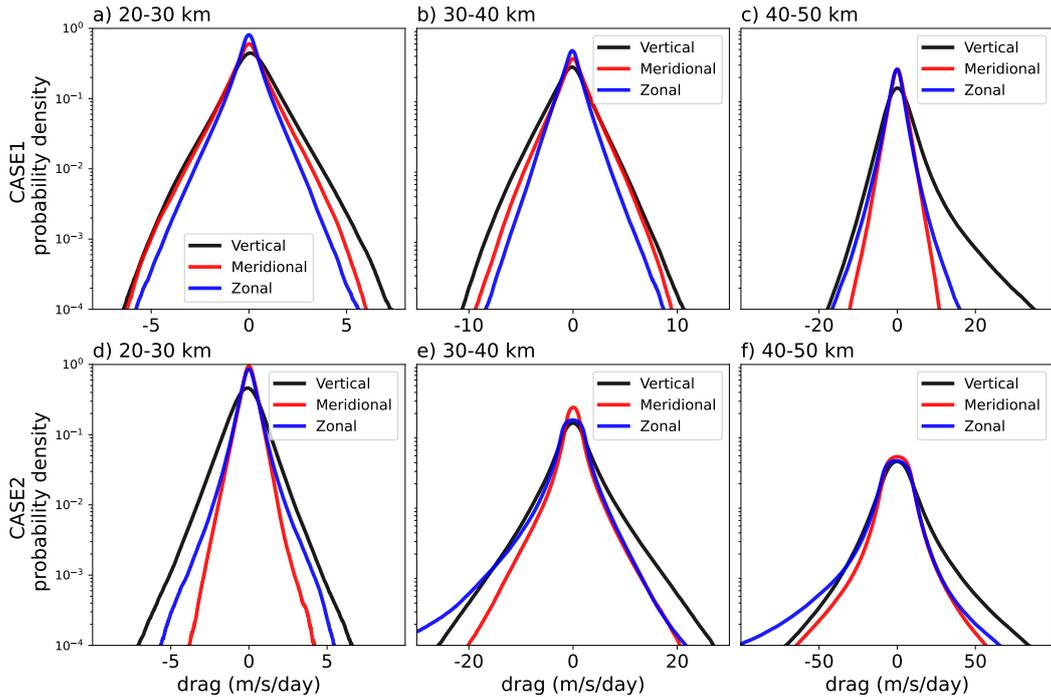


Figure 8. Probability density function (PDF) of zonal SGS GWD due to SGS zonal (blue), meridional (red), and vertical (black) momentum fluxes for both cases. Upper row: Case 1; Lower panel: Case 2. Left column, a & d: 20-30 km. Middle column, b & e: 30-40 km. Right column, c & f: 40-50 km. Note the difference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

497 So far, we have used the UUGS_D method to calculate the GWD due to SGS horizontal
 498 fluxes (Figs. 7-8). Compared to the vertical fluxes shown earlier, calculations of SGS GWD
 499 due to horizontal momentum fluxes have a much stronger sensitivity to the choice of the
 500 method. Figure 9 shows the time-mean, zonal mean SGS GWD associated with the meridional
 501 fluxes of the zonal momentum, calculated using HELM_D, REYN_D, and UUGS_D
 502 (the second term in Eq. (11)), respectively. Different from Fig. 4, the results here strongly
 503 depend on the method, even after averaging over time (simulation period) and space (zonal
 504 direction). This suggests that if we want to include the lateral propagation of GWs in the
 505 data-driven GWP schemes, then we must carefully examine the GWD extraction methodol-
 506 ogy. The PDFs in Fig. 10 show the same story. The SGS GWD induced by the lateral fluxes
 507 are much larger if calculated using the UUGS method compared to the other two (note the
 508 logarithmic color bar). It is clear that drag due to Reynolds stress is not the dominant term
 509 anymore when we consider GWD due to the SGS lateral fluxes. One explanation for this is
 510 that there are fundamental differences between the scales and amplitudes of the horizontal
 511 winds (u, v) and the vertical winds (w), as already shown in Fig. 2. The vertical velocity
 512 is dominated by small-scale features with negligible signal at the resolved scales in GCMs,
 513 which results in weak interactions between the resolved scales and the small scales. On the
 514 contrary, the horizontal winds are dominated by winds at the resolved scales, which means
 515 much stronger interactions between the resolved scales and the unresolved scales, and hence
 516 the large differences between UUGS_D and REYN_D.

517 Moreover, while Figs. 9 and 10 suggest similarities between the SGS GWD associated
 518 with the SGS horizontal fluxes calculated using the HELM_D and the REYN_D methods
 519 (Fig. 9a and 9b), substantial differences can exist even between the SGS GWD patterns
 520 extracted using these two methods. Figure 11 shows the correlation between instantaneous

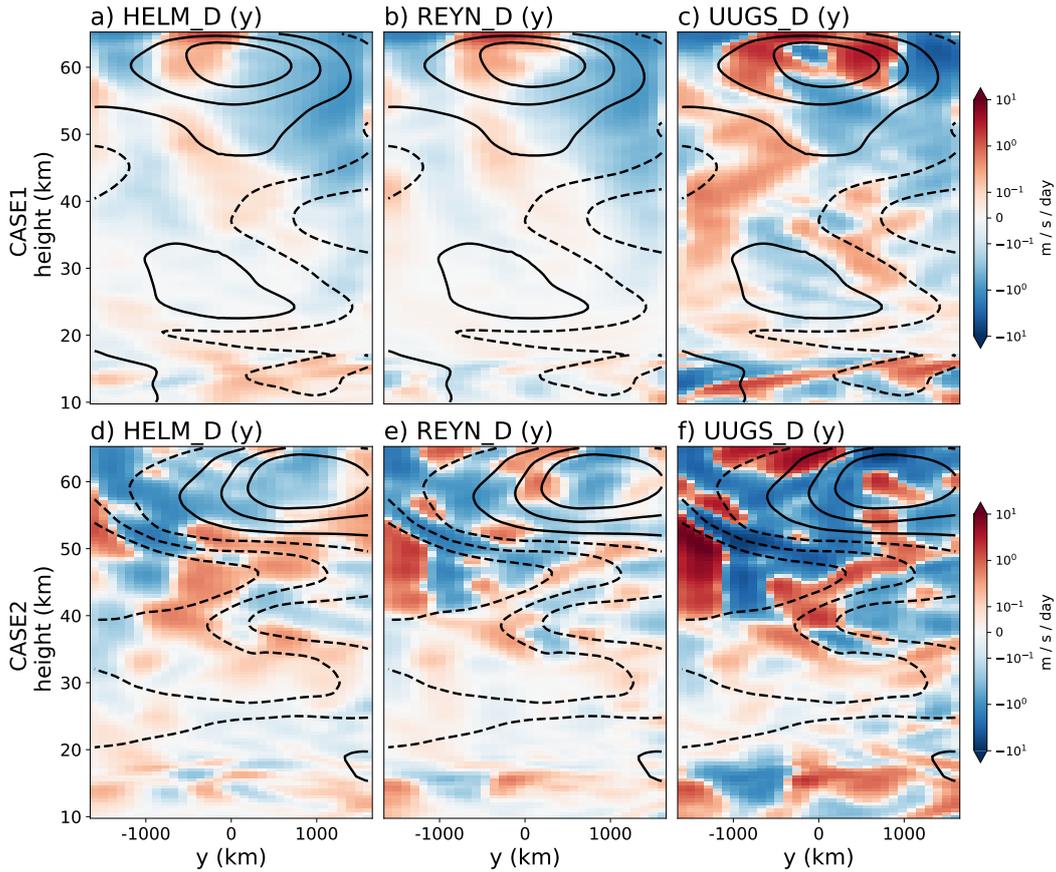


Figure 9. Zonal-mean, time-mean zonal SGS GWD, similar to Fig. 4, but for the GWD due to the un-/under-resolved (SGS) horizontal fluxes. Only the meridional direction is shown here; examining the zonal direction shows a similar story.

521 GWD calculated using the HELM_D and REYN_D method. For the SGS GWD due to
 522 vertical fluxes, as already discussed, there is a good match between these two methods.
 523 However, for SGS GWD due to the horizontal fluxes, the correlation is fairly weak, even
 524 though mathematically similar expressions are used for GWD in both methods. These
 525 results, again, show the high sensitivity of the lateral momentum fluxes and the resulting
 526 GWD to the details of the extraction method.

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

530 3.4 Sensitivity to the filter type/size and the GCM resolution

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

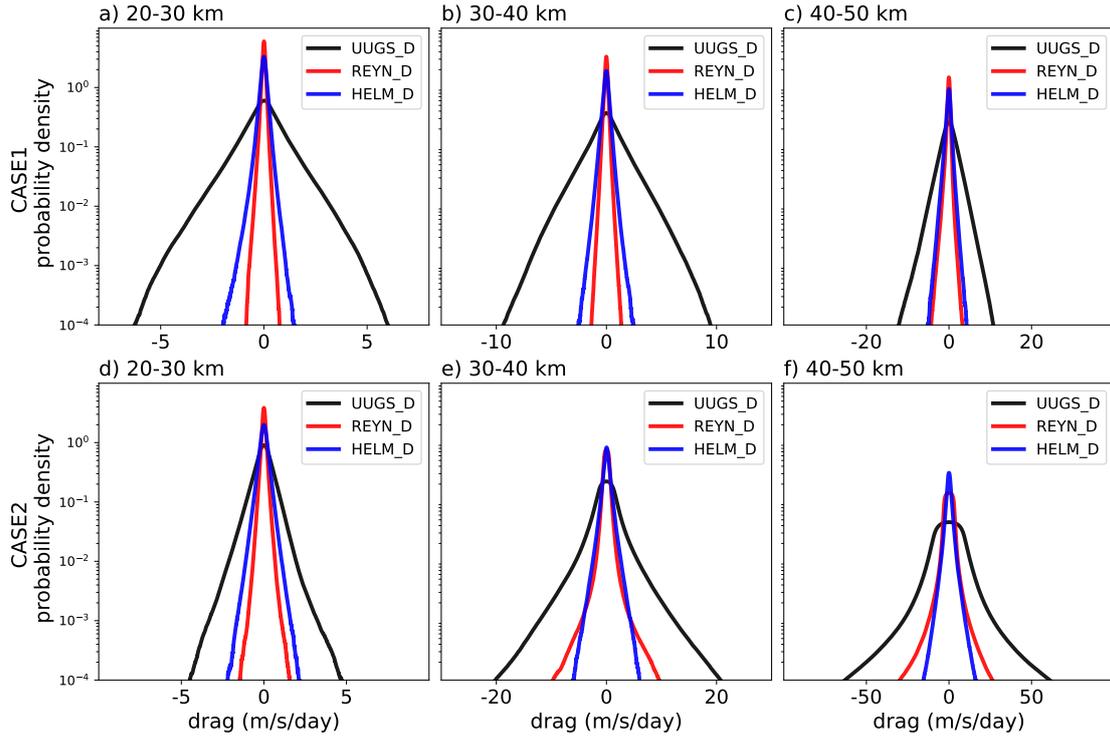


Figure 10. Same as Fig. 6, but for the un- and under-resolved zonal GWD due to SGS meridional momentum fluxes. Note the difference in the x -axes. See Fig. S4 for the PDFs of zonal GWD due to the SGS zonal momentum fluxes. See Fig. S5 for the PDFs of data from all 20 cases combined.

539 As mentioned earlier, Fig. B1 shows the power spectrum of the zonal winds before and
 540 after these low-pass filters are applied. This figure demonstrates the overall similarities
 541 between the outcome of the Gaussian and top-hat filters, at least up to the filtering scale, and
 542 major differences with the outcome of the sharp-spectral filter. Figure B2 shows snapshots
 543 of the SGS vertical momentum flux (Reynolds stress and total stress) extracted using these
 544 3 filters and $\Delta = 700 \text{ km}$. Again, we see that the Gaussian and top-hat filters overall
 545 yield fairly similar results. The outcomes of the sharp-spectral filter on the other hand,
 546 show differences in both amplitude and pattern, though the degree of difference is more
 547 pronounced for the Reynolds stress.

548 All the results shown so far are with filter size $\Delta = 700 \text{ km}$, coarse-grained to the GCM
 549 resolution of 100 km . However, this choice of 700 km is rather subjective, as there is no well-
 550 defined physical scale separation for GWs. Moreover, the appropriate filtering scale depends
 551 on the capability of a given GCM to resolve the GWs larger than the GCM's grid spacing,
 552 i.e., it depends on "effective resolution" of the GCM, which in turn depends on the GCM's
 553 numerical schemes and choices of grid-scale filters, like hyperdiffusion (e.g., Klaver et al.,
 554 2020). The $\Delta = 700 \text{ km}$ used here is based on studies showing that the effective resolution
 555 of WRF for GWs is 7 times the grid spacing (Skamarock, 2004). Admittedly, Δ should be
 556 chosen based on the effective resolution of the target GCM, not that of the GW-resolving
 557 model. To examine the sensitivity of the results to this choice, below we also present analysis
 558 with filtering scale that is twice the GCM grid spacing (i.e., $\Delta = 200 \text{ km}$), which is based
 559 on the LES literature (Pope, 2000; Sagaut, 2006; Guan et al., 2022). Furthermore, with the
 560 increase in computing power, some GCMs now have grid spacing of 0.5° or even smaller.
 561 Therefore, below, we also show results with for a GCM with the grid spacing of 30 km .

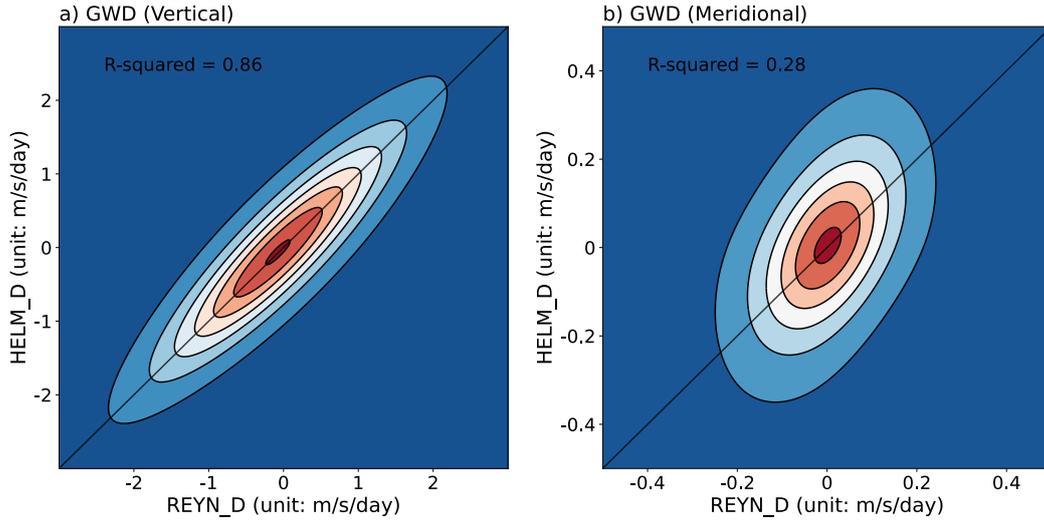


Figure 11. Joint PDFs between the SGS GWD calculated using the HELM.D and REYN.D methods for Case 1. a) GWD due to SGS vertical fluxes and b) GWD due to SGS meridional fluxes. The R-squared value is shown to measure the correlation between these two methods.

562 Figure 12 shows PDFs of the GWD from the SGS zonal, meridional, and vertical
 563 fluxes for 3 sets of choices: GCM grid spacing of 100 km and $\Delta = 700$ km (left column)
 564 and $\Delta = 200$ km (middle column), and GCM grid spacing of 30 km and $\Delta = 200$ km
 565 (right column). The second choice is meant to show the influences of effective resolution
 566 change while the third choice is meant to show what happens with higher-resolution GCMs.
 567 Although one might expect smaller SGS GW wind perturbations with reduced Δ , the zonal
 568 SGS GWD may not be necessarily reduced, as less averaging of the momentum flux is also
 569 applied with a smaller Δ . As a result, in both cases, for the zonal SGS GWD associated
 570 with vertical fluxes (solid lines), we find larger values when the filter length scale is reduced
 571 (compare the tails of the PDFs in the left and middle columns). The zonal SGS GWD
 572 associated with horizontal fluxes may become larger or smaller depending on the case.

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

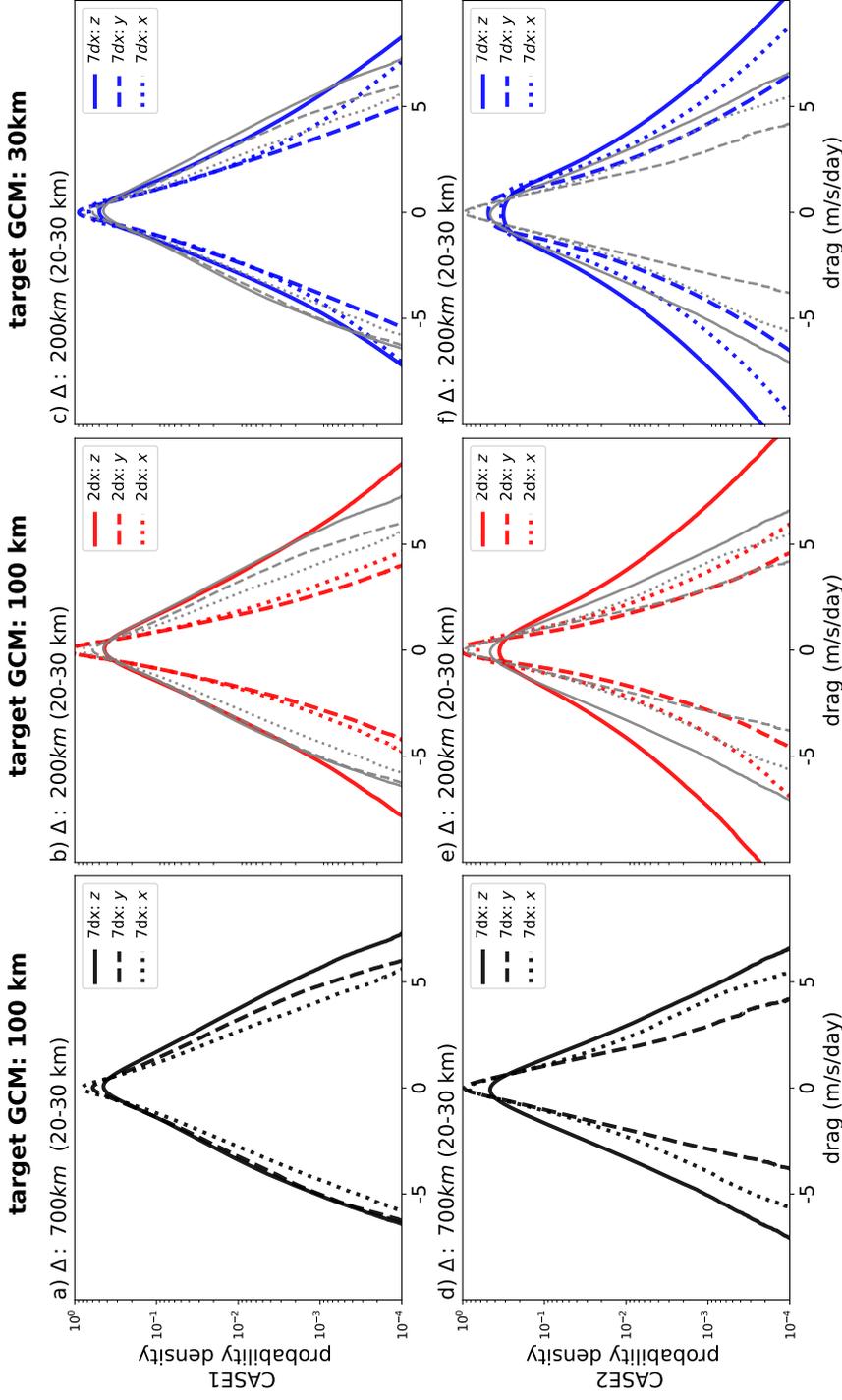


Figure 12. The PDFs of GWD at 20 – 30 km heights associated with the SGS zonal (dotted), meridional (dashed), and vertical (solid) fluxes extracted using the UUGS_D method (with Gaussian filter) for 3 different sets of filter size Δ and GCM grid spacing. Left column: $\Delta = 700$ km, 100 km grid spacing. 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 complexity, we have found that the GWD due to SGS lateral momentum fluxes could be sensitive to the methods used to extract them. Even the time-mean, zonal-mean GWD could be very different when different methods are used. The instantaneous GWD amplitudes from the UUGS_D method could be much

larger than those from the REYN_D method due to the strong cross-scale interactions between SGS GWs and the horizontal background flow that could be resolved by the GCMs. This suggests that to include the SGS GWD due to the horizontal fluxes in the data-driven GWP schemes, further research needs to be done on the extraction methodology.

