NOAA Technical Memorandum OAR GSL-74

https://doi.org/10.25923/v2kx-8e10

A Description of the Unified Gravity Wave Physics Package

April 2025

Michael D. Toy, Joseph B. Olson, Fanglin Yang, Bo Yang and Song-You Hong

National Oceanic and Atmospheric Administration
Oceanic and Atmospheric Research
Global Systems Laboratory
Boulder, Colorado
April 2025

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

A Technical Memorandum OAR GSL-74

A Description of the Unified Gravity Wave Physics Package

Michael D. Toy[1,2]
Joseph B. Olson[1]
Fanglin Yang[3]
Bo Yang[3,4]
Song-You Hong[2,5]

[1]National Oceanic and Atmospheric Administration, Global Systems Laboratory
[2]Cooperative Institute for Research in Environmental Sciences, CU Boulder
[3]National Oceanic and Atmospheric Administration, Environmental Modeling Center
[4]Lynker Corporation at Environmental Modeling Center, NCEP
[5]National Oceanic and Atmospheric Administration, Physical Sciences Laboratory

Acknowledgements

Funding for this work was provided by many sources over several years, each helping to develop different aspects/components of the orographic drag scheme. These agencies/programs include NOAA's Atmospheric Science for Renewable Energy (ASRE) program, the Federal Aviation Administration (FAA), and NOAA's Next Generation Global Prediction System (NGGPS). The views expressed are those of the authors and do not necessarily represent the official policy or position of any funding agency. We are grateful to the CCPP developers (especially Dominikus Heinzeller) for help integrating, testing, and maintaining this code for use in the Unified Forecast System. We also thank Valery A. Yudin, Jongil Han and Sajal K. Kar for their contributions to this work and for their helpful comments on the manuscript, as well as Jaymes Kenyon for early testing and verification work. Funding for this project was provided by the Joint Technology Transfer Initiative within the Weather Program Office (NOAA Research and Development Database ID number 22380). This work was also supported in part by NOAA cooperative agreements NA17OAR4320101 and NA22OAR4320151.

UNITED STATES DEPARTMENT OF COMMERCE
Howard Lutnick, Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
Laura Grimm, NOAA Administrator

Office of Oceanic and Atmospheric Research
Dr. Steven Thur, Assistant Administrator

Contents

1. Introduction

Topography exerts an opposing drag force on atmospheric flow via three main processes: 1) gravity waves that propagate upward and break at various levels, 2) blocking at low levels if the flow has insufficient kinetic energy to make it over the mountains, but must instead go around them, and 3) turbulent form drag in the planetary boundary layer (PBL) caused by turbulent pressure perturbations induced by the terrain. The first process, gravity wave drag (GWD), can be resolved by the dynamical cores of atmospheric models as long as the spatial resolution, both horizontally and vertically, are sufficient to resolve the gravity waves. GWD due to unresolved, subgrid scale terrain must be parameterized. The free atmosphere supports orographic gravity waves over a wide range of spatial scales from ~5 km to 100's of km in the horizontal, commensurate with the scales of the topography that generate them. This presents a challenge to the development of orographic GWD (OGWD) parameterizations due to this wide "gray scale", where GWD is partly resolved and partly parameterized (Sandu et al. 2019). Representing the "missing drag" from subgrid scale GWD is essential to accurately forecasting the zonal circulation in global models and alleviating the high westerly windspeed biases and associated "cold pole" problems that develop without parameterized GWD (Kim et al. 2003).

Topography with horizontal scales as small as ~1 km can support vertically propagating gravity waves under very stable conditions, such as those associated with nocturnal PBLs. These waves often break at the PBL top and impart their momentum at this level, causing a drag force. Recently, parameterizations that model this effect have been developed (e.g., Steeneveld et al. 2008; Tsiringakis et al. 2017), and these help to improve surface windspeed biases, particularly at night. Such small-scale terrain variations also exert form drag on the PBL winds. This is not to be confused with the tangential drag force calculated by surface layer parameterizations, but rather is the force normal to the terrain surface induced by turbulent pressure perturbations. These develop as a result of the terrain in such a way that the terrain slope and perturbation pressure become positively correlated, resulting in an opposing drag force (e.g., Beljaars et al. 2004). Finally, the parameterization of low-level flow blocking by subgrid-scale terrain is essential for accurate predictions of the near-surface winds, and has been deployed in Numerical Weather Prediction (NWP) models for a number of decades (e.g., Lott and Miller 1997; Kim and Doyle 2005).

Non-orographic, subgrid-scale gravity wave sources must also be parameterized to ensure realistic forecasts of winds in the middle atmosphere. These sources include deep convection, frontal instability, and stratified shear instability associated with the tropospheric jet. NOAA's NWP models have integrated parameterizations for these non-stationary gravity wave drag effects.

All of these drag-related impacts on the atmospheric flow are meant to be represented within a single physics parameterization suite for use in operational weather forecasting. This memorandum documents each component of the Unified Gravity Wave Physics (UGWP) Suite that is being developed within the framework of NOAA's Unified Forecast System (UFS). Examples of performance are included in this document. Related information, such as proper generation of the necessary static fields required to use the UGWP suite and configuration of the suite within the context of the Common Community Physics Package (CCPP) is also provided.

2. UGWP components

The UGWP suite is a set of physical parameterizations combined in a single CCPP physics module (see Heinzeller et al. (2022) and Zhang et al. (2022) for descriptions of the CCPP). The schemes include the non-stationary gravity wave drag parameterization used in the GFSv16, and the Global Systems Laboratory (GSL) orographic drag suite, which consists of the four orographic physical parameterizations, described in the Introduction. The GSL orographic drag suite is used in the operational RAPv5 and HRRRv4 models at NCEP, and is being tested in the latest prototype for the next version of the FV3GFS (version 17). The following subsections describe the highlights of each drag scheme, and point out changes we have made to them from their originally published forms. The references should be consulted for full details of the schemes.

In the following subsections, the momentum flux, or wave stress, $\boldsymbol{\tau}$, will be discussed. The tendency of the velocity in each of the *x* and *y* directions are given by the vertical divergence of the stress from each component, which can be written

$$\left(\frac{\partial u}{\partial t}\right)_{drag} = -\frac{1}{\rho}\frac{\partial \tau_x}{\partial z}, \qquad \left(\frac{\partial v}{\partial t}\right)_{drag} = -\frac{1}{\rho}\frac{\partial \tau_y}{\partial z}. \tag{1}$$

where ρ is the air density and τ_x and τ_y are the x- and y-components of the wave stress, respectively.

2.1 Mesoscale gravity wave drag parameterization

The OGWD parameterization developed by Kim and Arakawa (1995), and later modified by Kim and Doyle (2005, hereafter KD05) and Choi and Hong (2015), calculates the drag from gravity waves forced by flow over subgrid-scale topography. Gravity waves propagate vertically and amplify with height, as the atmospheric density decreases, until they reach a level at which they overturn and break. The scheme determines the heights at which these waves break and deposits momentum at the corresponding model levels, producing a drag force in the direction opposite to the low-level flow direction. Below the wave-breaking levels, the wave vertical flux of horizontal momentum is constant with height (Holton 2004), and there is no drag force. The horizontal wavelengths at which mesoscale gravity waves may propagate vertically as internal waves in the free atmosphere ranges from 100's of kilometers down to approximately 5 km (Beljaars 2004). For topographic variations smaller than ~5 km, the waves that develop are typically evanescent, and do not transfer momentum or energy upward (Holton 2004).

The first step of the algorithm is to calculate the subgrid GWD stress at a reference level h_{REF} above the surface, and to propagate this wave momentum flux upward model layer by model layer, using the resolved model sounding of the column, until a level is found at which the wave induces static instability. It is assumed that the wave begins to break at this level, and the momentum flux is reduced before passing upward to the next level. The "wave saturation" hypothesis of Lindzen (1981) provides the theoretical background for the momentum deposition calculation. Note that if no level is found at which the wave would break, then the GWD parameterization produces no drag in the column.

The reference level where the GW flux originates is defined as $h_{\mathrm{REF}} = \max(2\sigma_h, h_{\mathrm{PBL}})$, where σ_h is the standard deviation of the subgrid topography, and h_{PBL} is the PBL depth. Following KD05, the GWD stress at the reference level is given by

$$\tau_{GWD} = \rho_0 E \frac{m}{\lambda_{eff}} G \frac{|U_0|^3}{N_0}, \tag{2}$$

$$E \equiv (OA+2)^{\frac{C_E Fr_0}{Fr_c}}, \quad m \equiv (1+L_x)^{OA+1}, \qquad G \equiv \frac{Fr_0^2}{Fr_0^2 + C_G OC^{-1}}, \tag{3}$$

where ρ_0 is the low-level air density, N is the Brunt-Väisälä frequency, U_0 is the low-level horizontal wind speed, m is the "number of mountains", OA is a measure of the asymmetry of the subgrid topography (defined in KD05), L_x is the "effective orographic length" of the subgrid topography (defined in KD05), OC is the "orographic convexity", which is the kurtosis of the subgrid topography (a measure of the sharpness of the mountains, defined in KD05), and $C_E = 0.8$ *and* $C_G = 0.5$, which were calibrated with mesoscale simulations in Kim and Arakawa (1995). The Froude numbers in (3), which are actually inverse Froude numbers as per KD05, are expressed as

$$Fr_0 = h \frac{N_0}{U_0} OD, \tag{4}$$

where h=2σ_h , and OD is the "orographic direction" representing the orographic anisotropy (defined in KD05), and Fr_c (≈0.8) is a prescribed critical Froude number. Finally, λ_{eff} is the "effective grid length", which is basically used as a tuning coefficient.