4. The sensitivity of extracted SGS GWD to the length scale (Δ) of the filtering operation and the horizontal resolution of the GCM are also studied. Our results suggest that both the effective resolution of a GCM and its grid spacing have significant influences on the calculated SGS GWD (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

682 the GWs in the tropics (Jeevanjee, 2017; Bramberger et al., 2020). We have conducted
 683 limited simulations with 1 km and 3 km grid spacing at periods when observations from
 684 super-pressure balloons are available. We will use these simulations to examine the effect of
 685 resolution and validate the GW fluxes in the library. Also, currently our WRF simulations
 686 are limited to the tropics. In the future, we aim to extend the library to the mid-latitude
 687 and even polar regions. We also plan to conduct pseudo-global warming experiments to
 688 examine the response of the GWD to climate change and to expand the library to include
 689 training sets for transfer learning, such that generalizable data-driven GWP schemes could
 690 be developed (Guan et al., 2022; Subel et al., 2022).

691 Appendix A UUGS Drag Extraction

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

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

695 where (u, v, w) is the 3D wind fields; p is pressure; ρ is density; f is the Coriolis parameter;
 696 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:

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

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

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

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

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

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

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

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

To get an expression for \mathbf{X}_x^G , we can apply both spatial filtering and coarse-graining operators to each term in the original Eq. (A1), and assume that the operations are commutative ¹, which means, e.g.,

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

¹This 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

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

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

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

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

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

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

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

Similarly, for the meridional momentum equation, we obtain

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

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

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

700 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.,

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

701 We see that the Reynolds stress is one of the three components of the total SGS flux. The
 702 total SGS flux accounts for interactions among all scales, including scales resolved by the
 703 GCMs with the un- and under-resolved scales. The importance of Leonard term and cross
 704 term has long been shown in studies of turbulent flows (e.g., Leonard, 1975; Galmarini et al.,
 705 2000). We also note that the Reynolds term here based on spatial filtering is different with
 706 the traditionally temporal-based Reynolds average in which the flow is decomposed into a
 707 mean and fluctuating components. As pointed in Aluie et al. (2018), the time-mean flow is
 708 not synonymous with large-scale flow, nor does a temporal fluctuation directly correspond
 709 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).

710 **Appendix B The effects of filter type**

711 Figures B1 and B2 show the spectra and snapshots of zonal wind and SGS vertical flux
712 when the 3 different filters are applied with the same Δ . Note that because our WRF regional
713 domain is not periodic, we have used mirrored tiles to reduce the boundary effects. Still,
714 Fourier-based filters such as the sharp-spectral filter might suffer from Gibbs oscillations
715 and give non-physical results.

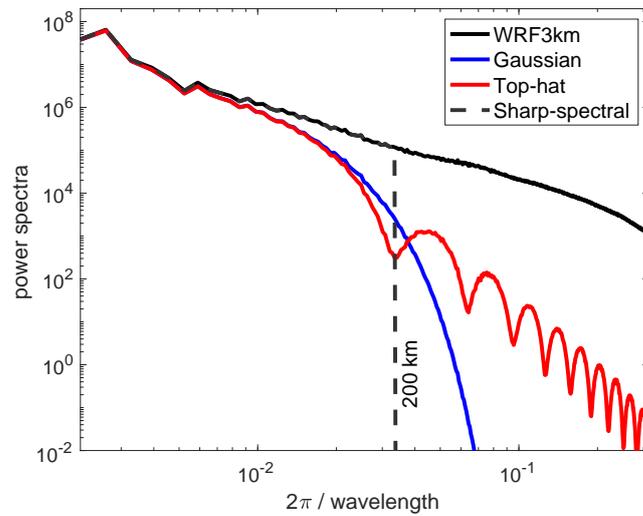


Figure B1. Example of the power spectrum of the zonal wind from the 3-*km* WRF simulations before and after the low-pass filtering using different kernels (with length scale $\Delta = 200$ *km*). The black line shows the spectrum before filtering, while the blue, red, and green lines show the the spectrum 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).

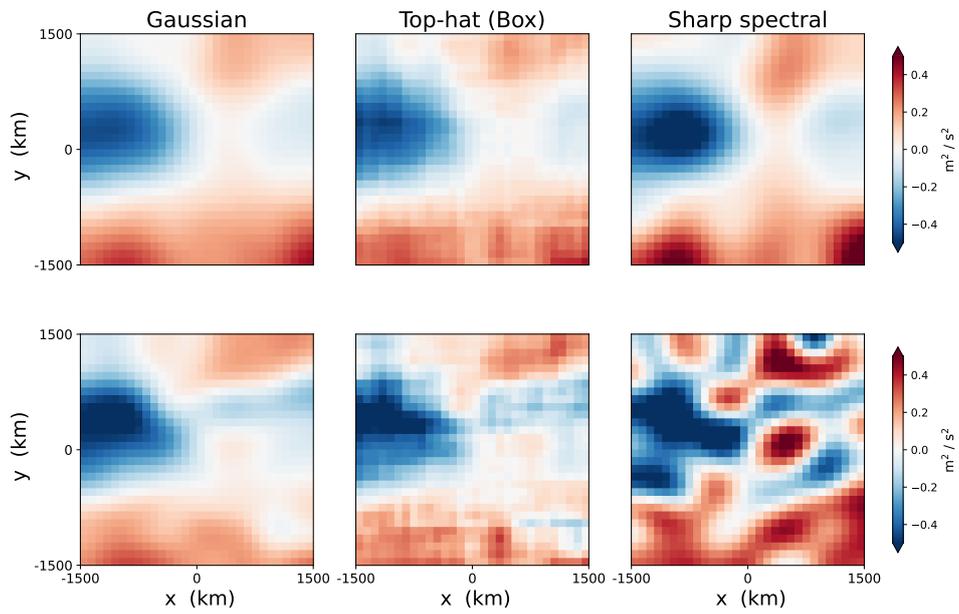


Figure B2. Snapshots of the SGS vertical momentum fluxes, calculated using the 3 different filters for Case 1 (at 40 km height). Top row: the Reynolds stress, $\overline{u'w'}$. Bottom: the total SGS stress, $\overline{uw} - \overline{u}\overline{w}$. The filter size is $\Delta = 700$ km.

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 CISE (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 climate models and the global distribution of gravity-wave momentum 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
- Amiranjadi, M., Plougonven, R., Mohebalhojeh, A. R., & Mirzaei, M. (2022). Using machine learning to estimate non-orographic gravity wave characteristics at source levels. *Journal of the Atmospheric Sciences*. 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
- Böloni, 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

- 766 Bölöni, G., Ribstein, B., Muraschko, J., Sgoff, C., Wei, J., & Achatz, U. (2016). The inter-
767 action between atmospheric gravity waves and large-scale flows: An efficient description
768 beyond the nonacceleration paradigm. *Journal of the Atmospheric Sciences*, *73*. doi:
769 10.1175/JAS-D-16-0069.1
- 770 Callies, J., Ferrari, R., & Bühler, O. (2014). Transition from geostrophic turbulence to
771 inertia-gravity waves in the atmospheric energy spectrum. *Proceedings of the National
772 Academy of Sciences of the United States of America*, *111*. doi: 10.1073/pnas.1410772111
- 773 Cao, J., Ran, L., & Li, N. (2014). An application of the helmholtz theorem in extracting
774 the externally induced deformation field from the total wind field in a limited domain.
775 *Monthly Weather Review*, *142*. doi: 10.1175/MWR-D-13-00311.1
- 776 Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Ma-
777 chine learning emulation of gravity wave drag in numerical weather forecasting. *Jour-
778 nal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002477. Retrieved
779 from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002477>
780 (e2021MS002477 2021MS002477) doi: <https://doi.org/10.1029/2021MS002477>
- 781 Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven super-
782 parameterization using deep learning: Experimentation with multiscale Lorenz 96 sys-
783 tems and transfer learning. *Journal of Advances in Modeling Earth Systems*, *12*(11),
784 e2020MS002084.
- 785 Chen, T.-C., & Wiin-Nielsen, A. C. (1976). On the kinetic energy of the divergent and
786 nondivergent flow in the atmosphere. *Tellus*, *28*. doi: 10.3402/tellusa.v28i6.11317
- 787 Clark, R. A., Ferziger, J. H., & Reynolds, W. C. (1979). Evaluation of subgrid-scale models
788 using an accurately simulated turbulent flow. *Journal of fluid mechanics*, *91*(1), 1–16.
- 789 Corcos, M., Hertzog, A., Plougonven, R., & Podglajen, A. (2021). Observation of gravity
790 waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical
791 Research: Atmospheres*, *126*. doi: 10.1029/2021JD035165
- 792 Dunbar, O. R., Garbuno-Inigo, A., Schneider, T., & Stuart, A. M. (2021). Calibration
793 and uncertainty quantification of convective parameters in an idealized gcm. *Journal of
794 Advances in Modeling Earth Systems*, *13*(9), e2020MS002454.
- 795 Eckermann, S. D., Ma, J., & Broutman, D. (2015). Effects of horizontal geometrical
796 spreading on the parameterization of orographic gravity wave drag. part i: Numer-
797 ical transform solutions. *Journal of the Atmospheric Sciences*, *72*(6), 2330 - 2347.
798 Retrieved from [https://journals.ametsoc.org/view/journals/atsc/72/6/jas-d-14-
799 -0147.1.xml](https://journals.ametsoc.org/view/journals/atsc/72/6/jas-d-14-0147.1.xml) doi: 10.1175/JAS-D-14-0147.1
- 800 Ehard, B., Kaifler, B., Dörnbrack, A., Preusse, P., Eckermann, S. D., Bramberger, M., ...
801 Rapp, M. (2017). Horizontal propagation of large-amplitude mountain waves into the
802 polar night jet. *Journal of Geophysical Research*, *122*. doi: 10.1002/2016JD025621
- 803 Espinosa, Z. I., Sheshadri, A., Cain, G. R., Gerber, E. P., & DallaSanta, K. J. (2022).
804 Machine Learning Gravity Wave Parameterization Generalizes to Capture the QBO and
805 Response to Increased CO₂. , *49*(8), e98174. doi: 10.1029/2022GL098174
- 806 Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., ... Williams,
807 D. N. (2016). Towards improved and more routine Earth system model evaluation in
808 CMIP. *Earth System Dynamics*, *7*(4), 813-830. doi: 10.5194/esd-7-813-2016
- 809 Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle
810 atmosphere. *Review of Geophysics*, *41*(1). doi: 10.1029/2001RG000106
- 811 Fritts, D. C., & Nastrom, G. D. (1992). Sources of Mesoscale Variability of Gravity Waves.
812 Part II: Frontal, Convective, and Jet Stream Excitation. *Journal of Atmospheric Sciences*,
813 *49*(2), 111-127. doi: 10.1175/1520-0469(1992)049<0111:SOMVOG>2.0.CO;2
- 814 Galmarini, S., Michelutti, F., & Thunis, P. (2000). Estimating the contribution of leonard
815 and cross terms to the subfilter scale from atmospheric measurements. *Journal of the
816 Atmospheric Sciences*, *57*, 2968-2976. doi: 10.1175/1520-0469(2000)057<2968:ETCOLA>
817 2.0.CO;2
- 818 Gentine, P., Eyring, V., & Beucler, T. (2021). Deep learning for the parametrization of sub-
819 grid processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive*

- 820 *Approach to Remote Sensing, Climate Science, and Geosciences*, 307–314.
- 821 Germano, M. (1986). A proposal for a redefinition of the turbulent stresses in the filtered
822 Navier-Stokes equations. *The Physics of fluids*, 29(7), 2323–2324.
- 823 Germano, M. (1992). Turbulence: the filtering approach. *Journal of Fluid Mechanics*, 238,
824 325–336.
- 825 Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R.,
826 ... Randel, W. J. (2019). The whole atmosphere community climate model version
827 6 (waccm6). *Journal of Geophysical Research: Atmospheres*, 124, 12380-12403. doi:
828 10.1029/2019JD030943
- 829 Grooms, I., Loose, N., Abernathy, R., Steinberg, J., Bachman, S. D., Marques, G., ...
830 Yankovsky, E. (2021). Diffusion-based smoothers for spatial filtering of gridded geophys-
831 ical data. *Journal of Advances in Modeling Earth Systems*, 13(9), e2021MS002552.
- 832 Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022). Stable a posteriori
833 LES of 2D turbulence using convolutional neural networks: Backscattering analysis and
834 generalization to higher Re via transfer learning. *Journal of Computational Physics*, 458,
835 111090.
- 836 Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2023). Learning physics-
837 constrained subgrid-scale closures in the small-data regime for stable and accurate les.
838 *Physica D: Nonlinear Phenomena*, 443, 133568. doi: [https://doi.org/10.1016/j.physd](https://doi.org/10.1016/j.physd.2022.133568)
839 .2022.133568
- 840 Hong, S.-Y., & Lim, J.-O. J. (2006). The wrf single-moment 6-class microphysics scheme
841 (wsm6). *Journal of the Korean Meteorological Society*, 42.
- 842 Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A New Vertical Diffusion Package with an
843 Explicit Treatment of Entrainment Processes. *Monthly Weather Review*, 134(9), 2318.
844 doi: 10.1175/MWR3199.1
- 845 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins,
846 W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the
847 aer radiative transfer models. *Journal of Geophysical Research Atmospheres*, 113. doi:
848 10.1029/2008JD009944
- 849 Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling*
850 *Earth Systems*, 9. doi: 10.1002/2017MS001059
- 851 Jiang, Q., Doyle, J. D., Eckermann, S. D., & Williams, B. P. (2019). Stratospheric trailing
852 gravity waves from new zealand. *Journal of the Atmospheric Sciences*, 76(6), 1565 - 1586.
853 Retrieved from [https://journals.ametsoc.org/view/journals/atasc/76/6/jas-d-18](https://journals.ametsoc.org/view/journals/atasc/76/6/jas-d-18-0290.1.xml)
854 [-0290.1.xml](https://journals.ametsoc.org/view/journals/atasc/76/6/jas-d-18-0290.1.xml) doi: 10.1175/JAS-D-18-0290.1
- 855 Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-
856 Bustamante, E. (2012). A revised scheme for the wrf surface layer formulation. *Monthly*
857 *Weather Review*, 140. doi: 10.1175/MWR-D-11-00056.1
- 858 Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmailzadeh, S., ... others
859 (2021). Physics-informed machine learning: case studies for weather and climate mod-
860 elling. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200093.
- 861 Kim, Y., Eckermann, S. D., & Chun, H. (2003). An overview of the past, present and
862 future of gravity-wave drag parametrization for numerical climate and weather prediction
863 models. *Atmosphere-Ocean*, 41, 65-98. Retrieved from [https://doi.org/10.3137/ao](https://doi.org/10.3137/ao.410105)
864 [.410105](https://doi.org/10.3137/ao.410105) doi: 10.3137/ao.410105
- 865 Kim, Y. H., Bölöni, G., Borchert, S., Chun, H. Y., & Achatz, U. (2021). Toward transient
866 subgrid-scale gravity wave representation in atmospheric models. part ii: Wave intermit-
867 tency simulated with convective sources. *Journal of the Atmospheric Sciences*, 78. doi:
868 10.1175/JAS-D-20-0066.1
- 869 Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high
870 resolution global atmospheric models for climate studies. *Atmospheric Science Letters*,
871 21. doi: 10.1002/asl.952
- 872 Kruse, C. G., Alexander, M. J., Hoffmann, L., Niekerk, A. V., Polichtchouk, I., Bacmeister,
873 J. T., ... Stein, O. (2022). Observed and modeled mountain waves from the surface to

- 874 the mesosphere near the drake passage. *Journal of the Atmospheric Sciences*, *79*. doi:
875 10.1175/JAS-D-21-0252.1
- 876 Kruse, C. G., & Smith, R. B. (2015). Gravity wave diagnostics and characteristics in
877 mesoscale fields. *Journal of the Atmospheric Sciences*, *72*. doi: 10.1175/JAS-D-15-0079.1
- 878 Kruse, C. G., Smith, R. B., & Eckermann, S. D. (2016). The midlatitude lower-stratospheric
879 mountain wave "valve layer". *Journal of the Atmospheric Sciences*, *73*. doi: 10.1175/
880 jas-d-16-0173.1
- 881 Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In
882 *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier.
- 883 Lott, F., Guez, L., & Maury, P. (2012). A stochastic parameterization of non-orographic
884 gravity waves: Formalism and impact on the equatorial stratosphere. *Geophysical Re-*
885 *search Letters*, *39*. doi: 10.1029/2012GL051001
- 886 Lynch, P. (1988). Deducing the wind from vorticity and divergence. *Monthly Weather*
887 *Review*, *116*. doi: 10.1175/1520-0493(1988)116<0086:DTWFVA>2.0.CO;2
- 888 Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the fidelity of
889 explainable artificial intelligence methods for applications of convolutional neural networks
890 in geoscience. *Artificial Intelligence for the Earth Systems*, *1*(4), e220012.
- 891 Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020).
892 Application of Deep Learning to Estimate Atmospheric Gravity Wave Parameters in Re-
893 analysis Data Sets. , *47*(19), e89436. doi: 10.1029/2020GL089436
- 894 Maulik, R., San, O., Rasheed, A., & Vedula, P. (2019). Subgrid modelling for two-
895 dimensional turbulence using neural networks. *Journal of Fluid Mechanics*, *858*, 122–
896 144.
- 897 McFarlane, N. A. (1987). The Effect of Orographically Excited Gravity Wave Drag on the
898 General Circulation of the Lower Stratosphere and Troposphere. *Journal of Atmospheric*
899 *Sciences*, *44*(14), 1775–1800. doi: 10.1175/1520-0469(1987)044<1775:TEOOEG>2.0.CO;2
- 900 Muraschko, J., Fruman, M. D., Achatz, U., Hickel, S., & Toledo, Y. (2015). On the
901 application of wenzel-kramer-brillouin theory for the simulation of the weakly nonlinear
902 dynamics of gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *141*.
903 doi: 10.1002/qj.2381
- 904 Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblus-Reyes, F. J., Jung, T., & Leutbecher, M.
905 (2005). Representing model uncertainty in weather and climate prediction. *Annual Review*
906 *of Earth and Planetary Sciences*, *33*. doi: 10.1146/annurev.earth.33.092203.122552
- 907 Palmer, T. N., Shutts, G. J., & Swinbank, R. (1986). Alleviation of a systematic westerly
908 bias in general circulation and numerical weather prediction models through an orographic
909 gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*,
910 *112*. doi: 10.1002/qj.49711247406
- 911 Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., . . .
912 others (2022). Fourcastnet: A global data-driven high-resolution weather model using
913 adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- 914 Pincus, R., Barker, H. W., & Morcrette, J. J. (2003). A fast, flexible, approximate technique
915 for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical*
916 *Research: Atmospheres*, *108*. doi: 10.1029/2002jd003322
- 917 Plougonven, R., & Zhang, F. (2014). Internal gravity waves from atmospheric jets and
918 fronts. *Reviews of Geophysics*, *52*. doi: 10.1002/2012RG000419
- 919 Polichtchouk, I., Wedi, N., & Kim, Y. H. (2022). Resolved gravity waves in the tropi-
920 cal stratosphere: Impact of horizontal resolution and deep convection parametrization.
921 *Quarterly Journal of the Royal Meteorological Society*, *148*. doi: 10.1002/qj.4202
- 922 Pope, S. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*,
923 *72*(2), 331–340.
- 924 Pope, S. (2000). *Turbulent flows*. Cambridge university press.
- 925 Qiao, W., Song, J., He, H., & Li, F. (2019). Application of different wind field models
926 and wave boundary layer model to typhoon waves numerical simulation in wavewatch
927 iii model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *71*. doi: 10.1080/

- 16000870.2019.1657552
- 928 Rai, S., Hecht, M., Maltrud, M., & Aluie, H. (2021). Scale of oceanic eddy killing by wind
929 from global satellite observations. *Science Advances*, 7(28), eabf4920.
- 930 Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid
931 processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39),
932 9684–9689.
- 933 Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simp-
934 son, I. R. (2020, 4). Progress in simulating the quasi-biennial oscillation in cmip models.
935 *Journal of Geophysical Research: Atmospheres*, 125. doi: 10.1029/2019JD032362
- 936 Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., . . . Yukimoto,
937 S. (2022, 4). Response of the quasi-biennial oscillation to a warming climate in global
938 climate models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1490-1518.
939 doi: 10.1002/qj.3749
- 940 Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity
941 wave source parameterization in a general circulation model. *Journal of the Atmospheric*
942 *Sciences*, 67. doi: 10.1175/2009JAS3112.1
- 943 Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Springer
944 Science & Business Media.
- 945 Sato, K., Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., & Takahashi, M. (2009).
946 On the origins of mesospheric gravity waves. *Geophysical Research Letters*, 36. doi:
947 10.1029/2009GL039908
- 948 Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A
949 blueprint for models that learn from observations and targeted high-resolution simula-
950 tions. *Geophysical Research Letters*, 44(24), 12–396.
- 951 Scinocca, J. F., & McFarlane, N. A. (2000). The parametrization of drag induced by
952 stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological*
953 *Society*, 126. doi: 10.1002/qj.49712656802
- 954 Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction sys-
955 tems. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric*
956 *sciences, applied meteorology and physical oceanography*, 131(612), 3079–3102.
- 957 Shutts, G. J., & Vosper, S. B. (2011). Stratospheric gravity waves revealed in nwp model
958 forecasts. *Quarterly Journal of the Royal Meteorological Society*, 137. doi: 10.1002/
959 qj.763
- 960 Sigmond, M., & Scinocca, J. F. (2010). The influence of the basic state on the northern
961 hemisphere circulation response to climate change. *Journal of Climate*, 23(6), 1434–1446.
- 962 Skamarock, W. C. (2004). Evaluating mesoscale nwp models using kinetic energy spectra.
963 *Monthly Weather Review*, 132. doi: 10.1175/MWR2830.1
- 964 Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model
965 for weather research and forecasting applications. *Journal of Computational Physics*, 227.
966 doi: 10.1016/j.jcp.2007.01.037
- 967 Skamarock, W. C., Snyder, C., Klemp, J. B., & Park, S. H. (2019). Vertical resolution
968 requirements in atmospheric simulation. *Monthly Weather Review*, 147. doi: 10.1175/
969 MWR-D-19-0043.1
- 970 Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., . . . Ziemer,
971 F. (2022, 4). Atmospheric energy spectra in global kilometre-scale models. *Tellus A:*
972 *Dynamic Meteorology and Oceanography*, 74, 280-299. doi: 10.16993/tellusa.26
- 973 Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H.
974 (2019). Intercomparison of gravity waves in global convection-permitting models. *Journal*
975 *of the Atmospheric Sciences*, 76. doi: 10.1175/JAS-D-19-0040.1
- 976 Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2022). Explaining the physics
977 of transfer learning in data-driven turbulence modeling. *arXiv preprint arXiv:2206.03198*.
- 978 Sun, Y. Q., & Zhang, F. (2016). Intrinsic versus practical limits of atmospheric predictability
979 and the significance of the butterfly effect. *Journal of the Atmospheric Sciences*, 73(3),
980 1419 - 1438. Retrieved from <https://journals.ametsoc.org/view/journals/atsc/>
981

- 982 73/3/jas-d-15-0142.1.xml doi: 10.1175/JAS-D-15-0142.1
- 983 Wang, P., Yuval, J., & O’Gorman, P. A. (2022). Non-local parameterization of atmospheric
984 subgrid processes with neural networks. *arXiv preprint arXiv:2201.00417*. doi: 10.48550/
985 ARXIV.2201.00417
- 986 Wei, J., Bölöni, G., & Achatz, U. (2019). Efficient modeling of the interaction of
987 mesoscale gravity waves with unbalanced large-scale flows: Pseudomomentum-flux con-
988 vergence versus direct approach. *Journal of the Atmospheric Sciences*, *76*. doi: 10.1175/
989 JAS-D-18-0337.1
- 990 Wei, J., Zhang, F., Richter, J. H., Alexander, M. J., & Sun, Y. Q. (2022). Global dis-
991 tributions of tropospheric and stratospheric gravity wave momentum fluxes resolved
992 by the 9-km ecmwf experiments. *Journal of the Atmospheric Sciences*, *79*. doi:
993 10.1175/JAS-D-21-0173.1
- 994 Yuval, J., & O’Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid
995 processes for climate modeling at a range of resolutions. *Nature communications*, *11*(1),
996 1–10.
- 997 Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures.
998 *Geophysical Research Letters*, *47*(17), e2020GL088376.
- 999 Zanna, L., & Bolton, T. (2021). Deep learning of unresolved turbulent ocean processes in
1000 climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to*
1001 *Remote Sensing, Climate Science, and Geosciences*, 298–306.
- 1002 Zhou, Z., He, G., Wang, S., & Jin, G. (2019). Subgrid-scale model for large-eddy simulation
1003 of isotropic turbulent flows using an artificial neural network. *Computers & Fluids*, *195*,
1004 104319.
- 1005 Žagar, N., Kasahara, A., Terasaki, K., Tribbia, J., & Tanaka, H. (2015). Normal-mode
1006 function representation of global 3-d data sets: Open-access software for the atmospheric
1007 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 Sun¹, Pedram Hassanzadeh¹, M. Joan Alexander², and Christopher

G. Kruse²

¹Rice University, Houston, Texas, United States

²NorthWest 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.

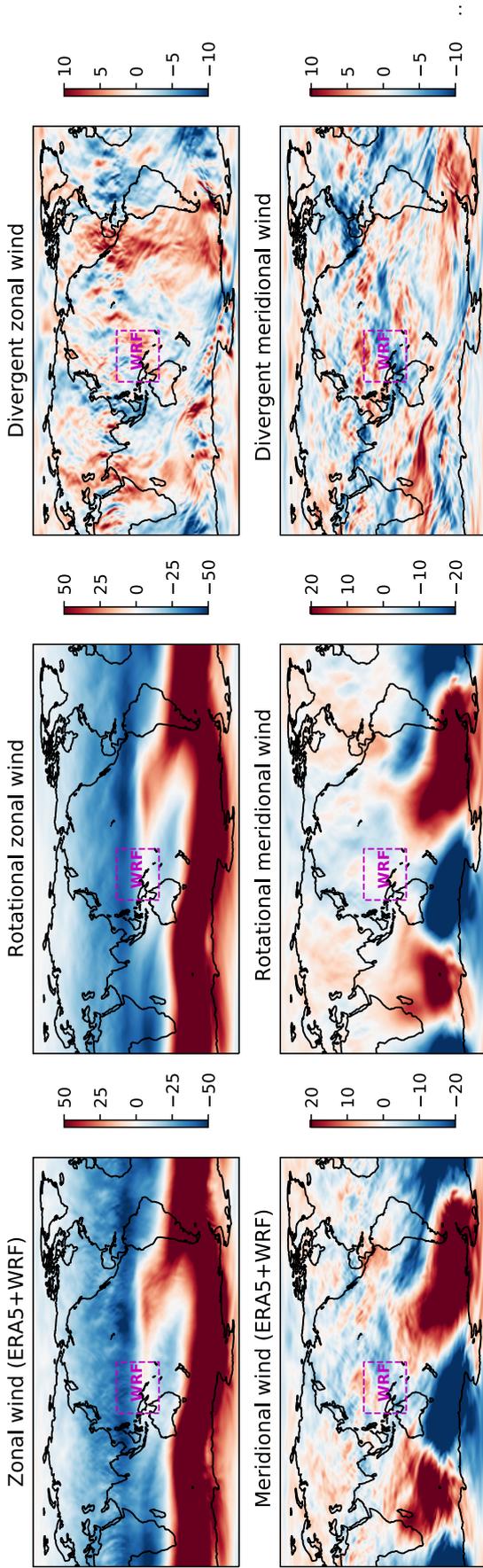


Figure S1. Snapshots from the Helmholtz decomposition approach used in this study.

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>

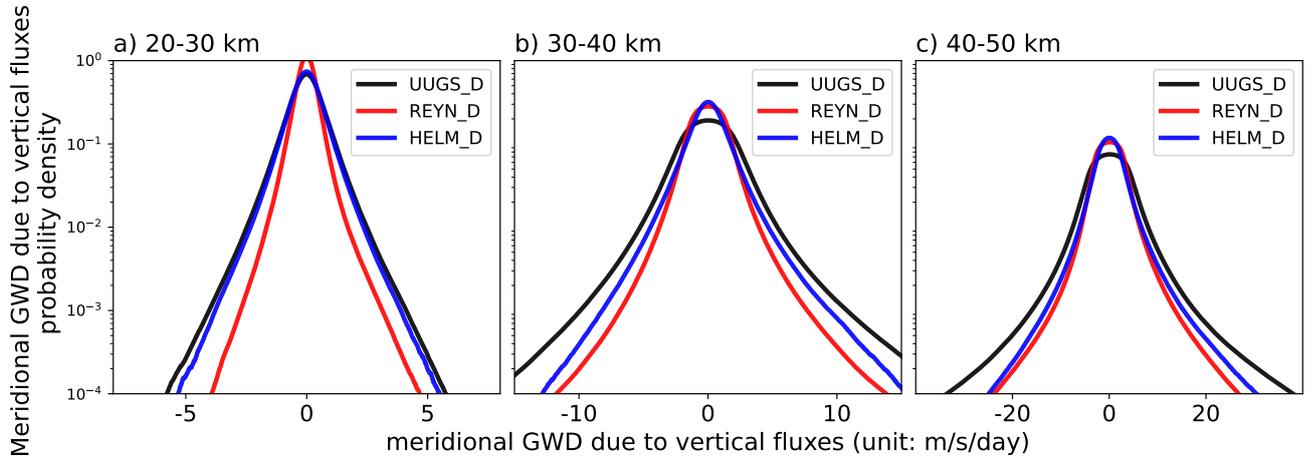


Figure S2. Same as Fig. 6, but for the meridional GWD due to vertical fluxes (last term of Eq. (A11)). Data from Cases 1 and 2 have been combined in making these PDFs.

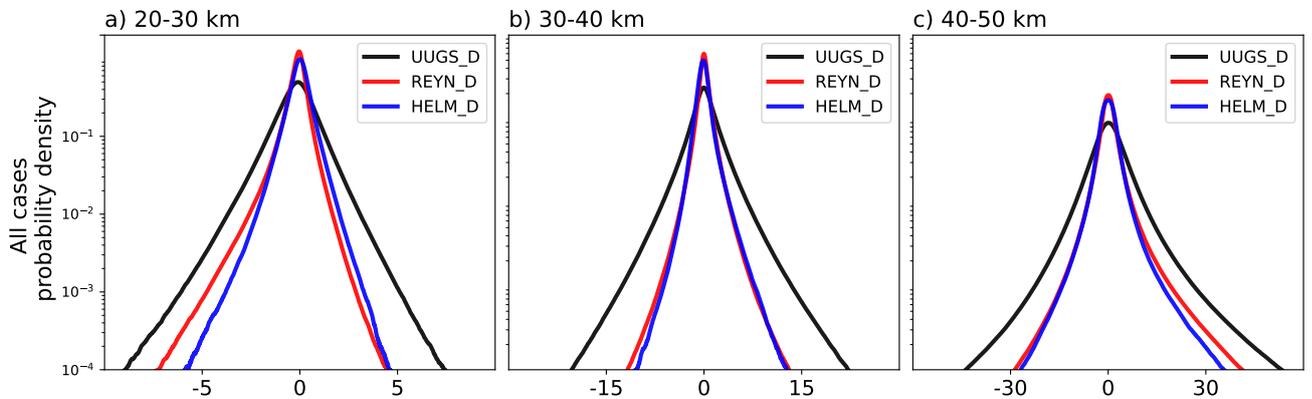


Figure S3. Same as Fig. 6, but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computational cost of the algorithm.

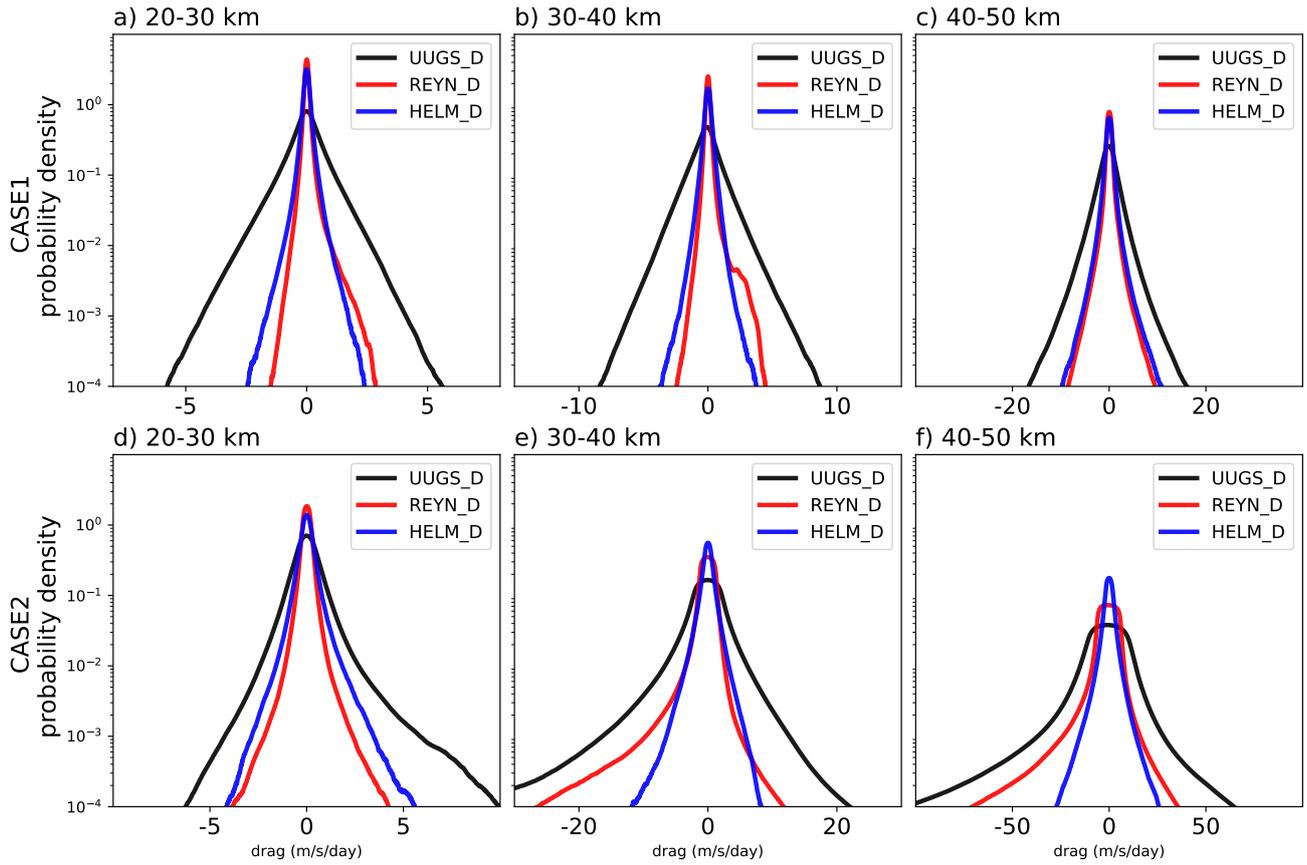


Figure S4. Same as Fig. 10, but for the zonal fluxes.

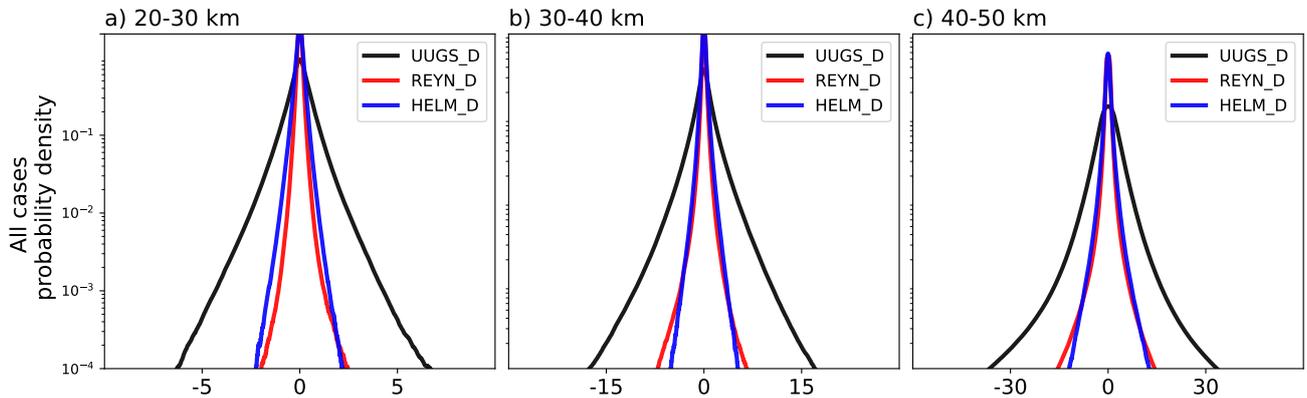


Figure S5. Same as Fig. 10 but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computational cost of the algorithm.