The subgrid wave stress is then propagated upward, and using the resolved atmospheric state that the wave encounters at each level in the column, it is determined whether the wave results in instability and breaking. Following Palmer et al. (1986) and Pierrehumbert (1986), the subgrid wave-induced vertical displacement height (h_d) of a fluid parcel at model level k, given the momentum flux τ at the level below (k-1), is given by

$$(h_d^2)_k = \frac{\Delta x}{m} \frac{\tau_{k-1}}{\rho_k N_k U_k}, \tag{5}$$

where Δx is the horizontal grid spacing, and $N_k = (g/\theta_v)_k(\partial\theta_v/\partial z)_k$, where θ_v is the virtual potential temperature. The value of h_d is then used to calculate the minimum Richardson number defined by

$$Ri_m = \frac{Ri(1-Fr_d)}{\left(1+\sqrt{Ri}Fr_d\right)^2}, \tag{6}$$

where $Fr_d \equiv h_d N/U$, and

$$Ri = \frac{\frac{g}{\theta}\frac{\partial\theta}{\partial z}}{\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2}. \tag{7}$$

Note that the horizontal wind speed U at all levels is that which is projected onto the low-level wind direction, i.e., where the wave is generated. The critical Richarson number, below which it is assumed there is subgrid-scale instability, is given by $Ri_c \simeq 0.25$. If $Ri_m > Ri_c$, there is no instability and τ is unchanged for the next model level. If $Ri_m \leq Ri_c$, then, following the wave saturation hypothesis of Lindzen (1981), the vertical wave displacement is limited to a critical value calculated by substituting the critical Richardson number ($Ri_c \simeq 0.25$) into (6) to give

$$(h_d)_c = \frac{U}{N}\left[2\left(2+\frac{1}{\sqrt{Ri}}\right)^{\frac{1}{2}} - 2\left(2+\frac{1}{\sqrt{Ri}}\right)\right]. \tag{8}$$

This value is used to calculate the wave stress τ at the next model level using (5) with the level indices advanced to k +1.

There are basically two outcomes from the above algorithm: 1) a critical level is met, i.e., where $U \lesssim 0$, and the wave breaks and deposits its momentum at that level, or 2) no critical level is met and there is no GWD in the column (this may occur due to increasing U with height, in which the wave can no longer propagate and becomes evanescent).

We have added a third outcome to the above algorithm by introducing a new step that prevents unlimited growth of the calculated surface stress with increasing near-surface wind speed. From linear theory (Holton 2004), when the near-surface wind speed U_0 is greater than N_0/k_S, where k_S is the topographic horizontal wave number, the gravity waves are evanescent, that is, their amplitude decays with height and they do not transfer momentum or energy to the mean flow. To test for this condition we assume a maximum subgrid mountain wavelength equal to half the horizontal grid spacing to calculate k_S. The OGWD is set to zero in the column when the condition is met.

The model code we use for the OGWD scheme originated from the Weather Research and Forecasting (WRF) model (Skamarock et al. 2019). We modified the scheme to be scale aware at the gray zone horizontal resolutions. The effects of the horizontal grid resolution on the strength of the parameterized GWD is accounted for in two ways: 1) the larger the grid box area, the larger the sample of subgrid topography within each box, and the larger the standard deviation σ_h, and 2) we multiply the surface stress calculated in (2) by the following coefficient, which tapers the subgrid stress when the horizontal resolution falls within the gray scale for subgrid GWD:

$$C_{\text{taper}} = \begin{cases} 1, & \text{for } \Delta x \geq (\Delta x)_{\max}; \\ \frac{1}{2}\left(\sin\left\{\pi \frac{\Delta x - \frac{1}{2}[(\Delta x)_{\max} + (\Delta x)_{\min}]}{(\Delta x)_{\max} - (\Delta x)_{\min}}\right\} + 1\right), & \text{for } (\Delta x)_{\max} > \Delta x \geq (\Delta x)_{\min}; \\ 0, & \text{for } \Delta x < (\Delta x)_{\min}; \end{cases} \qquad (9)$$

where $(\Delta x)_{\max}$ is the upper limit of the gray zone resolution (typically ≃13km) and $(\Delta x)_{\min}$ is the lower limit of the gray scale resolution (typically ≃3km). Note that 5-13km contains most of the variation of the strength of the parameterized drag, but it may extend to larger grid spacing (e.g., ~50km) as the mean characteristics of the subgrid terrain change with further coarsening of the grid spacing.

As orographically generated gravity waves propagate upward into the increasingly rarified atmosphere, the velocity tendency associated with breaking waves has the potential to become exceedingly large (see Eq. 1). In model testing with the 127-level FV3GFS, whose model top is at approximately 78 km, parameterized wind tendencies occasionally became excessive above the stratopause due to the low air density (~10^{-3} – 10^{-5} kg m^{-3}) and caused the wind to reverse direction in a single time step, leading to model crashes. Such crashes (and sudden flow reversals) were prevented by decreasing the physics time step by an order of magnitude, but of course, this is not a practical solution due to operational constraints. In order to avoid sudden flow reversals, we reduce the velocity tendency at each time step, if necessary, to a value that limits the percentage change of the velocity magnitude to a given threshold. Defining the prescribed maximum percentage change as μ ($0 \leq \mu \leq 1$), a tendency-limiting factor γ is calculated as

$$\gamma = \min\left(1, \mu \left|\frac{U}{\Delta t(dU/dt)}\right|\right), \qquad (10)$$

where dU/dt is the wind tendency calculated by the OGWD parameterization. The final tendency is multiplied by γ.

2.2 Low-level flow blocking

The representation of low-level flow blocking by subgrid-scale orography in the UGWP suite follows the scheme of KD05, as originally coded in the WRF-ARW. With the development of the GSL orographic drag suite, the gray-scale tapering of Eq. (9) has been introduced. As with the OGWD scheme, the first step is to calculate the surface stress, which, for blocking, is calculated as

$$\tau_{BLK} = \frac{1}{2}\rho_0 \frac{m}{\Delta_x^2} C_d \Delta_x^{\perp} L_x^{\perp} h_B |U_0|^2, \qquad (11)$$

where Δ_x^2 is the grid cell area, C_d is a bulk drag coefficient ($\cong$1), $\Delta_x^{\perp}$ is length of the grid cell in the cross-wind direction, $L_x^{\perp}$ represents the dominant subgrid-scale orography along the cross-wind direction, that is, the effective orographic length L_x perpendicular to the wind direction, and finally, h_B is the height of the blocking level (see Fig. 1). The blocking level is the height above the surface of the lowest upstream streamline that is able to flow over the subgrid-scale mountain, which is assumed to have a height of $H = 2\sigma_h$. The flow along this streamline has just sufficient kinetic energy (KE) to oppose the potential energy (PE) associated with lifting an air parcel against the buoyancy force from the blocking height to the mountain height. Below the blocking level, the flow must split and pass around the mountain, which imparts a drag force on the flow. Following Alpert (2004), the PE at a given model level k (below the mountain height H) is calculated by summing the buoyancy force times the layer thickness from k to k_{mtn}, where k_{mtn} is the model level corresponding to the mountain height H, that is

$$(PE)_k = \sum_{k'=k}^{k_{mtn}} N_{k'}^2 (H - z_{k'})(\delta z)_{k'}, \qquad (12)$$

where z_k is the model level height and δz is the model layer thickness. The KE at each level is

$$(KE)_k = \frac{1}{2}[(u_k)^2 + (v_k)^2]. \qquad (13)$$

The blocking height (h_B) is found when $(PE)_k > (KE)_k$, that is when the flow does not have sufficient kinetic energy to overcome the potential energy to make it over the mountain. The blocking height is used in (11) to calculate the surface stress. The stress is then tapered linearly from the surface to the blocking height as shown in Fig. 1. Note that the inclusion of moist processes in the calculation of blocking potential has not yet been considered, but this will be explored in future versions of the UGWP.

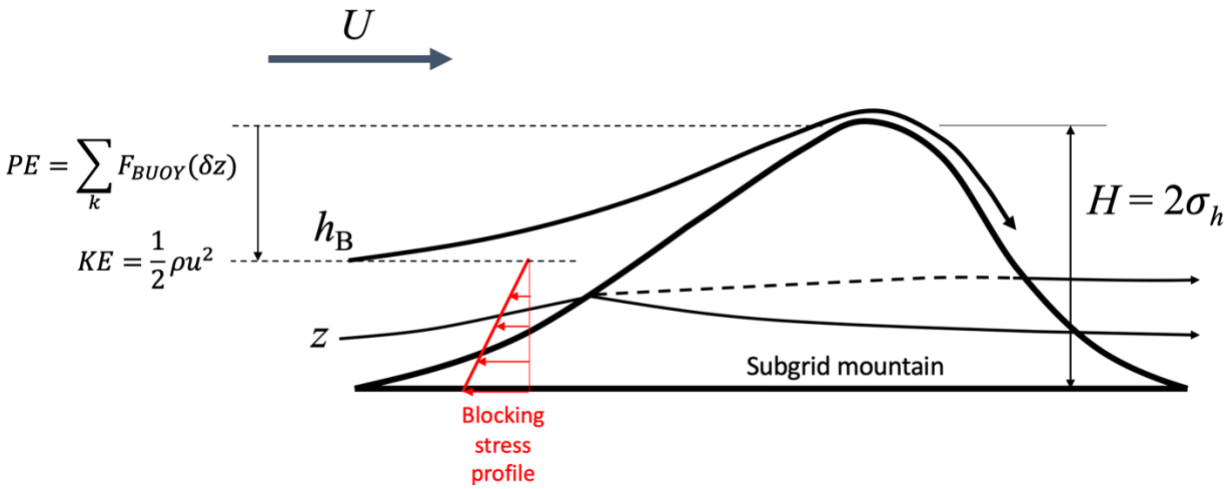


Figure 1. Schematic of low-level flow blocking features. Streamlines represent flow over and around the mountain.

2.3 Small-scale gravity wave drag

The small-scale GWD (SSGWD) scheme of Steenveld et al. (2008) and Tsiringakis et al. (2017) captures the effects of gravity waves produced by horizontal terrain variations on scales down to ~1 km in length. Such small-scale waves can propagate vertically under highly stable conditions, as in nocturnal PBLs. The scheme is active for all horizontal grid spacings.

The first step is to calculate the surface stress (τ_{wave}) induced by the SSGWD, which is approximated from linear theory (e.g., Holton 2004), by

$$\tau_{\text{wave}} = \begin{cases} \frac{1}{2}\rho_0 k_S m H^2 N U, & \text{if } \frac{N}{U} \geq k_S, \\ 0 & \text{if } \frac{N}{U} < k_S, \end{cases} \tag{14}$$

Where ρ_0 is the near-surface air density, k_S is the orographic wave number, H is the amplitude of the subgrid-scale topography, N is the Brunt-Väisälä frequency, and U is the mean wind speed at the PBL top. As in KD05, the topographic amplitude is calculated as twice the standard deviation of the subgrid topography (σ_h), i.e.,

$$H = 2\sigma_h. \tag{15}$$

Tsiringakis et al. (2017) use m, "number of mountains" from (3) for the orographic wave number, k_S. We instead provide a direct estimate of the topographic wave number, which can be used as a "tuning parameter". The subgrid terrain consists of a spectrum of wave numbers, so providing a single wave number is, of course, an oversimplification. As a starting point, we use $k_S = 0.001\pi$ m^{-1}, which corresponds to a horizontal wavelength of 2 km.

In order to include only the effects of small-scale topography within a grid cell, the SSGWD scheme is only active when the PBL height, h, is greater than the subgrid topographic height, 2 σ_h.

The momentum flux at the surface is linearly distributed in the vertical up to the PBL top by the relation:

$$\tau_{wave}(z) = \tau_{wave}(0)\left(1 - \frac{z}{h}\right), \qquad \text{for } z \leq h. \tag{16}$$

(Note that in Tsiringakis et al. (2017), the term on the RHS is raised to the second power.)

2.4 Turbulent orographic form drag

The turbulent orographic form drag (TOFD) parameterization is based on Beljaars et al. (2004), and accounts for drag due to horizontal topographic variations on scales of 5 km and smaller. Prior to implementation of the TOFD scheme, the turbulent effects of orography were often represented by increasing the roughness length in the PBL scheme in an ad hoc manner. As small-scale orography mainly effects the turbulent momentum transport, and not that of heat or moisture, the roughness lengths for these scalar transports are maintained at their nominal value based on the land surface parameters. The advantage of using the TOFD parameterization is that the orographic drag effects are physically

represented, and the roughness lengths for momentum transport may be returned to their standard value, based on the land and vegetation types.

The theory behind TOFD is that the small-scale topography effects the turbulent flow in the PBL in such a way that the turbulent pressure perturbation field becomes positively correlated with the slope of the topography, such that a net force in opposition to the flow develops. In Beljaars et al. (2004), the drag force is derived based on neutral PBL conditions, however, the form drag is applied under all conditions of stability. Note that TOFD is not a gravity wave phenomenon, as it does not involve the vertical transport of momentum and energy.

The expression for the form drag is based on an empirical breakdown of the topographic spectrum down to wavelengths on the order of 10 m. Since there is no global topographic data at such a high resolution, the Beljaars et al. (2004) extrapolates the power spectrum to this small scale from the available 1 km global data set (see below). The expression for the momentum tendency due to form drag is given as a function of height by

$$\frac{\partial}{\partial z}\left(\frac{\boldsymbol{\tau}}{\rho}\right) = -\alpha\beta C_{\mathrm{md}} C_{\mathrm{corr}} |\mathbf{U}(z)| \mathbf{U}(z) 2.109 e^{-1.5(z/1500)} a_2 z^{-1.2}, \qquad (17)$$

where α=12, β=1, C_{md}=0.005, C_{corr}=0.6, $\mathbf{U}$ is the horizontal wind vector, and $a_2 = a_1 k_1{}^{(n_1-n_2)}$, where k_1=0.003 m^{-1}, $a_1 = \sigma_{flt}^2 \left(I_H k_{flt}^{n_1}\right)^{-1}$, k_{flt}=0.00035 m^{-1}, n_1=-1.9, n_2=-2.8, I_{H}=0.00102 m^{-1}, and σ_{flt} is the standard deviation of the "filtered" subgrid topography described below in section 4.

2.5 Non-stationary gravity wave drag

Gravity waves generated by high-altitude, non-stationary sources, such as deep convection, frontal instability, and stratified shear instability associated with the tropospheric jet, produce drag forces where they dissipate in the upper stratosphere and mesosphere (e.g., Scinocca and Ford 2000; Scinocca 2003). These waves are essential drivers of the zonal mean wind and temperature structure of the middle atmosphere. Unlike their orographic counterparts, non-stationary gravity waves (NGW) have non-zero horizontal phase speeds. Since they cannot be fully resolved by the model resolution, the drag produced by NGW must be parameterized in numerical models.

The NGW drag parameterization in the UGWP was developed by Valery Yudin (Yudin et al. 2018) and is based on the work of Scinocca (2003). An early version ("version 0") is currently implemented in the GFSv16 model. Here we describe the latest "version 1" development planned for the next GFS release. The dispersion relation for linear, nonhydrostatic internal gravity waves with vertical wave number (m) and horizontal wave number (k), is given by

$$m^2 = \frac{N^2-\omega^2}{\omega^2-f^2} k^2, \qquad (18)$$

where N is the buoyancy frequency, f is the Coriolis frequency, and ω is the intrinsic frequency, i.e., the wave frequency with respect to the moving fluid, given by

$$\omega = \omega_0 - kU, \qquad (19)$$

where U is the mean wind in the direction of the horizontal wave propagation, and ω_0 is the wave frequency with respect to the ground. Note that $\omega_0 = 0$ for stationary, orographic GWs. By contrast, the dispersion relation for hydrostatic gravity waves in the absence of rotation is

$$m^2 = \frac{k^2 N^2}{\omega^2} = \frac{N^2}{c^2}, \quad (20)$$

Where c is the intrinsic phase speed of the gravity wave.

The NGW scheme of Scinocca (2003) is based on a central assumption that the background launch spectrum of NGWs is empirically determined, azimuthally isotropic, and independent of time and geographic location. In this scheme, NGWs are launched from a height of 275 hPa, and their upward propagation is governed by the atmospheric sounding above this level. As the NGWs propagate, they transfer momentum between atmospheric layers, depositing force at the level where they dissipate in an equal and opposite manner to the force imparted at the source level.

Notably, the NGW parameterization accounts only for the upward propagation and dissipation of the momentum flux, neglecting the accelerating force on the mean flow below the launch level and the downward transfer of momentum. These effects are considered negligible compared to the pronounced effects on the middle atmosphere, particularly as air density diminishes with increasing altitude. Consequently, NGW parameterization schemes focus on three key elements: the specification of NGW sources, the propagation and dissipation of waves, and the resulting impacts on the resolved dynamics and thermodynamics at the level of dissipation.

Following the approach of Fritts and Vanzandt (1993), the empirical spectra for the launch energy of NGW, denoted as $E(\omega,m)$, and Vertical Momentum Flux (VMF), represented as $F(\omega,m)$, are expressed as functions of the vertical wave number, m, and intrinsic frequency, ω, i.e.,

$$E(\omega, m) = E_0 C_A E_A(m) C_B E_B(\omega), \quad (21)$$

and

$$F(\omega, m) = -\rho \left(\frac{k}{m}\right) E(\omega, m), \quad (22)$$

where

$$E_A(m) = \frac{(\frac{m}{m_*})^s}{1+(\frac{m}{m_*})^{s+3}}, \quad (23)$$

$$C_A = 1/\int_{m_e}^{m_h} E_A(m) dm, \quad (24)$$

$$E_B(\omega) = \omega^{-p}, \quad (25)$$

and

$$C_B = 1/\int_{\omega_1}^{\omega_2} E_B(\omega) d\omega. \quad (26)$$

Note that $m_* = 2\pi/2km$ represents a characteristic vertical wave number and $s = 1$ corresponds to the typical shape of the NGW energy spectrum. The variables k, E_0, C_A, C_B, E_A, and E_B denote the horizontal wave number, total wave energy excluding the air density factor, and the normalized constant of empirical spectra of vertical wave number and frequency, respectively. m_e and m_h are determined by the maximum and minimum of vertical wave length, λ_{max}, λ_{min} (or phase velocity, c_{max}, c_{min}, Eq. 20). The NGW source spectrum is divided into $n = 25$ phase speed bins, discretized within the specified range of phase velocity (along with their corresponding vertical wave lengths),

$$c_{min} = 2.5m/s\ (\lambda_{min} = 0.75km), \quad c_{max} = 52.5m/s(\lambda_{max} = 15.75km),$$

and n_ϕ equally spaced azimuths. Details of the empirical NGW energy and momentum spectrum specification is available in Scinocca (2003) and Yudin et al. (2018).

In addition to the normalized empirical VMF spectrum, the final yet crucial parameter to completing the NGW source specification is the prescribed VMF amplitude (τ_{NGW}) at the launch level, which is parameterized based on latitude, exhibiting three peaks at the equator and north/south storm tracks, respectively. Specifically,

$$\tau_{\mathrm{NGW}} = \begin{cases} 1.25\tau_0, & |lat| \le 3^\circ, \\ 1.25\tau_0 e^{-\left(\frac{|lat|-3}{12.5}\right)^2}, & 3^\circ < |lat| < 15.3^\circ, \\ 0.25\tau_0, & 15.3^\circ \le |lat| < 31^\circ, \\ \tau_0 e^{-\left(\frac{|lat|-60}{23}\right)^2}, & 31^\circ \le |lat| < 60^\circ, \\ \tau_0 e^{-\left(\frac{|lat|-60}{25}\right)^2}, & lat \ge 60^\circ, \\ \tau_0 e^{-\left(\frac{|lat|-60}{30}\right)^2}, & lat \le -60^\circ. \end{cases} \tag{27}$$

It was recognized that the prescribed VMF amplitude at the launch level should diminish as the resolution increases, aiming to maintain the total (resolved plus subgrid) momentum flux at a roughly constant level. As a result, τ_0 is set to 1.5e-3, 0.8e-3, 0.5e-3 Pa for resolutions C192, C384 and C768 respectively.

The NGW spectrum is integrated over the possible values of intrinsic frequency and vertical wavenumber. Based on the resolved model sounding, the fate of each monochromatic vertically propagating NGW is evaluated on a level-by-level basis. During this process, the total momentum flux, $\boldsymbol{\tau}_{\mathrm{NGW}} = (\tau_{x,\mathrm{NGW}}, \tau_{y,\mathrm{NGW}})$, is determined. Various outcomes are possible for NGWs. Once scenario involves reflection at a level where vertical propagation is prohibited, meaning the intrinsic frequency becomes imaginary, causing the wave to reflect downward without inducing drag. Another occurs when waves encounter critical levels, where the vertical wave number becomes infinite, leading to their removal from the wave spectrum and the transfer of their momentum to the mean flow, thereby generating a drag force. Waves can also break and induce drag, typically through wave saturation. In this case, the wave amplitude grows until static instability occurs. Nonlinear dissipation is modeled empirically by constraining the gravity wave energy spectrum at large vertical wave number (short wave lengths) within the observed saturation tail, which follows a proportionality of m^{-3}. The saturated energy density spectrum is

$$E_{AS}(m) = \left(\frac{m}{m^*}\right)^{-3}, \tag{28}$$

which is equivalent to E_A in Eq. (23) but at asymptotically large m. The saturation momentum flux F_s is calculated based on Eqs. (21) and (22), with E_A superseded by E_{AS} correspondingly. F_s diminishes with height due to decreasing density.