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

4 **Y. Qiang Sun¹, Pedram Hassanzadeh¹, M. Joan Alexander², Christopher G.**
5 **Kruse²**

6 ¹Rice University, Houston, Texas, United States

7 ²NorthWest Research Associates, Boulder, Colorado, United States

8 **Key Points:**

- 9 • In a library of WRF simulations, we compare methods for estimating 3D gravity wave
10 drag force that are un- and under-resolved by GCMs.
11 • For drag associated with vertical fluxes, different methods agree on time- and zonal-
12 mean but not on instantaneous spatiotemporal patterns.
13 • Drag associated with horizontal fluxes is significant but is very sensitive to the esti-
14 mation 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 ($\Delta x = 3 \text{ km}$) 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 $O(1 \text{ km})$ to $O(100 \text{ km})$ 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 $\sim 1 \text{ km}$ 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-

65 opment, gravity wave parameterization (GWP) is now a critical component of GCMs to
 66 enable them to reproduce realistic atmospheric circulation mean, variability, and response
 67 to climate change (e.g., McFarlane, 1987; Scinocca & McFarlane, 2000; Y. Kim et al., 2003;
 68 Beres, 2004; Alexander et al., 2010; Richter et al., 2010; Lott et al., 2012; Plougonven &
 69 Zhang, 2014). For example, generating quasi-biennial oscillation (QBO) in GCMs requires
 70 skillful GWPs (e.g., Richter et al., 2020).

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

89 Recently, Machine Learning (ML) techniques have emerged as alternative tools for
 90 developing parameterizations for climate models. They have been used in parameterizing a
 91 variety of SGS processes with promising results (e.g., Schneider et al., 2017; Rasp et al., 2018;
 92 Bolton & Zanna, 2019; Maulik et al., 2019; Chattopadhyay et al., 2020; Yuval & O’Gorman,
 93 2020; Kashinath et al., 2021; Gentine et al., 2021; Guan et al., 2022). Matsuoka et al.
 94 (2020) were among the first to apply ML to GWs. Focused on the orographic GWs over the
 95 Hokkaido region of Japan, they trained a convolutional neural network to connect the large-
 96 scale tropospheric state and the small-scale GW wind fluctuations in the lower stratosphere.
 97 Recently, Amiramjadi et al. (2022) also found success in reconstructing the non-orographic
 98 GWs in the ERA5 dataset with a random forest regressor. Both of these studies identified
 99 fluctuations associated with GWs using a simple moving-box average and demonstrated the
 100 feasibility of using ML to represent GWs. However, these studies only focused on learning
 101 GWs or momentum fluxes at one level (100 *hPa*), without further calculating the GWD,
 102 which is required to develop GWPs for GCMs. A number of other studies have also shown
 103 the power of ML for GWP through emulating current GWP schemes (Chantry et al., 2021;
 104 Espinosa et al., 2022). These emulation efforts provide valuable insight on various promises
 105 and challenges of using ML for GWPs, though a number of key challenges, e.g., related to
 106 GWD extraction and lateral GW propagation, cannot be investigated through emulation (see
 107 below).

108 One key challenge for the data-driven approach is the availability of sufficient obser-
 109 vationally constrained data of GW momentum transport for training the ML algorithms.
 110 With limited availability of observations of GWs and the challenges associated with sparsity
 111 and noise, high-resolution GW-resolving model simulations must play a critical role in gen-
 112 erating the training data. A number of case studies have verified that high-resolution models
 113 are able to capture the key characteristics of observed GWs (Bramberger et al., 2020; Kruse
 114 et al., 2022). The second key challenge in the most common data-driven approach (the
 115 so-called “supervised” or “offline” learning) is the need to extract, from the high-resolution
 116 simulations, the true GWD due to the un- and under-resolved GWs; hereafter, we refer
 117 to this collectively as the *SGS drag*. This SGS GWD is what has to be added to a low-

118 resolution GCM to properly account for the un- and under-resolved GWs (note that this
 119 depends on the GCM's *effective resolution*; more discussions to follow later). In the GW
 120 modeling community, a number of methods have been used in the past to separate GWs
 121 from the large-scale flow and quantify the SGS fluxes or GWD (e.g., G. J. Shutts & Vosper,
 122 2011; Kruse & Smith, 2015; Žagar et al., 2015; Stephan et al., 2019; Matsuoka et al., 2020;
 123 Amiramjadi et al., 2022; Polichtchouk et al., 2022). Some studies pursued a simple box-
 124 averaging method (e.g., Matsuoka et al., 2020) or a cut-off low-pass filter (e.g., Polichtchouk
 125 et al., 2022). There are also more rigorous methods to separate the balanced large-scale
 126 components and the unbalanced GWs based on linear wave theory and Helmholtz decom-
 127 position (e.g., Callies et al., 2014; Žagar et al., 2015). Stephan et al. (2019) computed the
 128 resolved GW pseudo-momentum fluxes in month-long global convection-permitting simula-
 129 tions with two other methods. These methods showed overall good agreement on the general
 130 shape of the longitudinal profiles of GW fluxes, but systematic differences were found for
 131 the amplitudes of the pseudo-momentum fluxes even after averaging over the 30-day period,
 132 implying the importance of the extraction method if we were to use these high-resolution
 133 data for training the ML algorithms.

134 The third key challenge is related to the 3D propagation of GWs and the resulting
 135 3D SGS GWD. The aforementioned studies focused only on the vertical momentum fluxes
 136 of GWs, as the current operational GWP schemes ignore lateral propagation of the waves
 137 and the resulting lateral fluxes and their contribution to the total GWD. However, there
 138 is growing evidence that horizontal propagation of GWs has to be considered in GWP
 139 to produce a realistic atmospheric circulation (e.g., Sato et al., 2009; Muraschko et al.,
 140 2015; Ehard et al., 2017). Only few studies have discussed the lateral momentum fluxes
 141 in high-resolution simulations (Eckermann et al., 2015; Jiang et al., 2019). More recently,
 142 through analyzing the lateral momentum flux in the high-resolution simulations over the
 143 Drake Passage, Kruse et al. (2022) showed that the meridionally propagating mountain
 144 waves significantly enhanced the zonal drag. Additionally, their work suggested that not
 145 accounting for these meridional fluxes would result in GWD in the wrong direction at and
 146 below the polar night jet.

147 In this paper, we use data from a library of 20 convection-permitting (3 km) tropical
 148 WRF simulations to

- 149 1. Compare 3 methods that are commonly used in the GW and large-eddy simula-
 150 tion (LES) literature to quantify the SGS fluxes and drags. These methods are i)
 151 Helmholtz decomposition, ii) Spatial filtering to compute the full SGS stress and the
 152 resulting GWD, and iii) Same as (ii) but only for the Reynolds stress.
- 153 2. Quantify the contribution of both vertical and horizontal fluxes of horizontal mo-
 154 mentum to the total GWD to investigate if the latter should be included in SGS
 155 parameterizations too.

156 Item 1 is crucial because any data-driven method, ML or otherwise, is as good as the data
 157 used for the training. Note that the challenges associated with extracting the SGS terms
 158 for ML training are not limited to the GW applications, and are in fact relevant to many
 159 climate/turbulence processes and currently an active area of research (e.g., Zhou et al., 2019;
 160 Zanna & Bolton, 2021; Grooms et al., 2021; Beck & Kurz, 2021; Guan et al., 2022).

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

166 In the rest of the paper, we will first introduce the high-resolution data and the 3 meth-
 167 ods used for SGS GWD extraction. We will then compare the results for the SGS vertical

168 momentum flux and GWD, followed by similar analyses for the SGS GWD associated with
 169 lateral momentum fluxes. Concluding remarks and discussions are in the last section.

170 2 Data and methods

171 2.1 WRF setup and data

172 All data used in this study are generated using WRF, with initial conditions from re-
 173 analysis data and boundaries nudged towards reanalysis data. For the purpose of this work,
 174 the WRF model is modified according to Kruse et al. (2022) to support a deep configura-
 175 tion that runs up to 1 *Pa* (~ 80 *km*). Two key modifications of the released WRF version
 176 4.1 model are made to achieve the high model-top here. First, low-order interpolators are
 177 used to prevent the over-/under-shooting of default higher-order interpolators, preventing
 178 the intersection of analysis levels near complex topography after horizontal interpolation.
 179 Second, the default lateral relaxation is replaced with grid-point nudging confined to the
 180 lateral boundaries for the model to run stably.

181 For now, the library only includes the tropical regions (see the domains in Fig. 1). We
 182 have conducted a total of 20 simulations in 6 domains, where the dates of the week-long
 183 runs are chosen to sample the seasonal cycle, QBO phases, and precipitation distribution
 184 (Fig. 1c). Two of these simulations, one from the 2016 summertime all-ocean West Pacific
 185 (WP), which is in the westerly phase of QBO, and one from the 2020 wintertime land-ocean
 186 Indian Ocean (IO), which is in the easterly phase of QBO, are chosen as representative cases 1
 187 and 2, respectively. The first day of all simulations is treated as spin-up periods and not used
 188 in analyses. The horizontal domain size is 3600 *km* \times 3600 *km*. The simulations are done at
 189 3 *km* grid spacing. There are 180 vertical model levels in total. The vertical grid spacing is
 190 close to 200 *m* near the lower boundary and gradually increases to a maximum of 600 *m* near
 191 the model top. For these tropical simulations, we largely use the ‘‘Tropical’’ WRF physics
 192 suite (e.g., Qiao et al., 2019), but with a different surface layer scheme. The parameterization
 193 set includes the WRF Single-Moment 6-class (WSM6) microphysics scheme (Hong & Lim,
 194 2006), the Yonsei University planetary boundary layer scheme (Hong et al., 2006), the
 195 RRTM (Rapid Radiative Transfer Model) for longwave and shortwave radiation (Iacono et
 196 al., 2008; Pincus et al., 2003), and the revised surface layer scheme developed in Jim3nez
 197 et al. (2012). Note that no cumulus scheme is used given the 3 *km* grid spacing and, most
 198 importantly, no GWP is used.

199 Both the initial condition and the boundary condition come from the fifth-generation
 200 European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA5).
 201 As we are using the same method as in Kruse et al. (2022) to nudge the simulation boundaries
 202 towards the ERA5 data (hence limiting the GW signals there), the model output data near
 203 the domain boundary (< 300 *km*) are neglected when conducting analysis for the GWs.
 204 The analysis domains are hence 3000 *km* \times 3000 *km*.

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

209 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 $(\bar{\cdot})$, and b) coarse-graining, denoted with $(\tilde{\cdot})$. For any variable $\phi(\mathbf{x}, t)$, spatial filtering
 is defined as (e.g., Sagaut, 2006; Grooms et al., 2021; Guan et al., 2022)

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

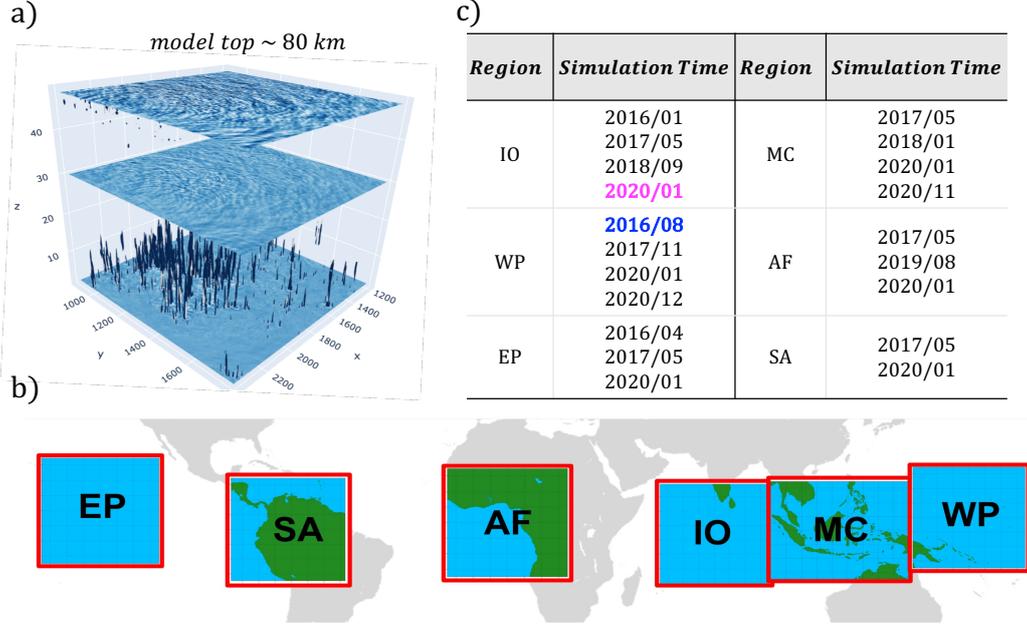


Figure 1. a) A snapshot of vertical velocity in a WRF simulation. b) Tropical regional domains where the WRF simulations are conducted. c) The list of 20 week-long WRF simulations conducted in this regional tropical channel setup, among which two representative cases are chosen (blue color represents 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 $\phi(\mathbf{x}, t)$ can be separated into two components

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

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

Two key points need to be clarified here. One is that following the convention used in recent literature (e.g., 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 $\tilde{\phi} \neq \bar{\phi}$ 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, \mathbf{r} and \mathbf{k} are coordinates in physical space and spectral space, respectively, with $\mathbf{r} = (r_x, r_y)$, and $\mathbf{k} = (k_x, k_y)$. $\hat{(\cdot)}$ is the Fourier transform, and Δ is the filter size as in Eq. (1).

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

229 the un-resolved and under-resolved GWs from the resolved GWs, or both. The choice of
 230 filter type (e.g., Gaussian, top-hat or box, sharp-spectral) can affect the extracted SGS
 231 terms, as already shown in a number of past studies including in the context of geophysical
 232 turbulence (e.g., Leonard, 1975; Zanna & Bolton, 2021; Beck & Kurz, 2021). Figure B1
 233 shows an example of the effect of filter type on the spectrum of zonal wind from our WRF
 234 simulations. Different low-pass filters (e.g., top-hat and sharp-spectral) have been used in
 235 previous studies to separate the GWs from the large-scale background (Kruse & Smith,
 236 2015; Matsuoka et al., 2020; Polichtchouk et al., 2022), though a systematic study on the
 237 effect of filter type and the potential implications for the extracted SGS terms is lacking.

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

250 To systematically quantify the effects of filter type and sizes, here, we use 3 filter
 251 types and two length scales $\Delta = 200$ km and 700 km to help with understanding the scale-
 252 awareness when building a data-driven GWP in the future (Table 1). Note that these choices
 253 of Δ are motivated by assuming that the low-resolution GCM has grid spacing of 100 km
 254 ($\sim 1^\circ$ resolution). $\Delta = 200$ km is based on the common choice for Δ in the LES literature,
 255 i.e., twice the low-resolution model’s grid spacing (Pope, 2000; Sagaut, 2006; Guan et al.,
 256 2022). $\Delta = 700$ km is based on the effective-resolution study of Skamarock (2004); this is
 257 the filter size used for the presented results, unless indicated otherwise.

258 Once resolved fluxes are quantified point-wise on the original grid, the effective fluxes
 259 within a hypothetical GCM grid cell must be computed on a coarse GCM grid. As mentioned
 260 before, we refer to this operation as coarse-graining. Admittedly, this terminology has
 261 not been uniformly adopted in the literature, though it has been recommended by several

262 recent studies (Grooms et al., 2021; Guan et al., 2022). Also, note that in some studies
 263 filtering and coarse-graining are done via just one operation, rather than two separate ones
 264 (e.g., Brenowitz & Bretherton, 2018; Yuval & O’Gorman, 2020). With all these issues in
 265 mind, here, we use one commonly used coarse-graining strategy: we simply truncate the
 266 wavenumbers greater than the cut-off wavenumber corresponding to the GCM grid spacing
 267 (100 km in this case). Note that in this study, for computational efficiency, both filtering and
 268 coarse-graining are done in the spectral (Fourier) space, and mirrored tiles are added around
 269 the original domain following Sun & Zhang (2016) to reduce problems with non-periodic
 270 boundaries.

271 To better illustrate the effects of these filtering and coarse-graining operations, Fig. 2
 272 shows examples of the high-resolution WRF snapshots, and filtered (Gaussian with $\Delta =$
 273 700 km) and coarse-grained 3D velocity fields at 30 km height. The full u, v, w in the
 274 3000 km \times 3000 km domain are shown in the left column. After the filtering operation,
 275 the velocity fields are separated into the large-scale (second column) and the perturbation
 276 (third column) components. We also apply coarse-graining operators to these fields (fourth
 277 and fifth columns) to transfer them to a 30 \times 30 grid, similar to that of a GCM with a
 278 grid spacing of 100 km. From this plot, we notice systematic differences between horizontal
 279 winds and vertical winds. For the horizontal winds u and v , the large-scale background (\tilde{u}
 280 and \tilde{v}) are much larger in amplitudes than the small-scale perturbations u' and v' , whereas
 281 for the vertical velocity, the large-scale background is almost negligible, with all the signal
 282 at small scales w' . Moreover, notice that there can be significant differences between $\tilde{\phi}$ and
 283 $\bar{\phi}$, which implies that $\tilde{\phi}' \neq 0$ with the Gaussian filtering applied here.

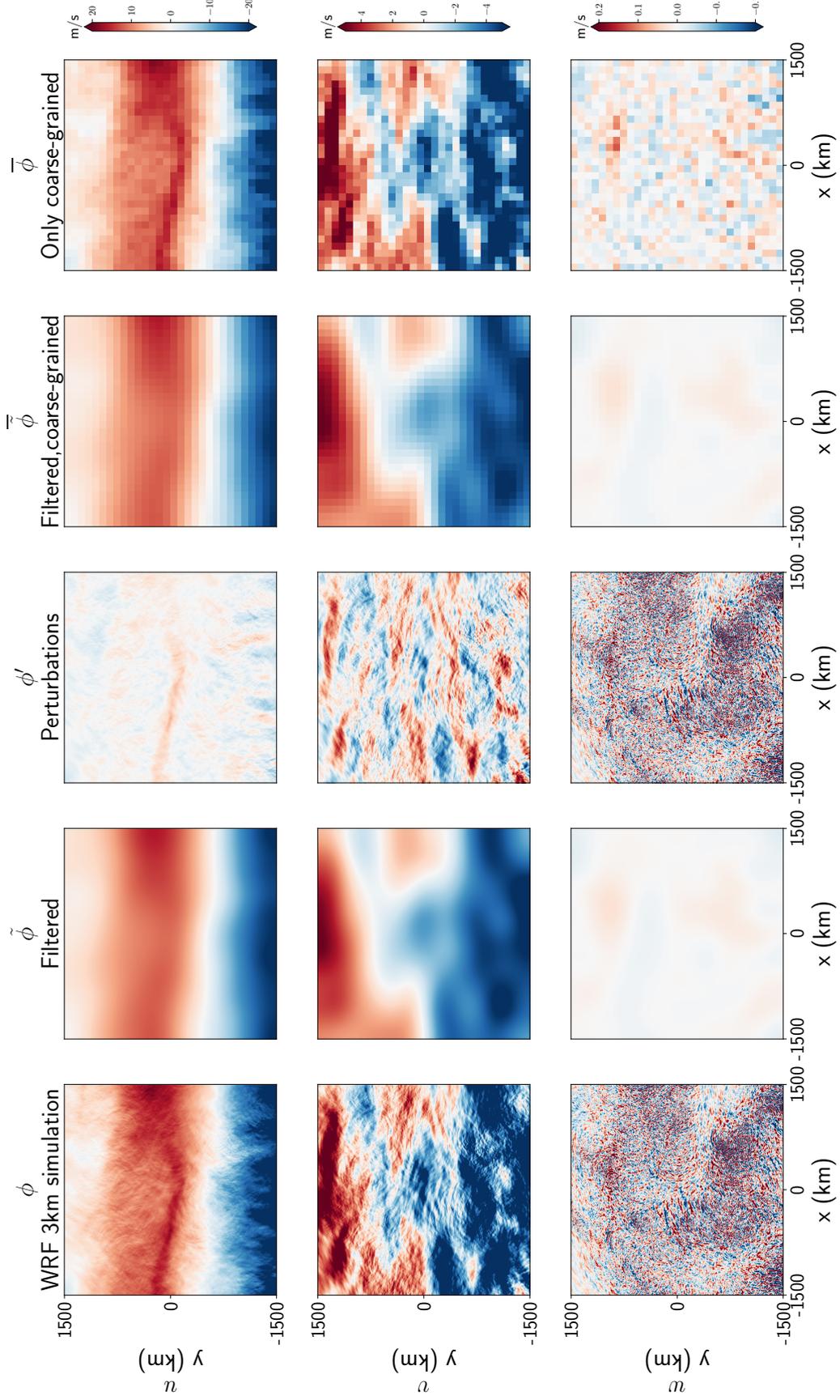


Figure 2. Examples of the effects of low-pass filtering and coarse-graining operators used in this study. The snapshots are from Case 1, on August 3rd 2016, 12:00 UTC, at 30 km height. A Gaussian filter with $\Delta = 700$ km is used. The coarse-graining is done by truncating all wavenumbers greater 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.nc1.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

333 waves and large-scale GWs could be resolved by the GCM, an additional high-pass
 334 spatial filter is applied to remove the signals that could be fully resolved by the low-
 335 resolution GCM grid. Here, we use a Gaussian high-pass filter with filter size of
 336 700 km.