Critical level filtering and nonlinear dissipation are examined for each azimuthal bin, at each level above the launch level, and for each phase speed. Notably, the NGW scheme updates the temperature tendency resulting from both NGW dissipation and the momentum eddy-diffusivity. In atmospheric models that extend into the mesosphere and thermosphere, turbulent and molecular diffusion contributes additional dissipation of NGW energy. Energy conservation requires that all vertical diffusion applied to the resolved flow must also damp the parameterized NGWs as well. As a result, the portion of the NGW spectrum that propagates upward without being removed by critical level filtering or nonlinear dissipation

is also subject to diffusion damping. At each model level, both the divergence of the VMF summed over the NGW spectrum and the vertical diffusion of parameterized NGWs contribute to the momentum tendencies in the zonal and meridional directions, affecting resolved horizontal winds. The divergence of the net eastward, $\overline{F_E}$, and northward, $\overline{F_N}$ momentum fluxes are $g\frac{\partial \overline{F_E}}{\partial p}$ and $g\frac{\partial \overline{F_N}}{\partial p}$, respectively. $\overline{F_E}$ and $\overline{F_N}$ are derived by summing the total momentum flux (i.e. integrated over all phase speed bins) in each azimuthal ϕ_i projected onto the east and north directions, respectively, i.e.,

$$\overline{F_E} = \sum_{i=1}^{n_\phi} F(\phi_i)\ cos\phi_i \tag{29}$$

and

$$\overline{F_N} = \sum_{i=1}^{n_\phi} F(\phi_i)\ sin\phi_i. \tag{30}$$

This approach has been implemented in the GFS, allowing for local cooling and heating by NGW-induced mixing and GW energy dissipation (see also Becker and McLandress 2009).

3. Semi-implicit time differencing

The SSGWD and TOFD schemes require a semi-implicit treatment of the time differencing for numerical stability, especially with the long physics time steps used in the FV3GFS. Here we provide a brief description of the algorithm. We start by writing the time-continuous tendency due to a given physics parameterization as

$$\left(\frac{\partial u}{\partial t}\right)_{PHYS} = -uf(u), \tag{31}$$

where $f(u)$ represents the grid-resolved contribution to the tendency by a given scheme as a function of the u-wind. We can write the time-discrete tendency equation as

$$\frac{u^{n+1}-u^n}{\Delta t} = -u^{n+1}f(u^n), \tag{32}$$

where n and n+1 are the current and next time steps, respectively. Note that with the SSGWD and TOFD schemes, the unknown n+1 u-wind is only needed at one vertical level, which greatly simplifies the solution for u^{n+1}. Rearranging (32), we get

$$u^{n+1} = \frac{u^n}{1+f^n\Delta t}, \tag{33}$$

where $f^n = f(u^n)$. Given that physics tendencies are treated in a forward Euler manner, that is,

$$\left(\frac{\partial u}{\partial t}\right)_{PHYS} = \frac{u^{n+1}-u^n}{\Delta t}, \tag{34}$$

we can use (33) in (34) to arrive at

$$\left(\frac{\partial u}{\partial t}\right)_{PHYS} = -u^n\frac{f^n}{1+f^n\Delta t}, \tag{35}$$

which is the contribution to the overall physics tendency by a given scheme.

4. Topographic statistics fields

The static data used by the orographic drag parameterizations are calculated from high-spatial resolution topographic data sets. Two sets of static fields are provided: one for the mesoscale GWD and mountain blocking schemes (stored in files named *_oro_data_ls.tilex.nc), and the other for the small-scale GWD and TOFD schemes (stored in files named *_oro_data_ss.tilex.nc). The difference between the two sets is that the former is calculated from a smoothed field on a global 2.5 minute lat-lon grid, and the latter is calculated from a band-pass filtered topographic field on a global 30 second lat-lon grid. In each case, the subgrid statistical parameters described in Section 2, are calculated from the data sets. The parameters are listed and described in Table 1.

Parameter	Description	Associated variable in Sect. 2 equations
stddev	Standard deviation of subgrid topography	σ_h
convexity	Convexity of subgrid topography	OC
oa1	Orographic asymmetry in west direction	OA
oa2	Orographic asymmetry in south direction	OA
oa3	Orographic asymmetry in south-west direction	OA
oa4	Orographic asymmetry in north-west direction	OA
ol1	Orographic effective length for westerly flow	L_x
ol2	Orographic effective length for southerly flow	L_x
ol3	Orographic effective length for south-westerly flow	L_x
ol4	Orographic effective length for north-westerly flow	L_x

Table 1. Subgrid topographic statistical parameters stored in data sets for use by orographic drag parameterizations.

This section describes how the high-resolution topographic data is preconditioned before these parameters are calculated. The Fortran code used to precondition the topographic data and calculate the statistical

parameters reside in the "develop" branch of the UFS GitHub repository at: https://github.com/ufs-community/UFS_UTILS/tree/develop/sorc/orog_mask_tools.fd/orog_gsl.fd.

4.1 Data preparation for mesoscale GWD and blocking fields

The smallest horizontal wavelength of topography that generates vertically propagating gravity wave in the free atmosphere is generally considered to be approximately 5km (e.g., Beljaars et al. 2004). This cutoff is also applied to the low-level blocking scheme. To calculate the statistics of the subgrid topography, we therefore use a 2.5 minute lat-lon topographic dataset, such that variations smaller than ~5km are removed. The source for the high-resolution data is the U. S. Geological Survey's (USGS) GMTED2010 30 second lat-lon dataset (Danielson and Gesch 2011). This data is interpolated on to the 2.5 minute grid using the WRF Preprocessing System (WPS) (Skamarock et al. 2019).

Our goal is to represent the subgrid topography within each model grid cell in such a way that information about the resolved topography is not included. For example, if there were no subgrid variations of topography, such as with a smooth, sloping plain that is fully resolved by the model (coarse) grid, the calculation of the standard deviation of the high-resolution topography within a model grid cell would be non-zero, resulting in a non-zero drag force calculated by the subgrid parameterization. Since the drag should be fully resolved by the model dynamics, the addition of parameterized subgrid drag would be redundant. To eliminate such redundancy, we "detrend" the high-resolution data by subtracting out the resolved coarse-grid topography, as shown in Fig. 2. Note that this method is also used at ECMWF (e.g., Wedi 2016) and is described in detail in Scinocca and McFarlane (2000). The resolved coarse-grid field is calculated via bilinearly interpolating the coarse-grid heights on to the high-resolution (2.5 minute) lat-lon grid.

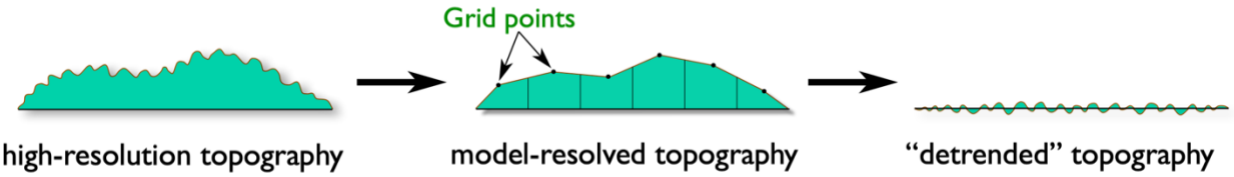


Figure 2. Steps to determine subgrid (detrended) topography from high-resolution topographic data. A typical profile of topography is shown on the left. The topography, as it is resolved by the horizontal model grid, is shown in the center. The detrended topography on the right is calculated by subtracting the model-resolved topography from the high-resolution topography.

4.2 Data preparation for small-scale fields

The small-scale GWD and TOFD schemes both represent the effects of variations in subgrid topography down to the 1 km horizontal scale. The statistical parameters used for the schemes are based on the GMTED2010 30-second lat-lon global topographic datasets. Following Beljaars et al. (2004), the high-resolution elevation data are passed through a spectral band-pass filter to isolate wavelengths in the range ~50km to ~2km, using the filter shown in Fig. 3.

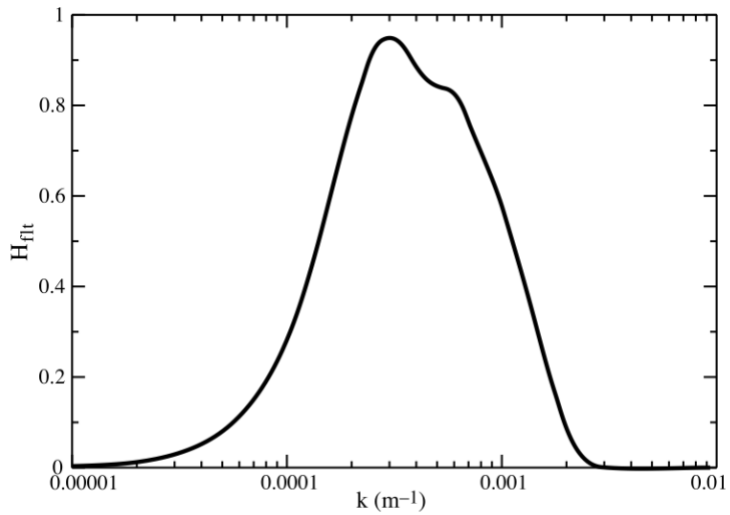


Figure 3. Spectral filter applied to global GMTED2010 30-second topographic data set for calculation of statistical parameters for small-scale GWD and TOFD schemes. [Figure from Beljaars et al. (2004).]

5. Results

5.1 Rapid Refresh (RAP)

Here we present results of tests with the 13km horizontal resolution RAPv5 that show the sensitivity of forecast skill to the various components of the GSL drag suite. The tests consist of a series of reforecasts initialized every 3 hours starting at 0000 UTC 2 Feb 2019 and continuing through 0000 UTC 15 Feb 2019. Eight experiments were run with different combinations of the orographic drag components of the GSL drag suite (note that the non-stationary GWD drag was not active). Experiment "0" is the control run with all of the components active. Experiment "1" is the case with no GWD. Experiments "2-7" implement various combinations of the drag components as shown in the table within Fig. 4, where "MS" is mesoscale GWD, "BL" is low-level blocking, "SS" is small-scale GWD, and "FD" is turbulent orographic form drag. The number "1" indicates the schemes that are active in each experiment. Figure 4 shows vertical profiles of RMS errors and bias of windspeed with respect to radiosonde observations (RAOBS) over the full RAP domain at forecast hour 27. The control experiment, with all the GSL drag components active, generally gives the best results, with Experiment "1" with no GWD parameterizations giving the worst results. The experiments with partial use of the drag components give intermediate results.