337 The outcome of step (e) is the GW perturbations for the horizontal winds (u'_H and
 338 v'_H). The subscript H here denotes the use of Helmholtz decomposition in the process. The
 339 vertical winds w are not involved here in the Helmholtz decomposition of horizontal winds.
 340 Given that vertical wind w is dominated by small scales and has negligible large-scale signals
 341 (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:

$$MF_x = [MF_{xx} \quad MF_{yx} \quad MF_{zx}] \quad (3)$$

where

$$MF_{xx} = \tilde{\rho} \widetilde{u'_H u'_H} \quad (4)$$

$$MF_{yx} = \tilde{\rho} \widetilde{u'_H v'_H} \quad (5)$$

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

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

346 The first two components in Eq. (3) are the zonal and meridional flux of zonal mo-
 347 mentum due to SGS GWs, respectively. They will also be referred to as lateral momentum
 348 fluxes. The last component in Eq. (3) is the vertical flux of zonal momentum due to SGS
 349 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 ,

$$\overline{MF}_x = [\overline{MF}_{xx} \quad \overline{MF}_{yx} \quad \overline{MF}_{zx}]. \quad (7)$$

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 :

$$\overline{GWD}_{xx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{xx}}{\partial x} \quad (8)$$

$$\overline{GWD}_{yx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{yx}}{\partial y} \quad (9)$$

$$\overline{GWD}_{zx} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{MF}_{zx}}{\partial z}. \quad (10)$$

350 Note that Eq. (10), the vertical divergence of the vertical flux of zonal momentum due to
 351 SGS GWs, is often considered to be the dominant component in previous studies, and the
 352 only term that is conventionally represented in existing SGS parameterizations. This has
 353 been the case in the development of physics-based GWP, and in the past efforts focused
 354 on extracting SGS GWD from high-resolution simulations (e.g., Alexander et al., 2010;
 355 Matsuoka et al., 2020; Polichtchouk et al., 2022). Yet, as shown in Kruse et al. (2022), this
 356 is not always the case, and the lateral divergence of lateral momentum fluxes (Eqs. (8) and
 357 (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

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

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 $\overline{u'w'}$ is a part of the (but not the entire) total SGS vertical flux, $\overline{u\tilde{w}} - \tilde{u}\tilde{w}$. The other two components (e.g., $\overline{u\tilde{v}}$) arise because as mentioned before, in spatial filtering and coarse-graining, terms like 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 $\overline{u'u'}$ and $\overline{u'v'}$ as well as other stresses, including $\overline{u'u'}$ and $\overline{v'u'}$. 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

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

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

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

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.

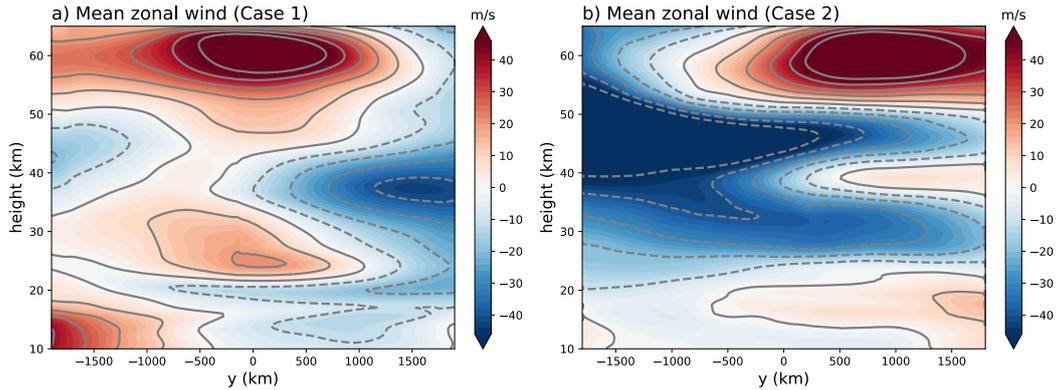


Figure 3. Zonal-mean, time-mean zonal wind u for two representative WRF simulation cases. The contour interval is 10 m , with dash lines implying zero and negative values. a) Case 1 with westerly QBO winds; b) Case 2 with easterly QBO winds .

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 $\sim 20\text{ m/s}$ in the tropical stratosphere ($\sim 25\text{ km}$). The semiannual oscillation (SAO) can also be seen near the stratopause ($\sim 60\text{ km}$), 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 ($\Delta = 700\text{ km}$) 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

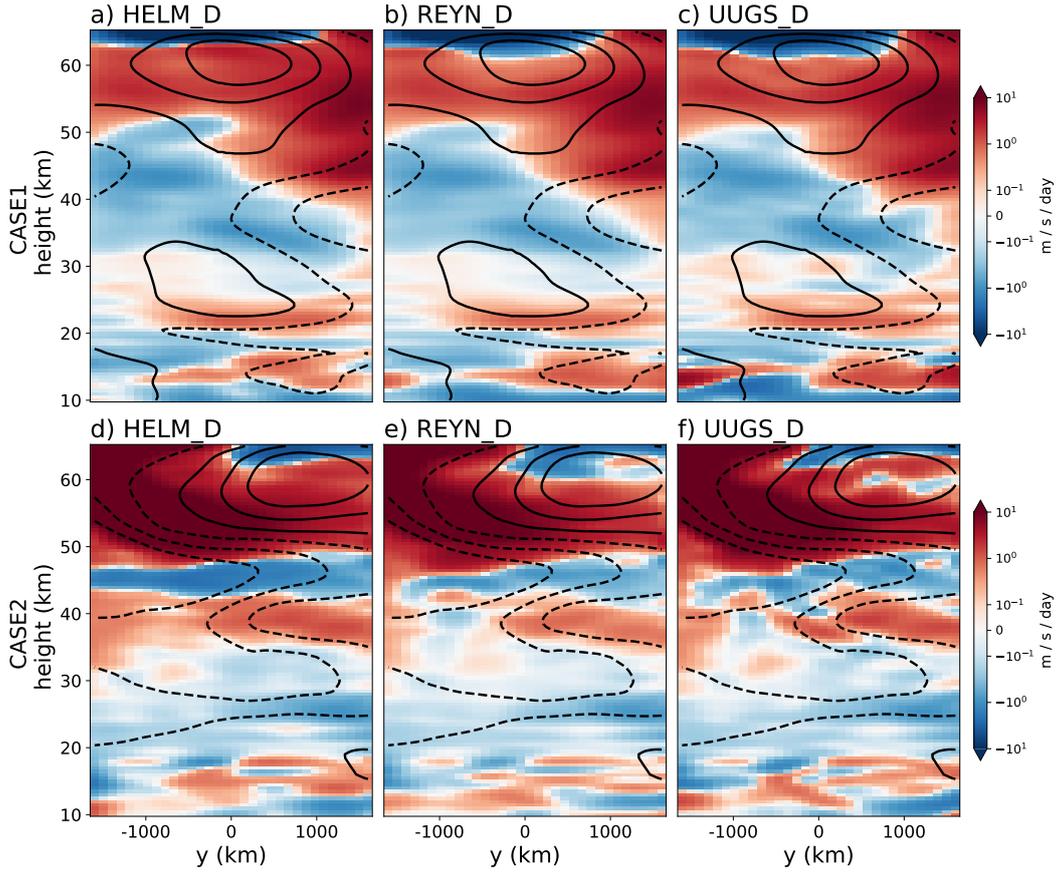


Figure 4. Zonal-mean, time-mean zonal SGS GWD (shading) due to the un-/under-resolved (SGS) vertical fluxes derived using three methods for both representative cases (upper row: Case 1, lower row: Case 2). Left column, a & d: Helmholtz decomposition method (HELM_D). Middle column, b & e: Reynolds stress 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 .

422 GWs help maintain and drive the zonal wind there (Alexander et al., 2010). Moreover, for
 423 the QBO region, the maximum drag is below the wind maximum (e.g., Case 1 in Fig. 4),
 424 implying the role of SGS GWs in the downward propagation of the zonal winds.

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

437 Figure 6 shows the probability density functions (PDFs) for the zonal SGS GWD as-
 438 sociated with vertical fluxes using these three methods, as another way of presenting the

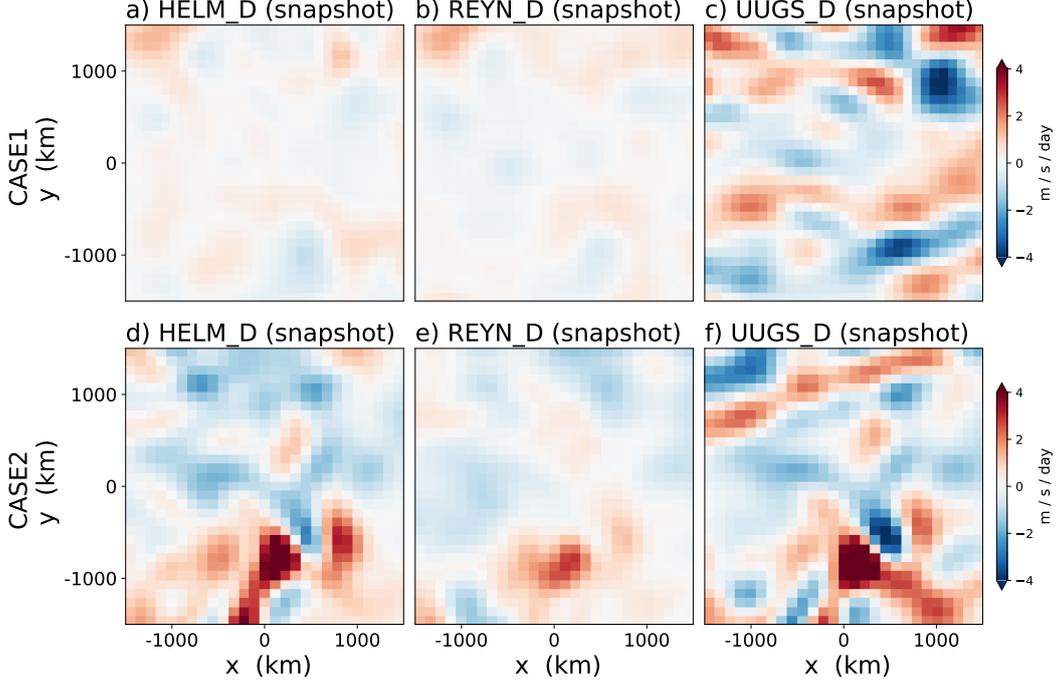


Figure 5. Snapshots of zonal SGS GWD due to vertical fluxes at 30 km height calculated using three methods. Upper row: Case 1. Lower row: Case 2. Left column, a & d: Helmholtz decomposition method $\frac{1}{\bar{\rho}} \frac{\partial \bar{\rho}'_H w'}{\partial z}$. Middle column, b & e: Reynolds stress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

439 differences among them. While the PDFs we obtain using the HELM_D and REYN_D
 440 methods are fairly similar, the PDFs from the UUGS method often have a clearly wider
 441 distribution. For both cases studied here, we find that the GWD from the UUGS method
 442 in general has higher variability, both temporal and spatial, compared with the other meth-
 443 ods, that are based on the Reynolds stress term alone (e.g., Eq. (14)), which is one of the
 444 three components of the total stress estimated in the UUGS method (see Appendix A). The
 445 degree of differences in PDFs depend on the case and height. For example, the differences
 446 are smaller at 40-50 km for Case 1, possibly due to the weak zonal winds there (Fig. 4).

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

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

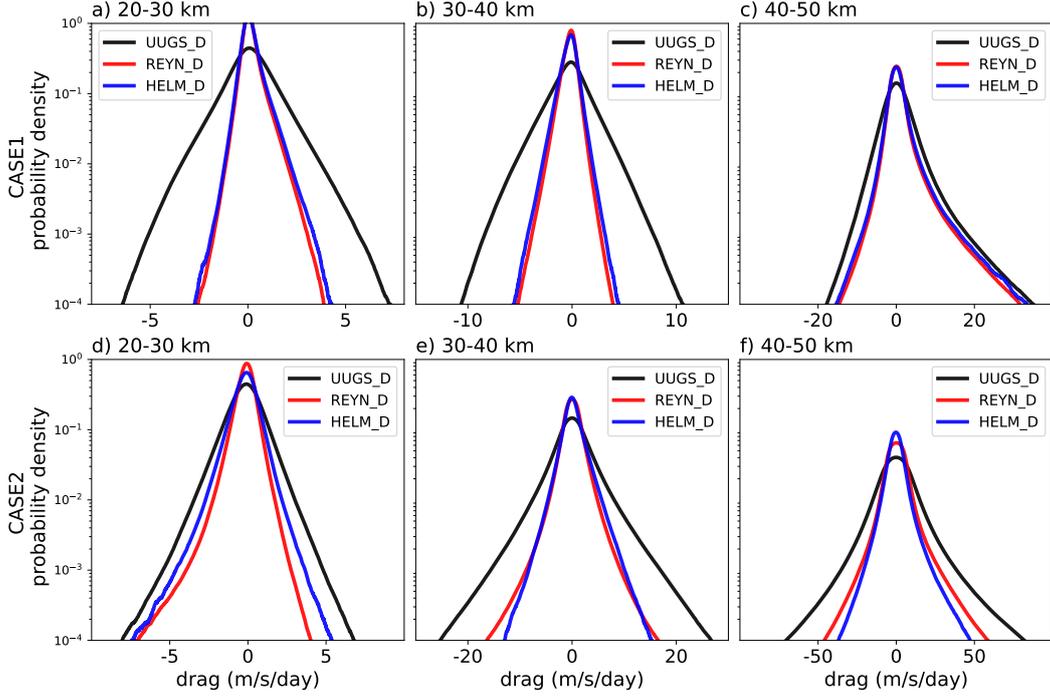


Figure 6. Probability density function (PDF) of zonal SGS GWD due to vertical fluxes calculated using three methods at different heights. Upper row: Case 1. Lower row: Case 2. Left column, a & d: 20-30 *km*. Middle column, b & e: 30-40 *km*. Right column, c & f: 40-50 *km*. Note the differences in the *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.

461

3.3 Horizontal flux of zonal momentum due to SGS GWs

462

463

464

465

466

467

468

469

470

471

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.

472

473

474

475

476

477

478

479

480

481

482

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

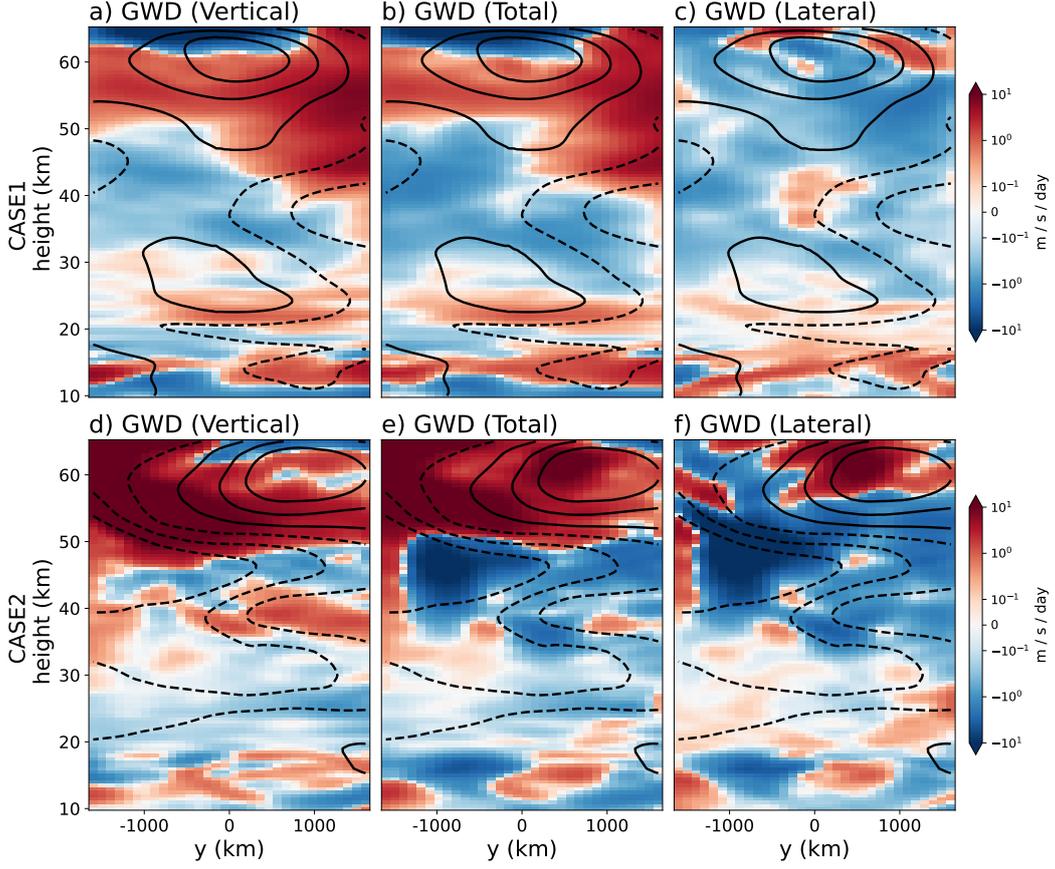


Figure 7. Zonal-mean, time-mean zonal SGS GWD due to only the vertical fluxes, total fluxes, and the lateral fluxes calculated using the UUGS_D method (Eq. (11)). Left column: Only the vertical component $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}]$. Middle column: All terms $-\frac{1}{\bar{\rho}} \frac{\partial}{\partial x} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial y} [\bar{\rho}(\overline{uv}) - \bar{u}\bar{v}] - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}(\overline{uw}) - \bar{u}\bar{w}]$. 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.

483 another example, in Case 2, at levels below the SAO (~ 50 km), it is clear that the lateral
 484 lateral momentum fluxes dominate the GWD there, even leading to a change of direction of the
 485 total zonal SGS GWD.