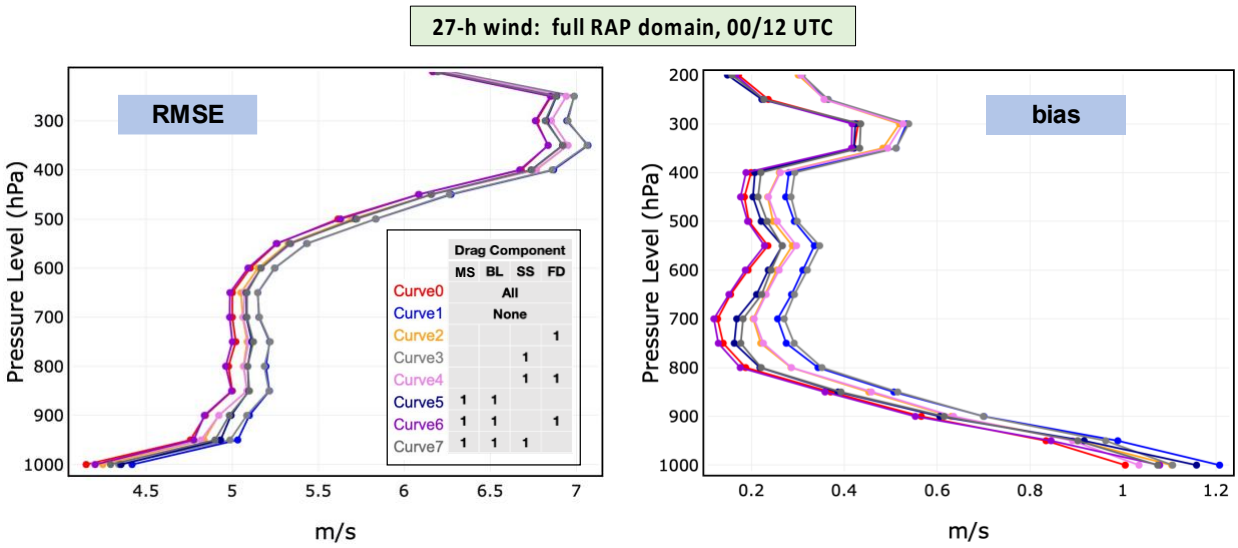


Figure 4. Vertical profiles of RMS windspeed error and bias over the full RAP domain compared to RAOBS at 00 and 12 UTC at forecast hour 27. The table shows the configurations of the eight experiments with "1" indicating that a given drag scheme is turned on. The control experiment, with all schemes turned on is denoted by "Curve0", and the "Curve1" is a run with no GWD parameterization. The drag components are "MS" (mesoscale GWD), "BL" (low-level flow blocking), "SS" (small-scale GWD) and "FD" (turbulent orographic form drag).

The effects of the drag parameterizations on the surface winds are shown in the 10-m windspeed RMSE and bias plots of Figs. 5 through 7. Figure 5 shows the errors over the full RAP domain at forecast hour 21, with the largest errors generally associated with the run with no GWD parameterization, and the least errors from the control run.

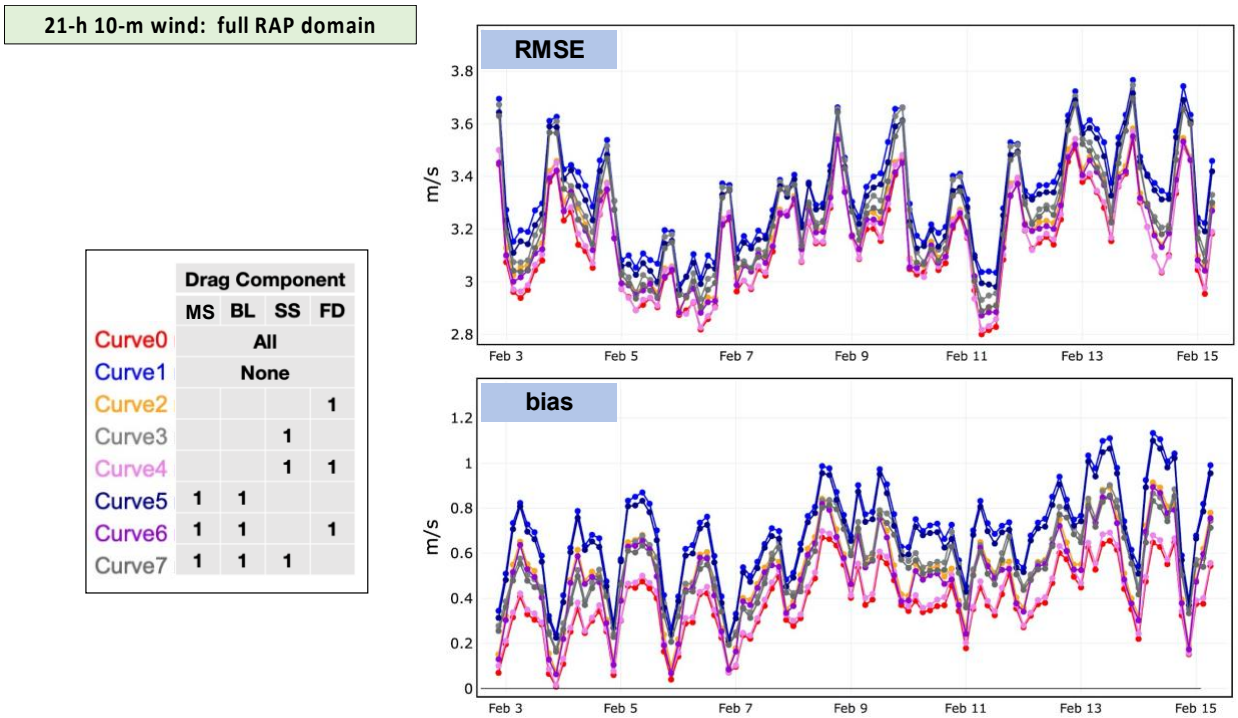


Figure 5. RMS error and bias of 10-m windspeed at forecast hour 21 compared to METAR observations over the full RAP domain. The table shows the configurations of the eight experiments with "1" indicating that a given drag scheme is turned on. The control

experiment, with all schemes turned on is denoted by "Curve0", and the "Curve1" is a run with no GWD parameterization. The drag components are "MS" (mesoscale GWD), "BL" (low-level flow blocking), "SS" (small-scale GWD) and "FD" (turbulent orographic form drag).

The results over the western CONUS, shown in Fig. 6, exhibit similar results to the full RAP domain, with a large reduction in the 10m windspeed bias that sometimes becomes negative with the drag suite in use.

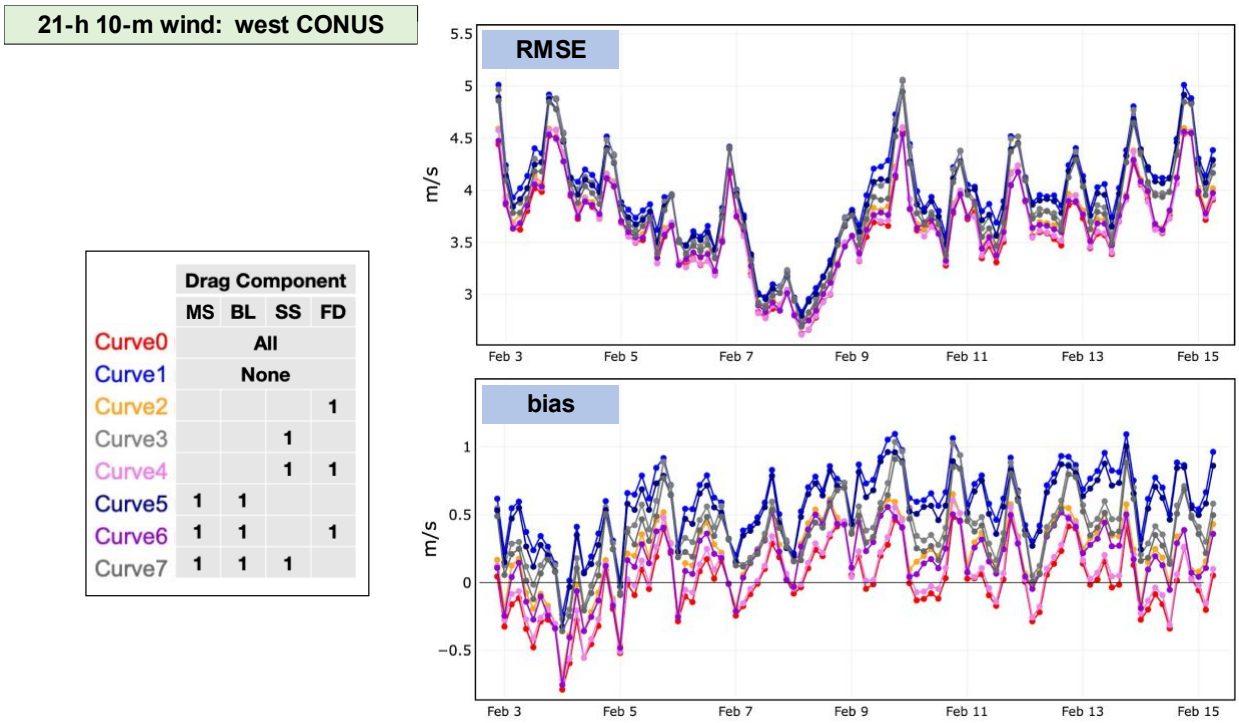


Figure 6. As in Fig. 5 except over the western CONUS.

Figure 7 shows that the RMSE of the 10m windspeed in the eastern CONUS is generally smaller than that of the western CONUS, while the BIAS tends to be larger. The effects of the drag parameterizations are smaller than those over the western CONUS due to the smaller subgrid terrain variations in the east. However, the effects of the parameterizations are favorable (comparing 'Curve0' to 'Curve1' in both RMSE and bias).