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

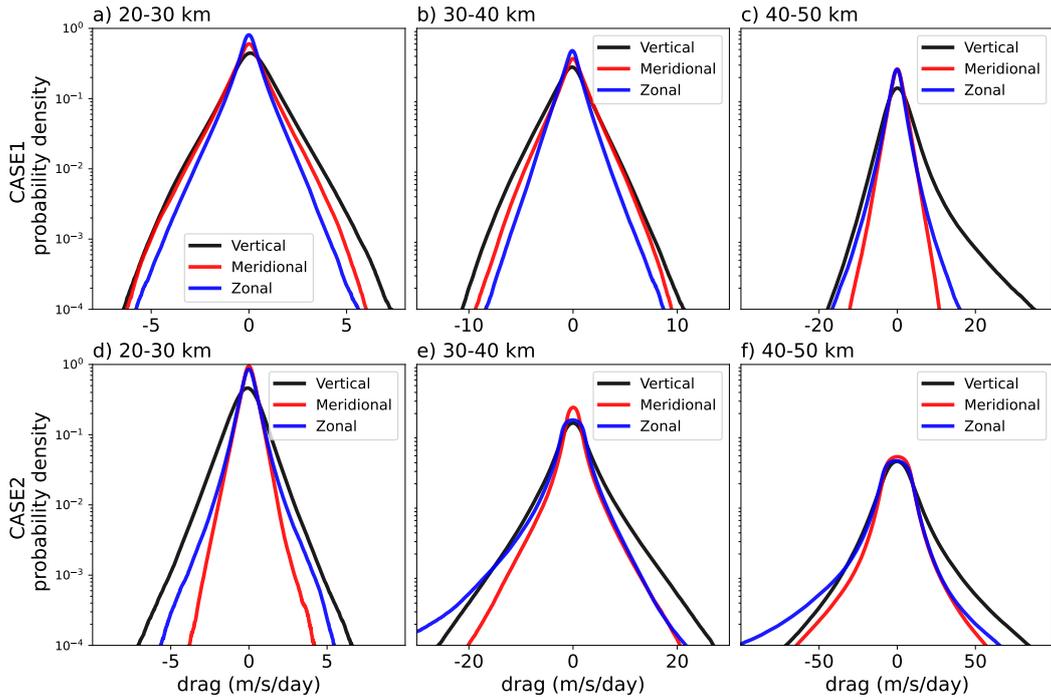


Figure 8. Probability density function (PDF) of zonal SGS GWD due to SGS zonal (blue), meridional (red), and vertical (black) momentum fluxes for both cases. Upper row: Case 1; Lower panel: Case 2. Left column, a & d: 20-30 km. Middle column, b & e: 30-40 km. Right column, c & f: 40-50 km. Note the difference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

497 So far, we have used the UUGS_D method to calculate the GWD due to SGS horizontal
 498 fluxes (Figs. 7-8). Compared to the vertical fluxes shown earlier, calculations of SGS GWD
 499 due to horizontal momentum fluxes have a much stronger sensitivity to the choice of the
 500 method. Figure 9 shows the time-mean, zonal mean SGS GWD associated with the meridional
 501 fluxes of the zonal momentum, calculated using HELM_D, REYN_D, and UUGS_D
 502 (the second term in Eq. (11)), respectively. Different from Fig. 4, the results here strongly
 503 depend on the method, even after averaging over time (simulation period) and space (zonal
 504 direction). This suggests that if we want to include the lateral propagation of GWs in the
 505 data-driven GWP schemes, then we must carefully examine the GWD extraction methodol-
 506 ogy. The PDFs in Fig. 10 show the same story. The SGS GWD induced by the lateral fluxes
 507 are much larger if calculated using the UUGS method compared to the other two (note the
 508 logarithmic color bar). It is clear that drag due to Reynolds stress is not the dominant term
 509 anymore when we consider GWD due to the SGS lateral fluxes. One explanation for this is
 510 that there are fundamental differences between the scales and amplitudes of the horizontal
 511 winds (u, v) and the vertical winds (w), as already shown in Fig. 2. The vertical velocity
 512 is dominated by small-scale features with negligible signal at the resolved scales in GCMs,
 513 which results in weak interactions between the resolved scales and the small scales. On the
 514 contrary, the horizontal winds are dominated by winds at the resolved scales, which means
 515 much stronger interactions between the resolved scales and the unresolved scales, and hence
 516 the large differences between UUGS_D and REYN_D.

517 Moreover, while Figs. 9 and 10 suggest similarities between the SGS GWD associated
 518 with the SGS horizontal fluxes calculated using the HELM_D and the REYN_D methods
 519 (Fig. 9a and 9b), substantial differences can exist even between the SGS GWD patterns
 520 extracted using these two methods. Figure 11 shows the correlation between instantaneous

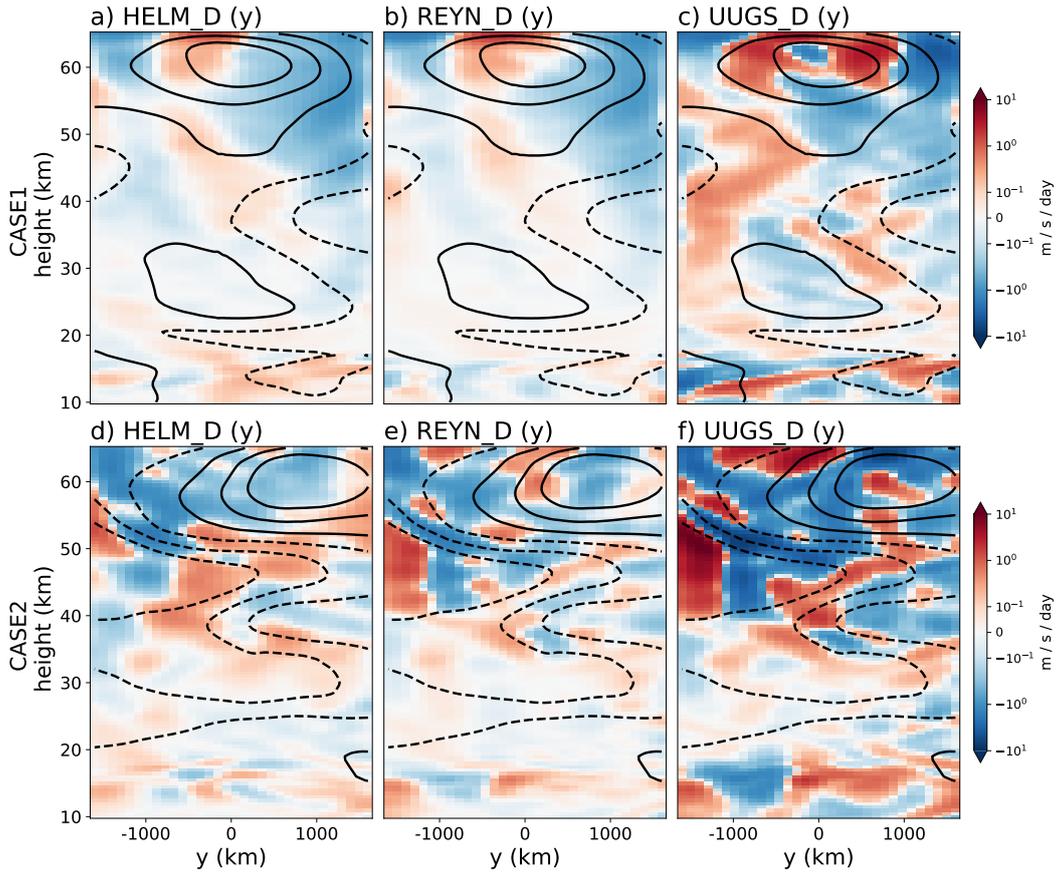


Figure 9. Zonal-mean, time-mean zonal SGS GWD, similar to Fig. 4, but for the GWD due to the un-/under-resolved (SGS) horizontal fluxes. Only the meridional direction is shown here; examining the zonal direction shows a similar story.

521 GWD calculated using the HELM_D and REYN_D method. For the SGS GWD due to
 522 vertical fluxes, as already discussed, there is a good match between these two methods.
 523 However, for SGS GWD due to the horizontal fluxes, the correlation is fairly weak, even
 524 though mathematically similar expressions are used for GWD in both methods. These
 525 results, again, show the high sensitivity of the lateral momentum fluxes and the resulting
 526 GWD to the details of the extraction method.

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

530 3.4 Sensitivity to the filter type/size and the GCM resolution

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

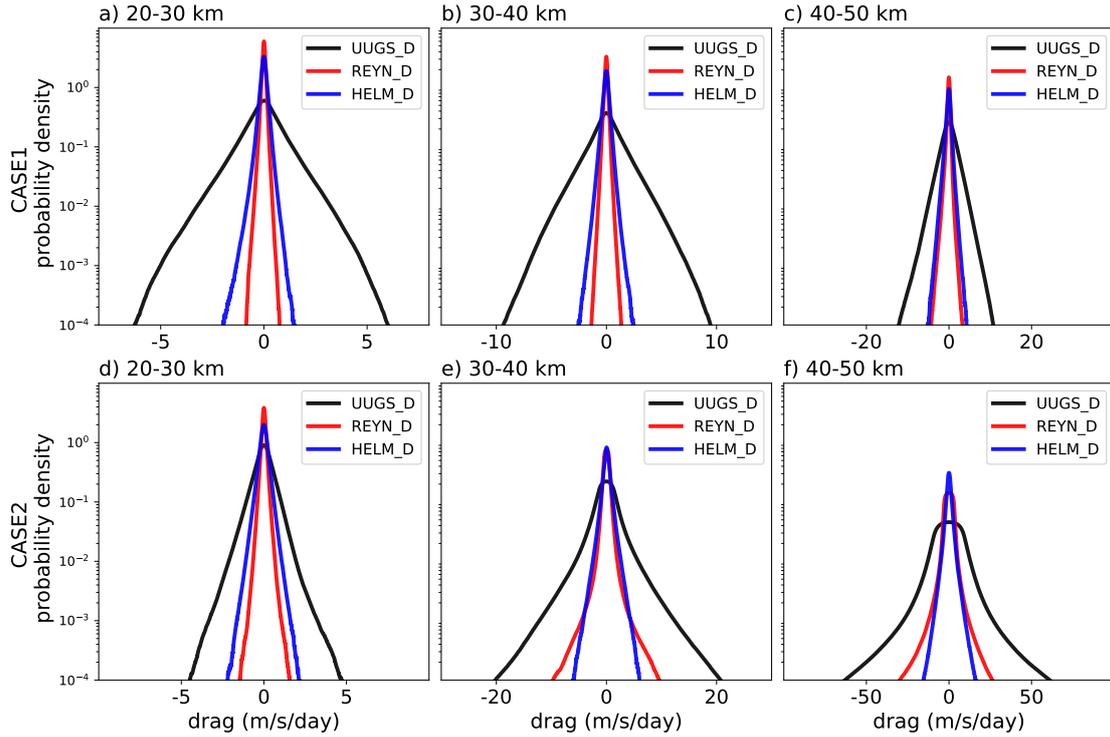


Figure 10. Same as Fig. 6, but for the un- and under-resolved zonal GWD due to SGS meridional momentum fluxes. Note the difference in the x -axes. See Fig. S4 for the PDFs of zonal GWD due to the SGS zonal momentum fluxes. See Fig. S5 for the PDFs of data from all 20 cases combined.

539 As mentioned earlier, Fig. B1 shows the power spectrum of the zonal winds before and
 540 after these low-pass filters are applied. This figure demonstrates the overall similarities
 541 between the outcome of the Gaussian and top-hat filters, at least up to the filtering scale, and
 542 major differences with the outcome of the sharp-spectral filter. Figure B2 shows snapshots
 543 of the SGS vertical momentum flux (Reynolds stress and total stress) extracted using these
 544 3 filters and $\Delta = 700 \text{ km}$. Again, we see that the Gaussian and top-hat filters overall
 545 yield fairly similar results. The outcomes of the sharp-spectral filter on the other hand,
 546 show differences in both amplitude and pattern, though the degree of difference is more
 547 pronounced for the Reynolds stress.

548 All the results shown so far are with filter size $\Delta = 700 \text{ km}$, coarse-grained to the GCM
 549 resolution of 100 km . However, this choice of 700 km is rather subjective, as there is no well-
 550 defined physical scale separation for GWs. Moreover, the appropriate filtering scale depends
 551 on the capability of a given GCM to resolve the GWs larger than the GCM's grid spacing,
 552 i.e., it depends on "effective resolution" of the GCM, which in turn depends on the GCM's
 553 numerical schemes and choices of grid-scale filters, like hyperdiffusion (e.g., Klaver et al.,
 554 2020). The $\Delta = 700 \text{ km}$ used here is based on studies showing that the effective resolution
 555 of WRF for GWs is 7 times the grid spacing (Skamarock, 2004). Admittedly, Δ should be
 556 chosen based on the effective resolution of the target GCM, not that of the GW-resolving
 557 model. To examine the sensitivity of the results to this choice, below we also present analysis
 558 with filtering scale that is twice the GCM grid spacing (i.e., $\Delta = 200 \text{ km}$), which is based
 559 on the LES literature (Pope, 2000; Sagaut, 2006; Guan et al., 2022). Furthermore, with the
 560 increase in computing power, some GCMs now have grid spacing of 0.5° or even smaller.
 561 Therefore, below, we also show results with for a GCM with the grid spacing of 30 km .

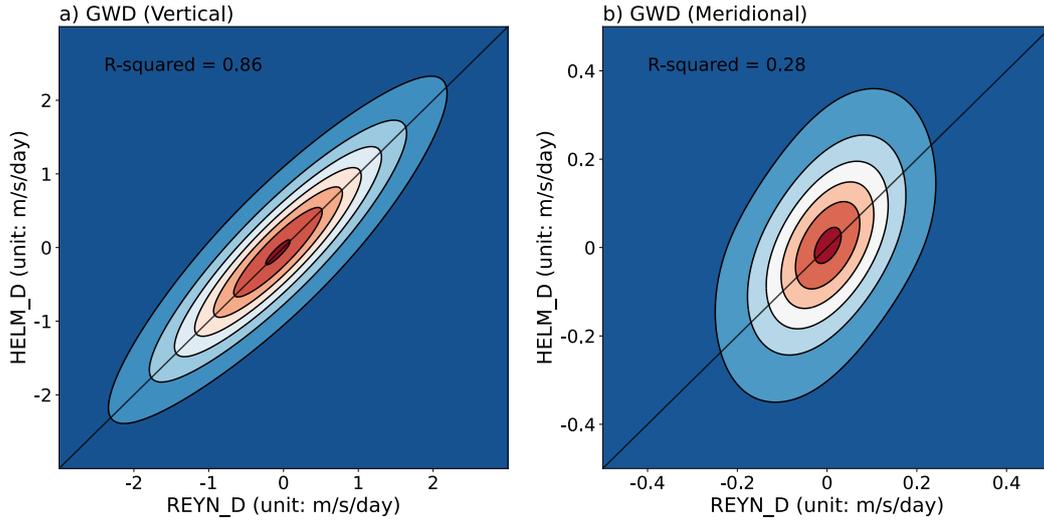


Figure 11. Joint PDFs between the SGS GWD calculated using the HELM.D and REYN.D methods for Case 1. a) GWD due to SGS vertical fluxes and b) GWD due to SGS meridional fluxes. The R-squared value is shown to measure the correlation between these two methods.

562 Figure 12 shows PDFs of the GWD from the SGS zonal, meridional, and vertical
 563 fluxes for 3 sets of choices: GCM grid spacing of 100 km and $\Delta = 700$ km (left column)
 564 and $\Delta = 200$ km (middle column), and GCM grid spacing of 30 km and $\Delta = 200$ km
 565 (right column). The second choice is meant to show the influences of effective resolution
 566 change while the third choice is meant to show what happens with higher-resolution GCMs.
 567 Although one might expect smaller SGS GW wind perturbations with reduced Δ , the zonal
 568 SGS GWD may not be necessarily reduced, as less averaging of the momentum flux is also
 569 applied with a smaller Δ . As a result, in both cases, for the zonal SGS GWD associated
 570 with vertical fluxes (solid lines), we find larger values when the filter length scale is reduced
 571 (compare the tails of the PDFs in the left and middle columns). The zonal SGS GWD
 572 associated with horizontal fluxes may become larger or smaller depending on the case.

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

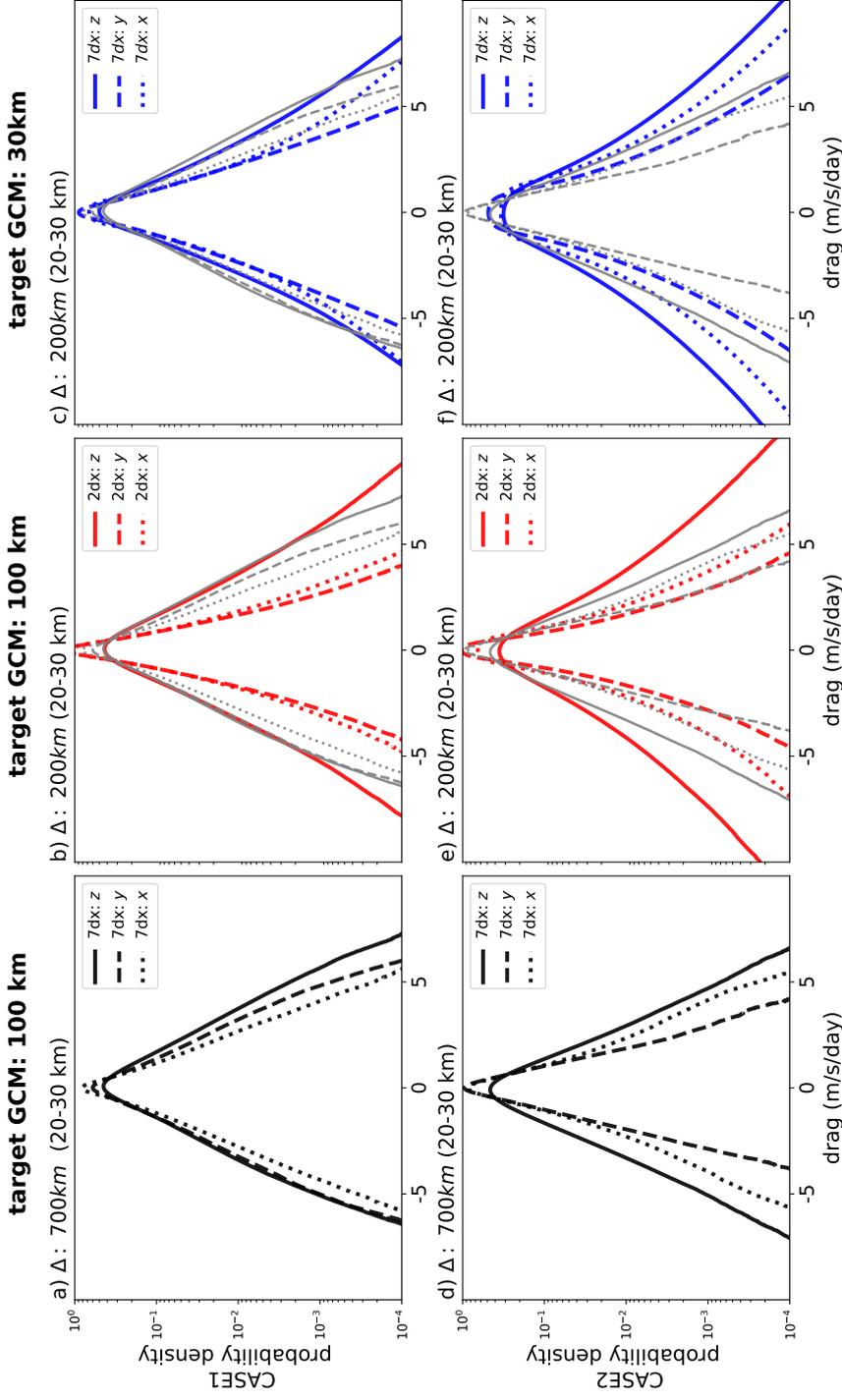


Figure 12. The PDFs of GWD at 20 – 30 km heights associated with the SGS zonal (dotted), meridional (dashed), and vertical (solid) fluxes extracted using the UUGS-D method (with Gaussian filter) for 3 different sets of filter size Δ and GCM grid spacing. Left column: $\Delta = 700$ km, 100 km grid spacing. 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 complexity, we have found that the GWD due to SGS lateral momentum fluxes could be sensitive to the methods used to extract them. Even the time-mean, zonal-mean GWD could be very different when different methods are used. The instantaneous GWD amplitudes from the UUGS_D method could be much

larger than those from the REYN_D method due to the strong cross-scale interactions between SGS GWs and the horizontal background flow that could be resolved by the GCMs. This suggests that to include the SGS GWD due to the horizontal fluxes in the data-driven GWP schemes, further research needs to be done on the extraction methodology.