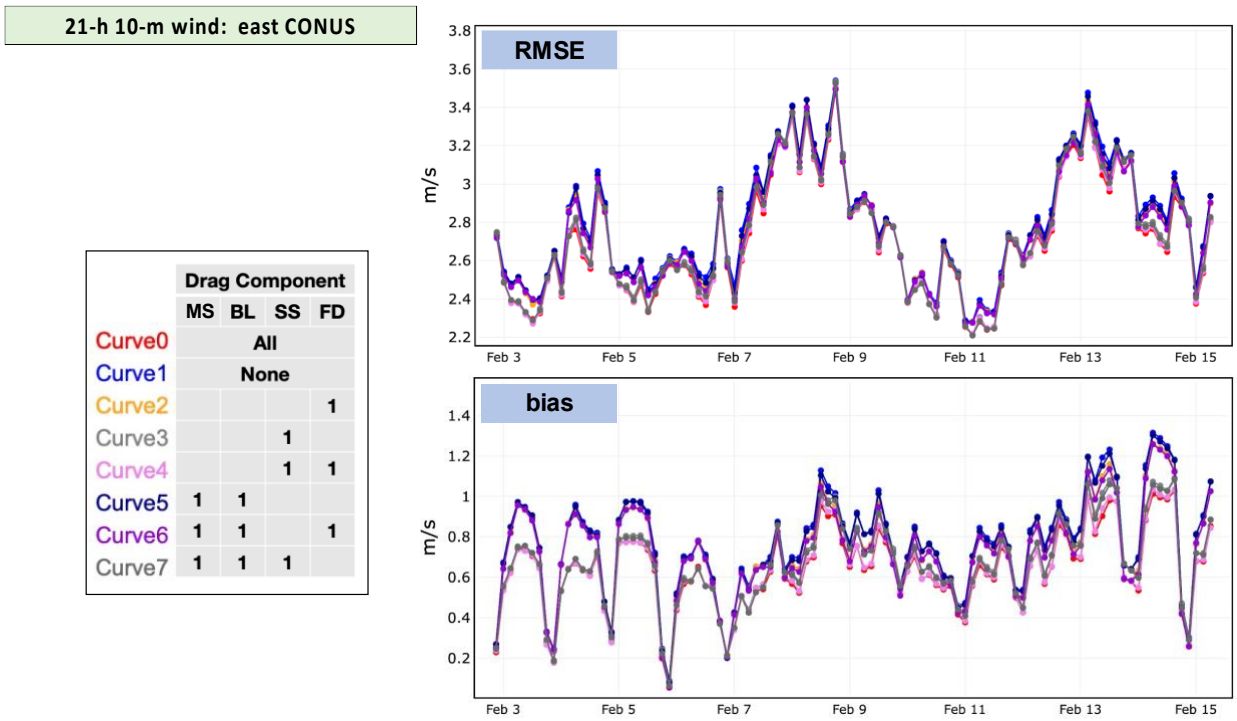


Figure 7. As in Fig. 5 except over the eastern CONUS.

Figures 8 and 9 show that the control experiment consistently gives the best results over the course of the diurnal cycle. Of note is the diurnal cycle of the effects of the small-scale GWD, especially over the eastern CONUS (Fig. 9). In experiments "0, 3, 4 and 7", the SSGWD is active, and during the nighttime and early morning hours (~3-15UTC) when it is most active due to the presence of highly-stable nocturnal PBLs, the windspeed bias is noticeably reduced.

21-h 10-m wind: west CONUS, diurnal composite

	Drag Component			
	MS	BL	SS	FD
Curve0	All			
Curve1	None			
Curve2				1
Curve3			1	
Curve4			1	1
Curve5	1	1		
Curve6	1	1		1
Curve7	1	1	1	

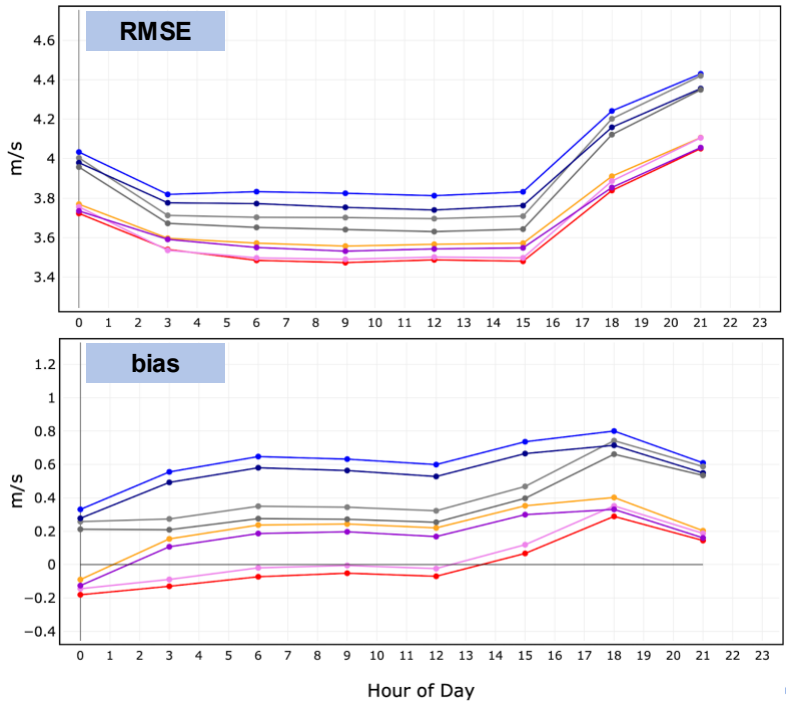


Figure 8. Diurnal composite of 10-m windspeed RMS error and bias at forecast hour 21 compared to METAR observations over the western CONUS. The table shows the configurations of the eight experiments as in Figures 4-7.

21-h 10-m wind: east CONUS, diurnal composite

	Drag Component			
	MS	BL	SS	FD
Curve0	All			
Curve1	None			
Curve2				1
Curve3			1	
Curve4			1	1
Curve5	1	1		
Curve6	1	1		1
Curve7	1	1	1	

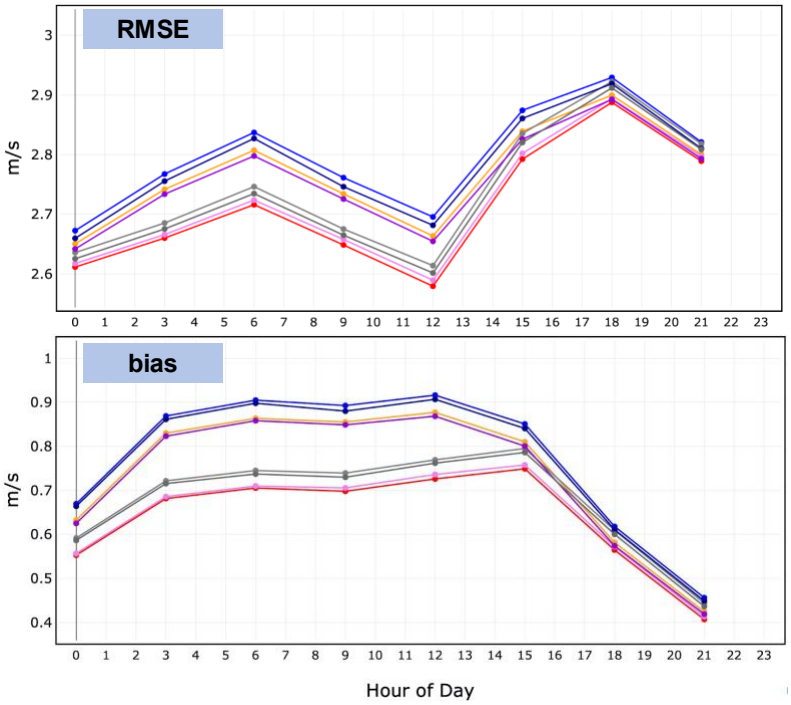
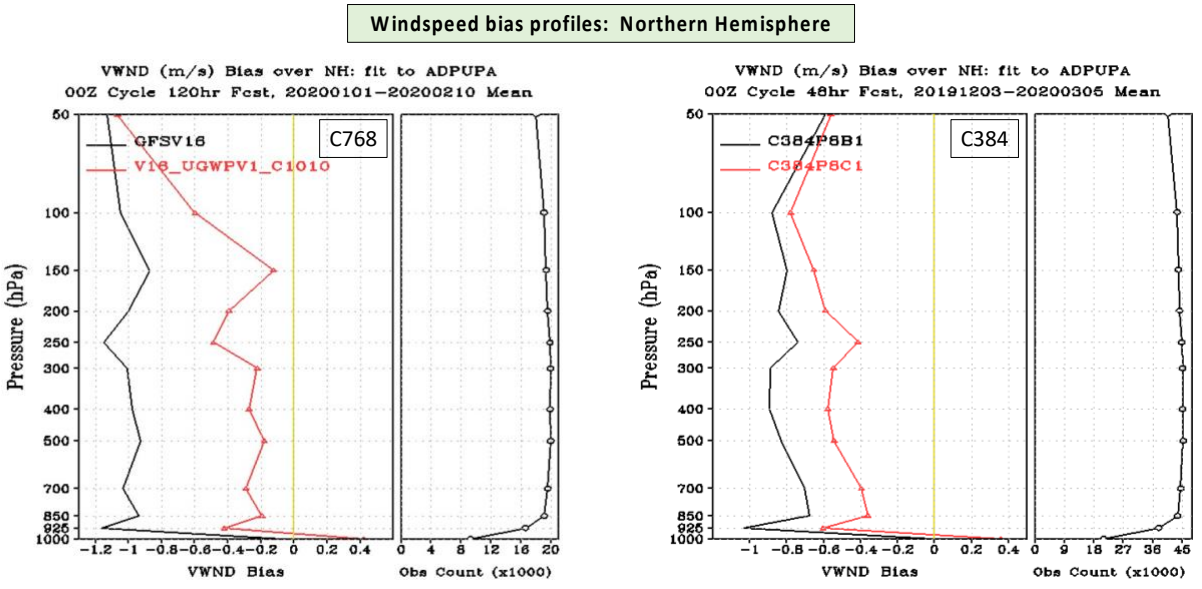


Figure 9. As in Fig. 8 except over the eastern CONUS.