4. The sensitivity of extracted SGS GWD to the length scale (Δ) of the filtering operation and the horizontal resolution of the GCM are also studied. Our results suggest that both the effective resolution of a GCM and its grid spacing have significant influences on the calculated SGS GWD (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

682 the GWs in the tropics (Jeevanjee, 2017; Bramberger et al., 2020). We have conducted
 683 limited simulations with 1 km and 3 km grid spacing at periods when observations from
 684 super-pressure balloons are available. We will use these simulations to examine the effect of
 685 resolution and validate the GW fluxes in the library. Also, currently our WRF simulations
 686 are limited to the tropics. In the future, we aim to extend the library to the mid-latitude
 687 and even polar regions. We also plan to conduct pseudo-global warming experiments to
 688 examine the response of the GWD to climate change and to expand the library to include
 689 training sets for transfer learning, such that generalizable data-driven GWP schemes could
 690 be developed (Guan et al., 2022; Subel et al., 2022).

691 Appendix A UUGS Drag Extraction

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

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

695 where (u, v, w) is the 3D wind fields; p is pressure; ρ is density; f is the Coriolis parameter;
 696 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:

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

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

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

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

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

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

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

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

To get an expression for \mathbf{X}_x^G , we can apply both spatial filtering and coarse-graining operators to each term in the original Eq. (A1), and assume that the operations are commutative¹, which means, e.g.,

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

¹This 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

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

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

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

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

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

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

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

Similarly, for the meridional momentum equation, we obtain

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

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

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

700 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.,

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

701 We see that the Reynolds stress is one of the three components of the total SGS flux. The
 702 total SGS flux accounts for interactions among all scales, including scales resolved by the
 703 GCMs with the un- and under-resolved scales. The importance of Leonard term and cross
 704 term has long been shown in studies of turbulent flows (e.g., Leonard, 1975; Galmarini et al.,
 705 2000). We also note that the Reynolds term here based on spatial filtering is different with
 706 the traditionally temporal-based Reynolds average in which the flow is decomposed into a
 707 mean and fluctuating components. As pointed in Aluie et al. (2018), the time-mean flow is
 708 not synonymous with large-scale flow, nor does a temporal fluctuation directly correspond
 709 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).

710 **Appendix B The effects of filter type**

711 Figures B1 and B2 show the spectra and snapshots of zonal wind and SGS vertical flux
712 when the 3 different filters are applied with the same Δ . Note that because our WRF regional
713 domain is not periodic, we have used mirrored tiles to reduce the boundary effects. Still,
714 Fourier-based filters such as the sharp-spectral filter might suffer from Gibbs oscillations
715 and give non-physical results.

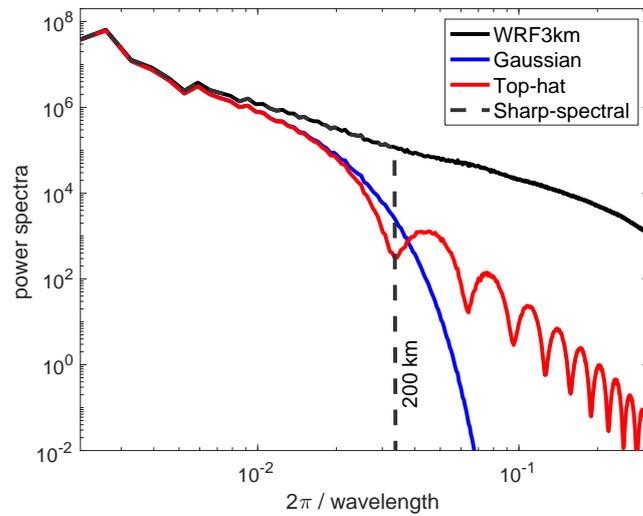


Figure B1. Example of the power spectrum of the zonal wind from the 3-*km* WRF simulations before and after the low-pass filtering using different kernels (with length scale $\Delta = 200$ *km*). The black line shows the spectrum before filtering, while the blue, red, and green lines show the the spectrum 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).

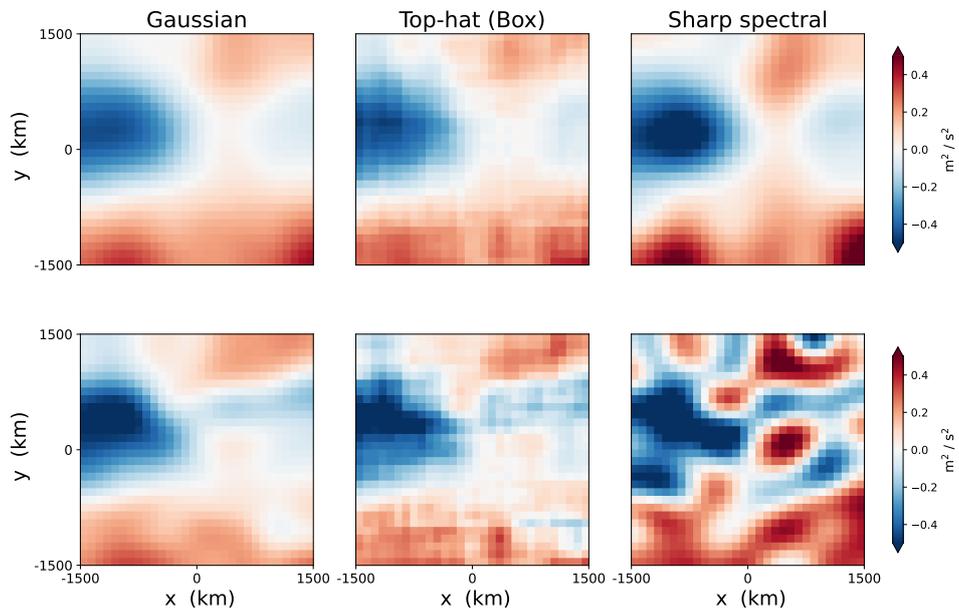


Figure B2. Snapshots of the SGS vertical momentum fluxes, calculated using the 3 different filters for Case 1 (at 40 km height). Top row: the Reynolds stress, $\overline{u'w'}$. Bottom: the total SGS stress, $\overline{uw} - \overline{u}\overline{w}$. The filter size is $\Delta = 700$ km.

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 CISE (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 climate models and the global distribution of gravity-wave momentum 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
- Amiranjadi, M., Plougonven, R., Mohebalhojeh, A. R., & Mirzaei, M. (2022). Using machine learning to estimate non-orographic gravity wave characteristics at source levels. *Journal of the Atmospheric Sciences*. 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
- Böloni, 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

- 766 Bölöni, G., Ribstein, B., Muraschko, J., Sgoff, C., Wei, J., & Achatz, U. (2016). The inter-
767 action between atmospheric gravity waves and large-scale flows: An efficient description
768 beyond the nonacceleration paradigm. *Journal of the Atmospheric Sciences*, *73*. doi:
769 10.1175/JAS-D-16-0069.1
- 770 Callies, J., Ferrari, R., & Bühler, O. (2014). Transition from geostrophic turbulence to
771 inertia-gravity waves in the atmospheric energy spectrum. *Proceedings of the National
772 Academy of Sciences of the United States of America*, *111*. doi: 10.1073/pnas.1410772111
- 773 Cao, J., Ran, L., & Li, N. (2014). An application of the helmholtz theorem in extracting
774 the externally induced deformation field from the total wind field in a limited domain.
775 *Monthly Weather Review*, *142*. doi: 10.1175/MWR-D-13-00311.1
- 776 Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Ma-
777 chine learning emulation of gravity wave drag in numerical weather forecasting. *Jour-
778 nal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002477. Retrieved
779 from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002477>
780 (e2021MS002477 2021MS002477) doi: <https://doi.org/10.1029/2021MS002477>
- 781 Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven super-
782 parameterization using deep learning: Experimentation with multiscale Lorenz 96 sys-
783 tems and transfer learning. *Journal of Advances in Modeling Earth Systems*, *12*(11),
784 e2020MS002084.
- 785 Chen, T.-C., & Wiin-Nielsen, A. C. (1976). On the kinetic energy of the divergent and
786 nondivergent flow in the atmosphere. *Tellus*, *28*. doi: 10.3402/tellusa.v28i6.11317
- 787 Clark, R. A., Ferziger, J. H., & Reynolds, W. C. (1979). Evaluation of subgrid-scale models
788 using an accurately simulated turbulent flow. *Journal of fluid mechanics*, *91*(1), 1–16.
- 789 Corcos, M., Hertzog, A., Plougonven, R., & Podglajen, A. (2021). Observation of gravity
790 waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical
791 Research: Atmospheres*, *126*. doi: 10.1029/2021JD035165
- 792 Dunbar, O. R., Garbuno-Inigo, A., Schneider, T., & Stuart, A. M. (2021). Calibration
793 and uncertainty quantification of convective parameters in an idealized gcm. *Journal of
794 Advances in Modeling Earth Systems*, *13*(9), e2020MS002454.
- 795 Eckermann, S. D., Ma, J., & Broutman, D. (2015). Effects of horizontal geometrical
796 spreading on the parameterization of orographic gravity wave drag. part i: Numer-
797 ical transform solutions. *Journal of the Atmospheric Sciences*, *72*(6), 2330 - 2347.
798 Retrieved from [https://journals.ametsoc.org/view/journals/atsc/72/6/jas-d-14-
799 -0147.1.xml](https://journals.ametsoc.org/view/journals/atsc/72/6/jas-d-14-0147.1.xml) doi: 10.1175/JAS-D-14-0147.1
- 800 Ehard, B., Kaifler, B., Dörnbrack, A., Preusse, P., Eckermann, S. D., Bramberger, M., ...
801 Rapp, M. (2017). Horizontal propagation of large-amplitude mountain waves into the
802 polar night jet. *Journal of Geophysical Research*, *122*. doi: 10.1002/2016JD025621
- 803 Espinosa, Z. I., Sheshadri, A., Cain, G. R., Gerber, E. P., & DallaSanta, K. J. (2022).
804 Machine Learning Gravity Wave Parameterization Generalizes to Capture the QBO and
805 Response to Increased CO₂. , *49*(8), e98174. doi: 10.1029/2022GL098174
- 806 Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., ... Williams,
807 D. N. (2016). Towards improved and more routine Earth system model evaluation in
808 CMIP. *Earth System Dynamics*, *7*(4), 813-830. doi: 10.5194/esd-7-813-2016
- 809 Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle
810 atmosphere. *Review of Geophysics*, *41*(1). doi: 10.1029/2001RG000106
- 811 Fritts, D. C., & Nastrom, G. D. (1992). Sources of Mesoscale Variability of Gravity Waves.
812 Part II: Frontal, Convective, and Jet Stream Excitation. *Journal of Atmospheric Sciences*,
813 *49*(2), 111-127. doi: 10.1175/1520-0469(1992)049<0111:SOMVOG>2.0.CO;2
- 814 Galmarini, S., Michelutti, F., & Thunis, P. (2000). Estimating the contribution of leonard
815 and cross terms to the subfilter scale from atmospheric measurements. *Journal of the
816 Atmospheric Sciences*, *57*, 2968-2976. doi: 10.1175/1520-0469(2000)057<2968:ETCOLA>
817 2.0.CO;2
- 818 Gentine, P., Eyring, V., & Beucler, T. (2021). Deep learning for the parametrization of sub-
819 grid processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive*

- 820 *Approach to Remote Sensing, Climate Science, and Geosciences*, 307–314.
- 821 Germano, M. (1986). A proposal for a redefinition of the turbulent stresses in the filtered
822 Navier-Stokes equations. *The Physics of fluids*, 29(7), 2323–2324.
- 823 Germano, M. (1992). Turbulence: the filtering approach. *Journal of Fluid Mechanics*, 238,
824 325–336.
- 825 Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R.,
826 ... Randel, W. J. (2019). The whole atmosphere community climate model version
827 6 (waccm6). *Journal of Geophysical Research: Atmospheres*, 124, 12380-12403. doi:
828 10.1029/2019JD030943
- 829 Grooms, I., Loose, N., Abernathy, R., Steinberg, J., Bachman, S. D., Marques, G., ...
830 Yankovsky, E. (2021). Diffusion-based smoothers for spatial filtering of gridded geophys-
831 ical data. *Journal of Advances in Modeling Earth Systems*, 13(9), e2021MS002552.
- 832 Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022). Stable a posteriori
833 LES of 2D turbulence using convolutional neural networks: Backscattering analysis and
834 generalization to higher Re via transfer learning. *Journal of Computational Physics*, 458,
835 111090.
- 836 Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2023). Learning physics-
837 constrained subgrid-scale closures in the small-data regime for stable and accurate les.
838 *Physica D: Nonlinear Phenomena*, 443, 133568. doi: [https://doi.org/10.1016/j.physd](https://doi.org/10.1016/j.physd.2022.133568)
839 .2022.133568
- 840 Hong, S.-Y., & Lim, J.-O. J. (2006). The wrf single-moment 6-class microphysics scheme
841 (wsm6). *Journal of the Korean Meteorological Society*, 42.
- 842 Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A New Vertical Diffusion Package with an
843 Explicit Treatment of Entrainment Processes. *Monthly Weather Review*, 134(9), 2318.
844 doi: 10.1175/MWR3199.1
- 845 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins,
846 W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the
847 aer radiative transfer models. *Journal of Geophysical Research Atmospheres*, 113. doi:
848 10.1029/2008JD009944
- 849 Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling*
850 *Earth Systems*, 9. doi: 10.1002/2017MS001059
- 851 Jiang, Q., Doyle, J. D., Eckermann, S. D., & Williams, B. P. (2019). Stratospheric trailing
852 gravity waves from new zealand. *Journal of the Atmospheric Sciences*, 76(6), 1565 - 1586.
853 Retrieved from [https://journals.ametsoc.org/view/journals/atasc/76/6/jas-d-18](https://journals.ametsoc.org/view/journals/atasc/76/6/jas-d-18-0290.1.xml)
854 -0290.1.xml doi: 10.1175/JAS-D-18-0290.1
- 855 Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-
856 Bustamante, E. (2012). A revised scheme for the wrf surface layer formulation. *Monthly*
857 *Weather Review*, 140. doi: 10.1175/MWR-D-11-00056.1
- 858 Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmailzadeh, S., ... others
859 (2021). Physics-informed machine learning: case studies for weather and climate mod-
860 elling. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200093.
- 861 Kim, Y., Eckermann, S. D., & Chun, H. (2003). An overview of the past, present and
862 future of gravity-wave drag parametrization for numerical climate and weather prediction
863 models. *Atmosphere-Ocean*, 41, 65-98. Retrieved from [https://doi.org/10.3137/ao](https://doi.org/10.3137/ao.410105)
864 .410105 doi: 10.3137/ao.410105
- 865 Kim, Y. H., Bölöni, G., Borchert, S., Chun, H. Y., & Achatz, U. (2021). Toward transient
866 subgrid-scale gravity wave representation in atmospheric models. part ii: Wave intermit-
867 tency simulated with convective sources. *Journal of the Atmospheric Sciences*, 78. doi:
868 10.1175/JAS-D-20-0066.1
- 869 Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high
870 resolution global atmospheric models for climate studies. *Atmospheric Science Letters*,
871 21. doi: 10.1002/asl.952
- 872 Kruse, C. G., Alexander, M. J., Hoffmann, L., Niekerk, A. V., Polichtchouk, I., Bacmeister,
873 J. T., ... Stein, O. (2022). Observed and modeled mountain waves from the surface to

- 874 the mesosphere near the drake passage. *Journal of the Atmospheric Sciences*, *79*. doi:
875 10.1175/JAS-D-21-0252.1
- 876 Kruse, C. G., & Smith, R. B. (2015). Gravity wave diagnostics and characteristics in
877 mesoscale fields. *Journal of the Atmospheric Sciences*, *72*. doi: 10.1175/JAS-D-15-0079.1
- 878 Kruse, C. G., Smith, R. B., & Eckermann, S. D. (2016). The midlatitude lower-stratospheric
879 mountain wave "valve layer". *Journal of the Atmospheric Sciences*, *73*. doi: 10.1175/
880 jas-d-16-0173.1
- 881 Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In
882 *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier.
- 883 Lott, F., Guez, L., & Maury, P. (2012). A stochastic parameterization of non-orographic
884 gravity waves: Formalism and impact on the equatorial stratosphere. *Geophysical Re-*
885 *search Letters*, *39*. doi: 10.1029/2012GL051001
- 886 Lynch, P. (1988). Deducing the wind from vorticity and divergence. *Monthly Weather*
887 *Review*, *116*. doi: 10.1175/1520-0493(1988)116<0086:DTWFVA>2.0.CO;2
- 888 Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the fidelity of
889 explainable artificial intelligence methods for applications of convolutional neural networks
890 in geoscience. *Artificial Intelligence for the Earth Systems*, *1*(4), e220012.
- 891 Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020).
892 Application of Deep Learning to Estimate Atmospheric Gravity Wave Parameters in Re-
893 analysis Data Sets. , *47*(19), e89436. doi: 10.1029/2020GL089436
- 894 Maulik, R., San, O., Rasheed, A., & Vedula, P. (2019). Subgrid modelling for two-
895 dimensional turbulence using neural networks. *Journal of Fluid Mechanics*, *858*, 122–
896 144.
- 897 McFarlane, N. A. (1987). The Effect of Orographically Excited Gravity Wave Drag on the
898 General Circulation of the Lower Stratosphere and Troposphere. *Journal of Atmospheric*
899 *Sciences*, *44*(14), 1775–1800. doi: 10.1175/1520-0469(1987)044<1775:TEOOEG>2.0.CO;2
- 900 Muraschko, J., Fruman, M. D., Achatz, U., Hickel, S., & Toledo, Y. (2015). On the
901 application of wenzel-kramer-brillouin theory for the simulation of the weakly nonlinear
902 dynamics of gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *141*.
903 doi: 10.1002/qj.2381
- 904 Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblus-Reyes, F. J., Jung, T., & Leutbecher, M.
905 (2005). Representing model uncertainty in weather and climate prediction. *Annual Review*
906 *of Earth and Planetary Sciences*, *33*. doi: 10.1146/annurev.earth.33.092203.122552
- 907 Palmer, T. N., Shutts, G. J., & Swinbank, R. (1986). Alleviation of a systematic westerly
908 bias in general circulation and numerical weather prediction models through an orographic
909 gravity wave drag parametrization. *Quarterly Journal of the Royal Meteorological Society*,
910 *112*. doi: 10.1002/qj.49711247406
- 911 Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., . . .
912 others (2022). Fourcastnet: A global data-driven high-resolution weather model using
913 adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- 914 Pincus, R., Barker, H. W., & Morcrette, J. J. (2003). A fast, flexible, approximate technique
915 for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical*
916 *Research: Atmospheres*, *108*. doi: 10.1029/2002jd003322
- 917 Plougonven, R., & Zhang, F. (2014). Internal gravity waves from atmospheric jets and
918 fronts. *Reviews of Geophysics*, *52*. doi: 10.1002/2012RG000419
- 919 Polichtchouk, I., Wedi, N., & Kim, Y. H. (2022). Resolved gravity waves in the tropi-
920 cal stratosphere: Impact of horizontal resolution and deep convection parametrization.
921 *Quarterly Journal of the Royal Meteorological Society*, *148*. doi: 10.1002/qj.4202
- 922 Pope, S. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*,
923 *72*(2), 331–340.
- 924 Pope, S. (2000). *Turbulent flows*. Cambridge university press.
- 925 Qiao, W., Song, J., He, H., & Li, F. (2019). Application of different wind field models
926 and wave boundary layer model to typhoon waves numerical simulation in wavewatch
927 iii model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *71*. doi: 10.1080/

- 16000870.2019.1657552
- 928 Rai, S., Hecht, M., Maltrud, M., & Aluie, H. (2021). Scale of oceanic eddy killing by wind
929 from global satellite observations. *Science Advances*, 7(28), eabf4920.
- 930 Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid
931 processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39),
932 9684–9689.
- 933 Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simp-
934 son, I. R. (2020, 4). Progress in simulating the quasi-biennial oscillation in cmip models.
935 *Journal of Geophysical Research: Atmospheres*, 125. doi: 10.1029/2019JD032362
- 936 Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., . . . Yukimoto,
937 S. (2022, 4). Response of the quasi-biennial oscillation to a warming climate in global
938 climate models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1490-1518.
939 doi: 10.1002/qj.3749
- 940 Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Toward a physically based gravity
941 wave source parameterization in a general circulation model. *Journal of the Atmospheric*
942 *Sciences*, 67. doi: 10.1175/2009JAS3112.1
- 943 Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Springer
944 Science & Business Media.
- 945 Sato, K., Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., & Takahashi, M. (2009).
946 On the origins of mesospheric gravity waves. *Geophysical Research Letters*, 36. doi:
947 10.1029/2009GL039908
- 948 Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A
949 blueprint for models that learn from observations and targeted high-resolution simula-
950 tions. *Geophysical Research Letters*, 44(24), 12–396.
- 951 Scinocca, J. F., & McFarlane, N. A. (2000). The parametrization of drag induced by
952 stratified flow over anisotropic orography. *Quarterly Journal of the Royal Meteorological*
953 *Society*, 126. doi: 10.1002/qj.49712656802
- 954 Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction sys-
955 tems. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric*
956 *sciences, applied meteorology and physical oceanography*, 131(612), 3079–3102.
- 957 Shutts, G. J., & Vosper, S. B. (2011). Stratospheric gravity waves revealed in nwp model
958 forecasts. *Quarterly Journal of the Royal Meteorological Society*, 137. doi: 10.1002/
959 qj.763
- 960 Sigmond, M., & Scinocca, J. F. (2010). The influence of the basic state on the northern
961 hemisphere circulation response to climate change. *Journal of Climate*, 23(6), 1434–1446.
- 962 Skamarock, W. C. (2004). Evaluating mesoscale nwp models using kinetic energy spectra.
963 *Monthly Weather Review*, 132. doi: 10.1175/MWR2830.1
- 964 Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model
965 for weather research and forecasting applications. *Journal of Computational Physics*, 227.
966 doi: 10.1016/j.jcp.2007.01.037
- 967 Skamarock, W. C., Snyder, C., Klemp, J. B., & Park, S. H. (2019). Vertical resolution
968 requirements in atmospheric simulation. *Monthly Weather Review*, 147. doi: 10.1175/
969 MWR-D-19-0043.1
- 970 Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., . . . Ziemer,
971 F. (2022, 4). Atmospheric energy spectra in global kilometre-scale models. *Tellus A:*
972 *Dynamic Meteorology and Oceanography*, 74, 280-299. doi: 10.16993/tellusa.26
- 973 Stephan, C. C., Strube, C., Klocke, D., Ern, M., Hoffmann, L., Preusse, P., & Schmidt, H.
974 (2019). Intercomparison of gravity waves in global convection-permitting models. *Journal*
975 *of the Atmospheric Sciences*, 76. doi: 10.1175/JAS-D-19-0040.1
- 976 Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2022). Explaining the physics
977 of transfer learning in data-driven turbulence modeling. *arXiv preprint arXiv:2206.03198*.
- 978 Sun, Y. Q., & Zhang, F. (2016). Intrinsic versus practical limits of atmospheric predictability
979 and the significance of the butterfly effect. *Journal of the Atmospheric Sciences*, 73(3),
980 1419 - 1438. Retrieved from <https://journals.ametsoc.org/view/journals/atsc/>
981

- 982 73/3/jas-d-15-0142.1.xml doi: 10.1175/JAS-D-15-0142.1
- 983 Wang, P., Yuval, J., & O’Gorman, P. A. (2022). Non-local parameterization of atmospheric
984 subgrid processes with neural networks. *arXiv preprint arXiv:2201.00417*. doi: 10.48550/
985 ARXIV.2201.00417
- 986 Wei, J., Bölöni, G., & Achatz, U. (2019). Efficient modeling of the interaction of
987 mesoscale gravity waves with unbalanced large-scale flows: Pseudomomentum-flux con-
988 vergence versus direct approach. *Journal of the Atmospheric Sciences*, *76*. doi: 10.1175/
989 JAS-D-18-0337.1
- 990 Wei, J., Zhang, F., Richter, J. H., Alexander, M. J., & Sun, Y. Q. (2022). Global dis-
991 tributions of tropospheric and stratospheric gravity wave momentum fluxes resolved
992 by the 9-km ecmwf experiments. *Journal of the Atmospheric Sciences*, *79*. doi:
993 10.1175/JAS-D-21-0173.1
- 994 Yuval, J., & O’Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid
995 processes for climate modeling at a range of resolutions. *Nature communications*, *11*(1),
996 1–10.
- 997 Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures.
998 *Geophysical Research Letters*, *47*(17), e2020GL088376.
- 999 Zanna, L., & Bolton, T. (2021). Deep learning of unresolved turbulent ocean processes in
1000 climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to*
1001 *Remote Sensing, Climate Science, and Geosciences*, 298–306.
- 1002 Zhou, Z., He, G., Wang, S., & Jin, G. (2019). Subgrid-scale model for large-eddy simulation
1003 of isotropic turbulent flows using an artificial neural network. *Computers & Fluids*, *195*,
1004 104319.
- 1005 Žagar, N., Kasahara, A., Terasaki, K., Tribbia, J., & Tanaka, H. (2015). Normal-mode
1006 function representation of global 3-d data sets: Open-access software for the atmospheric
1007 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 Sun¹, Pedram Hassanzadeh¹, M. Joan Alexander², and Christopher

G. Kruse²

¹Rice University, Houston, Texas, United States

²NorthWest 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.

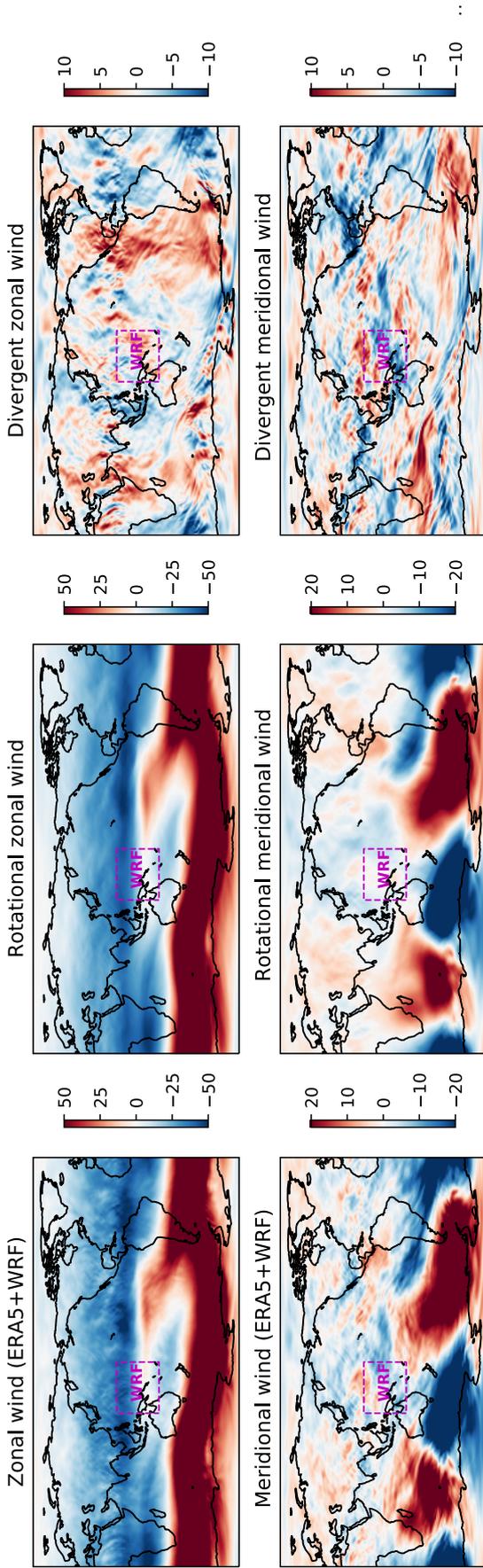


Figure S1. Snapshots from the Helmholtz decomposition approach used in this study.

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>

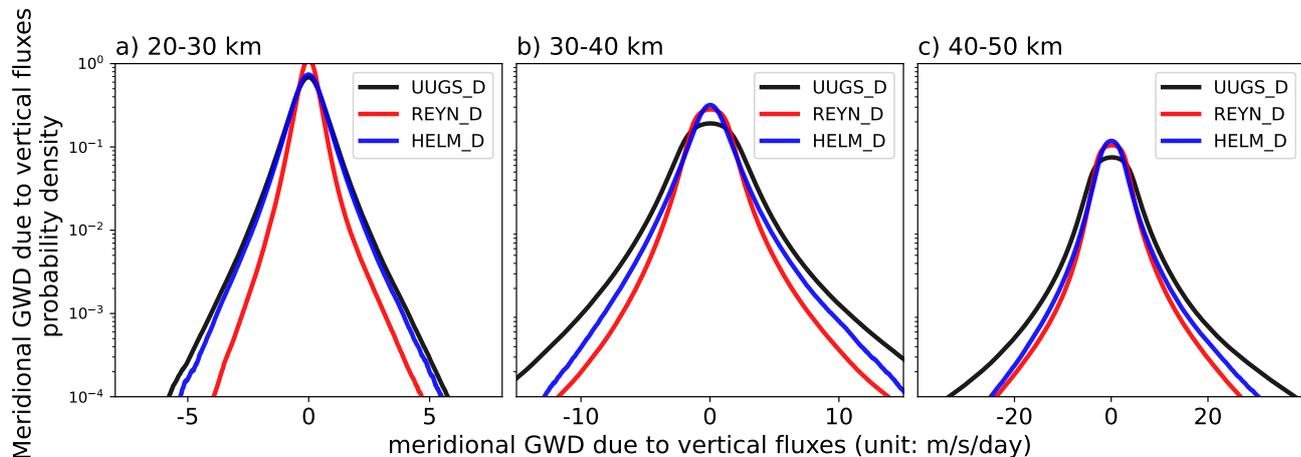


Figure S2. Same as Fig. 6, but for the meridional GWD due to vertical fluxes (last term of Eq. (A11)). Data from Cases 1 and 2 have been combined in making these PDFs.

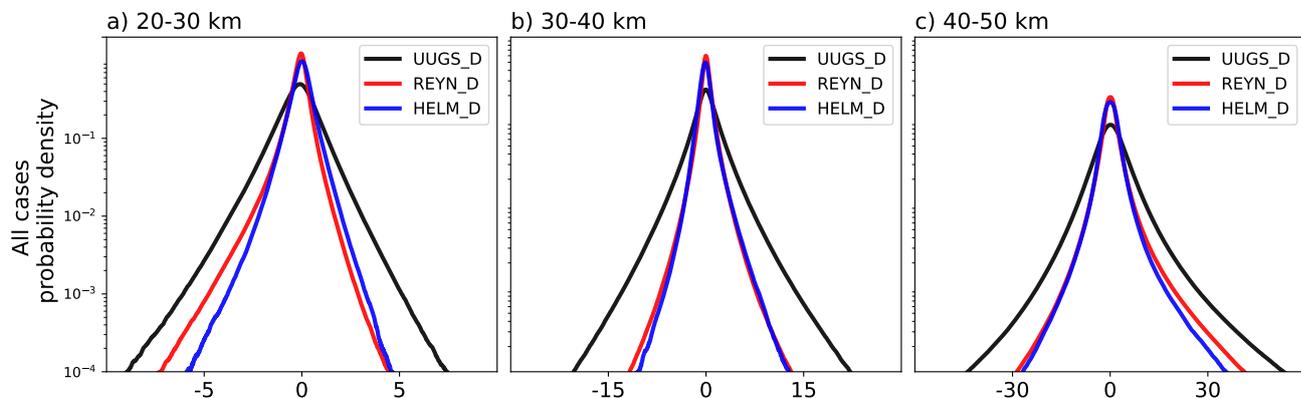


Figure S3. Same as Fig. 6, but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computational cost of the algorithm.

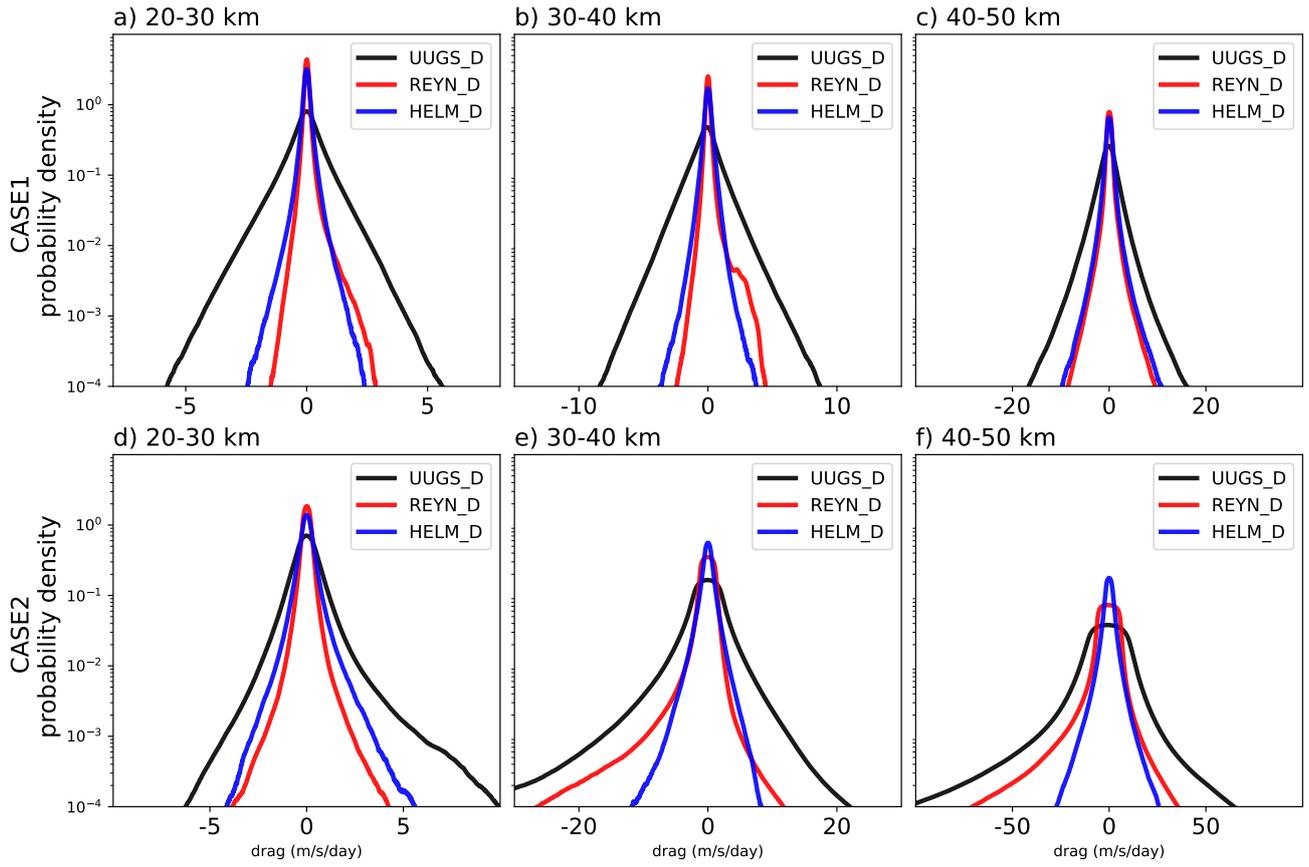


Figure S4. Same as Fig. 10, but for the zonal fluxes.

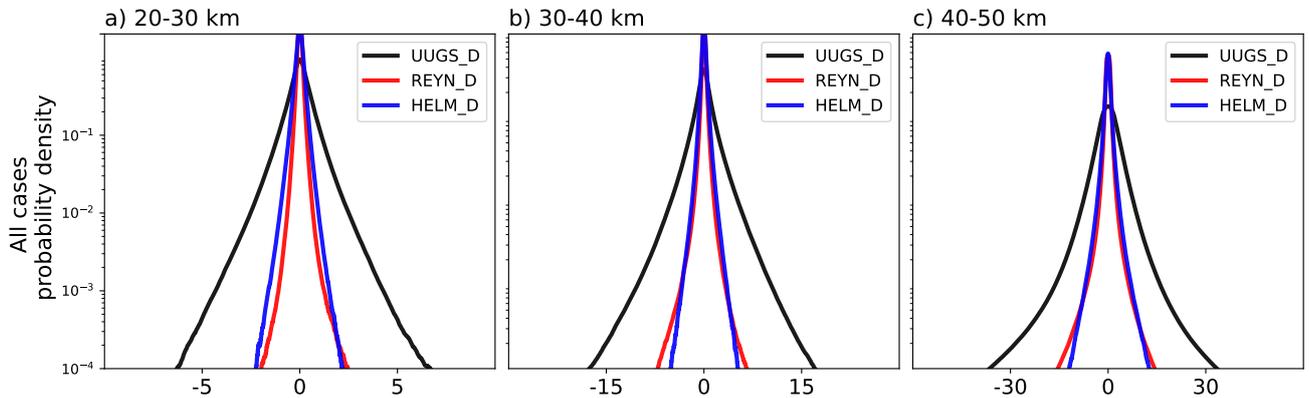


Figure S5. Same as Fig. 10 but for more cases. The UUGS_D and REYN_D results are based on data from all 20 cases combined. For HELM_D, data from only 4 cases have been due to the high computational cost of the algorithm.