5.2 Global FV3GFS

The UGWP has been tested in the global FV3GFS NWP model in reforecast experiments with the C768 (Δx~13km) and C384 (Δx~26km) cubed-sphere grids as part of the UFSR2O project (website: https://vlab.noaa.gov/web/ufs-r2o). The baseline (control) experiments with the C768 grid use the operational GFSv16 physics parameterization suite. For the sensitivity experiment, the same physics suite is used except that the GWD physics is replaced by the UGWP drag suite. The warm-start reforecast period for the C768 runs were 1 Jan 2020 to 10 Feb 2020, initialized at 00UTC every five days. For the C384 experiment, the baseline physics parameterizations were those of the GFSv17 Prototype 8b configuration (https://github.com/ufs-community/ufs-weather-model/tree/Prototype-P8), in which the GFSv16 mesoscale OGWD and blocking are used along with the small-scale OGWD and TOFD schemes of the GSL drag suite. The sensitivity test was to substitute the GFSv16 mesoscale schemes with the GSL mesoscale OGWD and blocking schemes. The C384 warm-start reforecasts were initialized every 5 days at 00 UTC from 3 Dec 2019 through 5 Mar 2020.

Fig. 10 shows the sensitivity of the northern hemisphere windspeed bias (as compared to RAOBS) to the changes in GWD schemes. The left panel shows improvement in the bias over the depth of the troposphere and lower stratosphere due to the use of the UGWP at C768 resolution. The right panel shows improvement in using the GSL mesoscale drag schemes in the C384 experiment.



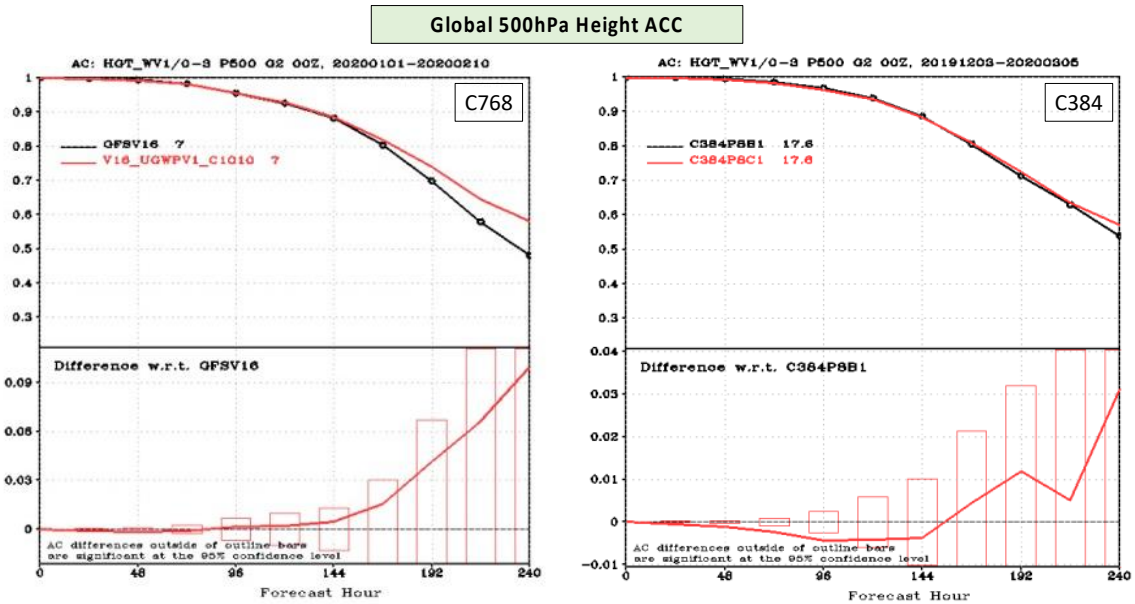
GFSv16
UGWP: cdmbgwd = (10,10)

GFSv16 OGWD+BLK
UGWP: cdmbgwd = (10,5)

Figure 10. Global FV3GFS northern hemisphere windspeed bias profiles. Left panel: C768 (~13km) grid – GFSv16 GWD (black curve) and UGWP with cdmbgwd=(10,10) (red curve). Right panel: C384 (~26km) grid – GFSv16 mesoscale OGWD and blocking (black curve) and UGWP with cdmbgwd=(10,5) (red curve). See Subsection 6.1 for an explanation of the cdmbgwd parameters.

The 500hPa height anomaly correlation coefficient (ACC) die-off curves are shown in Fig. 11. In the C768 run there is comparable performance to the GFSv16 GWD physics for the first 96 forecast hours,

with improvement with the UGWP after this forecast time. In the C384 runs, there is some degradation of the ACC for the first 144 hours, but afterward the GSL mesoscale drag and blocking exhibit improvement.



GFSv16
UGWP: cdmbgwd = (10,10)

GFSv16 OGWD+BLK
UGWP: cdmbgwd = (10,5)

Figure 11. Global FV3GFS 500hPa height anomaly correlation coefficient (ACC) die-off curves. Left panel: C768 (⊿x~13km) grid – GFSv16 GWD (black curve) and UGWP with cdmbgwd=(10,10) (red curve). (The lower left panel shows the difference between the UGWP and GFSv16 results.) Right panel: C384 (⊿x~26km) grid – GFSv16 mesoscale OGWD and blocking (black curve) and UGWP with cdmbgwd=(10,5) (red curve). (The lower right panel shows the difference between the UGWP and GFSv16 results.) See Subsection 6.1 for an explanation of the cdmbgwd parameters.

The effects on the low-level winds over the western CONUS are shown in Fig. 12. At both grid resolutions, there is improvement to the 10-m windspeed bias in the sensitivity experiments using the UGWP (C768) and GSL mesoscale OGWD and blocking schemes (C384).

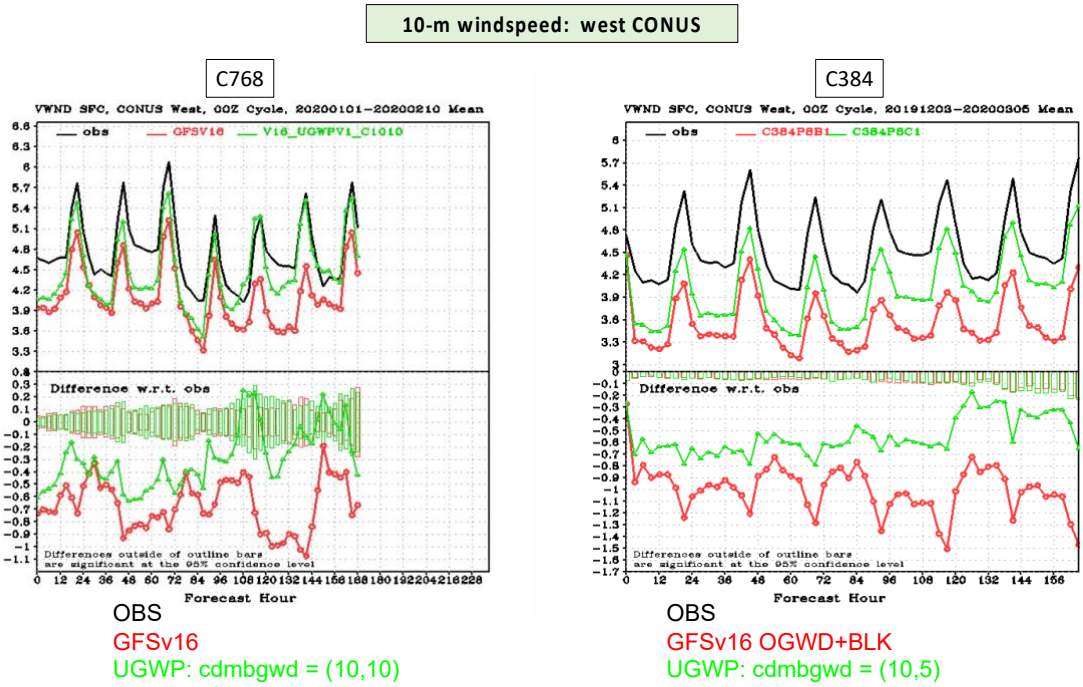


Figure 12. Global FV3GFS western CONUS 10=m windspeed. Left panel: C768 (⊿x~13km) grid – METAR observations (black curve), GFSv16 GWD (red curve), and UGWP with cdmbgwd=(10,10) (green curve). (The lower left panel shows the difference between the experimental results and observations.) Right panel: C384 (⊿x~26km) grid – METAR observations (black curve), GFSv16 mesoscale OGWD and blocking (red curve), and UGWP with cdmbgwd=(10,5) (green curve). (The lower right panel shows the difference between the experimental results and observations.) See Subsection 6.1 for an explanation of the cdmbgwd parameters.

6. Code description and model configuration

6.1 Code description and namelist options

The Unified Gravity Wave Physics (UGWP) suite of parameterizations is a physical package within the Common Community Physics Package library (Heinzeller et al. 2022). The Version 1 package name is *ugwpv1_gsldrag,* which is prescribed by the XML code of a CCPP Suite Definition File (SDF) by the following sequence of lines:

<scheme>GFS_GWD_generic_pre</scheme>
<scheme>ugwpv1_gsldrag</scheme>
<scheme>ugwpv1_gsldrag_post</scheme>
<scheme>GFS_GWD_generic_post</scheme>.

Note that the Fortran module for the package is contained in ccpp/physics/physics/ugwpv1_gsldrag.F90.

There are a series of runtime namelist logical flags that activate the various UGWP components. For a comprehensive list of namelist variables, see https://dtcenter.ucar.edu/GMTB/v6.0.0/sci_doc/_c_c_p_psuite_nml_desp.html. By default, all of the components are activated, but the user may turn any individual scheme off if desired, the exception being that the mesoscale OGWD and blocking schemes are tied together as one option, and are either both active

or not. The user also has the option to revert back to the GFSv16 (EMC) version of the mesoscale OGWD and blocking schemes. The logical flags are as follows:

gwd_opt	= 2 for unified_ugwp drag package
do_ugwp_v0	activates GFSv16 mesoscale OGWD, blocking and non-stationary GWD – default = .false. for GSL mesoscale OGWD and blocking
do_ugwp_v0_orog_only	activates GFSv16 mesoscale OGWD and blocking – default = .false. for GSL mesoscale OGWD and blocking
do_ugwp_v0_nst_only	activates GFSv16 non-stationary GWD – default = .false. for GSL mesoscale OGWD and blocking
do_ugwp_v1	activates the Version 1 UGWP non-stationary GWD scheme – default = .true.
do_gsl_drag_ls_bl	activates GSL drag suite mesoscale OGWD and blocking – default = .true.
do_gsl_drag_ss	activates GSL drag suite mesoscale OGWD – default = .true.
do_gsl_drag_tofd	activates GSL drag suite turbulent orographic form drag – default = .true.

The namelist variable *cdmbgwd* is a 1-dimensional real array of length 4 that serves as an array of "tuning knobs" to adjust the strength of the mesoscale OGWD and blocking schemes. The value of *cdmbgwd*(1) is a coefficient that multiplies the blocking stress at the reference level given by (2); the value of *cdmbgwd*(2) multiplies the mesoscale OGWD stress at the surface given by (11); and the values of *cdmbgwd*(3) and *cdmbgwd*(4) serve as logical flags and should be set to 1. See the Appendix for a discussion on tuning the blocking and OGWD schemes. The recommended values of *cdmbgwd*(1:2) at various horizontal resolutions are shown in Table 2.

Global Grid Configuration (approx. horizontal grid spacing)	*cdmbgwd*(1)	*cdmbgwd*(2)
C48 (208 km)	40.0	1.77
C96 (104 km)	20.0	2.5
C192 (52 km)	10.0	3.5
C384 (26 km)	5.0	5.0
C768 (13 km)	2.5	7.5
C1152 (8.7 km)	1.67	8.8
C3072 (3.25 km)	0.625	14.1

Table 2. Values of cdmbgwd(1:2) tuning parameters at various global grid resolutions.

The option to output diagnostic data such as the wind tendencies and surface stress from each of the UGWP drag components can be activated by setting the namelist flags *ldiag3d* and *ldiag_ugwp* to true.

The *alpha_fd* namelist variable is the coefficient α in Eq. (17), which can be used to adjust the strength of the turbulent orographic form drag. The default value is 12.

The *cires_ugwp_nml* namelist variables associated with the non-stationary GWD scheme, and recommended settings, are as follows:

knob_ugwp_solver — represents the spectral deterministic solver with background dissipation and spectral saturation
Default value: 2

knob_ugwp_wvspec — four-dimensional integer array that defines number of waves in each azimuthal propagation (as defined by knob_ugwp_azdir) for GWs excited due to the following four sources:

(1) subgrid orography
(2) convective
(3) frontal activity
(4) represents number of waves excited by dynamical imbalances that may mimic both convective and front-jet mechanisms of GW triggering.

knob_ugwp_azdir — four-dimensional integer array that defines number of azimuths for propagation of GWs triggered by four types of physics-based sources (orography, convection, front-jets, and dynamical imbalance). In UGWP v0, first two elements of the array control number of azimuths for OGW and NGWs respectively.

knob_ugwp_stoch — four-dimensional integer array that controls stochastic selection of GWs triggered by four types of physics-based sources.
Default values:0,0,0,0 - reflect deterministic selection of GW parameters without stochastic selection.

knob_ugwp_effac — four-dimensional real array that controls efficiency of GWs triggered by four types of physics-based sources.
Default values: 1.,1.,1.,1. - reflect that calculated GW-tendencies will be applied for the model state.

knob_ugwp_doaxyz — integer parameter that controls application of the momentum deposition for NGW-schemes
0: the momentum tendencies due to NGWs are calculated, but tendencies do not change the horizontal winds
1: default value; changes the horizontal momentum tendencies and horizontal winds

knob_ugwp_doheat — integer parameter that controls application of the heat deposition for NGW-schemes:
0: the temperature tendencies due to NGWs are calculated but tendencies do not change the temperature state
1: default value; changes the temperature tendencies and kinetic temperature

knob_ugwp_dokdis	integer parameter controls application of the eddy diffusion due to instability of NGWs: 0: the eddy diffusion tendencies due to NGWs are calculated but tendencies do not change the model state vector 1: computes eddy diffusion coefficient due to instability of NGWs; in UGWP v0, eddy viscosity, heat conductivity and tracer diffusion are not activated 2: default value; computes eddy diffusion coefficient and applies mixing due to instability of NGWs
knob_ugwp_ndx4lh	integer parameter that controls the selection of the horizontal wavenumber(wavelength) for NGW schemes: 1: selects the 4*dx sub-grid wavelength, where dx is the horizontal resolution of the model configuration (C96-400km; C768-52km)
knob_ugwp_version	integer parameter that selects a version of the UGWP implementation in FV3GFS-127L: 0: old NGW version 1: Version 1 UGWP-NGW – default
knob_ugwp_palaunch	pressure level in Pa for "launch" of NGWs Default: 275.0e2
knob_ugwp_nslope	typical shape of the NGW energy spectrum [variable *s* in Eq. (23)] Default: 1
knob_ugwp_lzmax	maximum vertical wavelength (m) (see Subsection 2.5) Default: 15.75e3
knob_ugwp_lzmin	minimum vertical wavelength (m) (see Subsection 2.5) Default: 0.75e3
knob_ugwp_lzstar	transitional vertical wavelength (see Subsection 2.5) Default: 2.0e3
knob_ugwp_taumin	minimum VMF source amplitude (Pa) Default: 0.25e-3
knob_ugwp_tauamp	VMF source amplitude [variable τ_0 in Eq. (27)]
knob_ugwp_lhmet	horizontal wavelength of NGWs (m) Default: 200.0e3

6.2 Use case configurations

The UGWP suite has some minor configuration requirements, mainly through namelist settings, when running various UFS cases, i.e., global (typically $\gtrapprox$ 13km horizontal grid spacing), regional (e.g., Rapid Refresh Forecast System, or RRFS, typically ≈ 3km horizontal grid spacing), and high-resolution simulations ($\lessapprox$ 1km horizontal grid spacing).

For global simulations, all four orographic drag schemes are typically active. The mesoscale GWD and blocking schemes are automatically tapered at gray-scale resolutions as described in Section 2.1. These two schemes require pre-determined values of the namelist variables *cdmbgwd*(1:2) for the horizontal grid spacing to be used. The optimal values for these will be updated and available in the UFS documentation.

In the convective-allowing regional grids, such as the RRFS, the cdmbgwd values are not used as the mesoscale GWD and blocking schemes are tapered to zero for the 3km grid spacing and below. The small-scale GWD and form drag schemes are active at this resolution, but are automatically tapered to zero at 1 km.

For grid spacings at 1 km and finer, all orographic drag parameterizations will be inactive. For these high resolutions, however, the orographic statistics static files may not be available due to the fact that the high-resolution topographic data used to generate these are on a 1 km grid. In this case, the UFS should be compiled without the UGWP suite by deleting the four lines of the SDF shown above in Section 6.1.

6.3 Sequential updating of column wind profile between mesoscale and small-scale drag calculations

By default, the calculation of the momentum tendencies from each of the five components of the drag suite are based on the current resolved-scale (*u* and *v*) winds, i.e., the winds updated by the dynamical core before the call to the physics parameterizations. To limit the possibility of numerical instability caused by an excess of accumulated wind tendencies by the drag schemes, there is an option to update the winds within the UGWP with the tendencies from the mesoscale GWD and blocking schemes, after they are called, for use by the small-scale GWD and TOFD schemes in the calculation of their tendencies. This option can be selected at compile time with the logical flag *ugwp_seq_update*, which is set in GFS_typedefs.F90 and is *.false*. by default.

6.4 Alternative orographic GWD and blocking configuration: *'do_gwd_opt_psl'* flag

During the development of the prototype GFSv17, it was found that small modifications to the OGWD and blocking schemes led to the improvement of various bulk skill scores, such as the geopotential height anomaly correlation coefficient (Hong et al. 2025). Instead of using the detrended high-resolution topographic dataset for calculating the static orographic input data used by the mesoscale OGWD and blocking schemes, as described in Subsection 4.1, the modified method uses the "raw" non-detrended topographic dataset. This results in generally larger values of the standard deviation (σ) of subgrid topography, especially in plain regions where the topography is well resolved by the model grid. The other scheme modifications are activated by a new namelist option, *do_gwd_opt_psl*, and are listed below:

1) The assumed subgrid topographic blocking height is effectively raised from 2σ to 2σ+*elvmax*, where *elvmax* is the maximum value of the subgrid topography within a model grid cell, which is typically equal to ~3σ. This has the effect of increasing the blocking surface stress and increasing the depth of the blocking layer from 2σ to 3σ.
2) The use of *cdmbgwd*(1) as a runtime tuning coefficient for low-level blocking is removed, and the tuning coefficient *cdmbgwd*(2), which scales the "effective grid length" for the OGWD scheme, is

replaced by a new namelist option *psl_gwd_dx_factor*, which by default is equal to 6.0 for all horizontal grid resolutions. The updated definition of the effective grid length is $0.5*(\Delta x+ \Delta y)$.

3) The blocking height over land ice is reduced from 2σ to σ in order to reduce the blocking drag over ice.

7. Summary and Future Work

The current comprehensive design of the UGWP suite offers a lot of flexibility for tuning but may also give rise to the inherent possibility of double-counting of drag effects due to the simultaneous application of each component, as well as to the limits of model error attribution for each physical process. The small-scale GWD (SSGWD) and mesoscale GWD (MSGWD) components are designed to parameterize the exact same drag physics but for gravity waves of distinct sizes. It is not yet known whether there is a missing gap, i.e., medium-scale gravity wave drag, that may require parameterizing, but the condensing of these drag processes onto a single scheme capable or representing the entire spectrum of gravity wave sizes and their drag effects would be desirable for limiting arbitrary scale cut-offs in the model code and removing the monochromatic assumption used in the current MSGWD component. Furthermore, this would help generalize the UGWP suite to be applied at any scale because the appropriate subgrid-scale gravity waves that need to be parameterized can be automatically set simply by the grid spacing of the application.

Preliminary work has been done to lay out the overall design of a condensed SSGWD/MSGWD into a spectral-GWD component. This scheme utilizes subgrid-scale terrain information that is decomposed by wavenumbers as shown in Fig. 13. The horizontal wavenumber of the terrain affects the sizes of the gravity waves likely to exist in a given model grid column and their size affects the characteristics of vertical propagation -- some wave numbers won't propagate at all, and the ones that do will break at different vertical levels. With a known power (or weighting) associated with each wavenumber, all relevant gravity waves can be represented, but the most significant wavenumbers will dominate the total drag. Similar work is being investigated by others, e.g., Smith and Kruse (2018) and Annelize vanNiekerk at the UKMet Office (personal communication). We intend to work in a collaborative manner with these international colleagues.

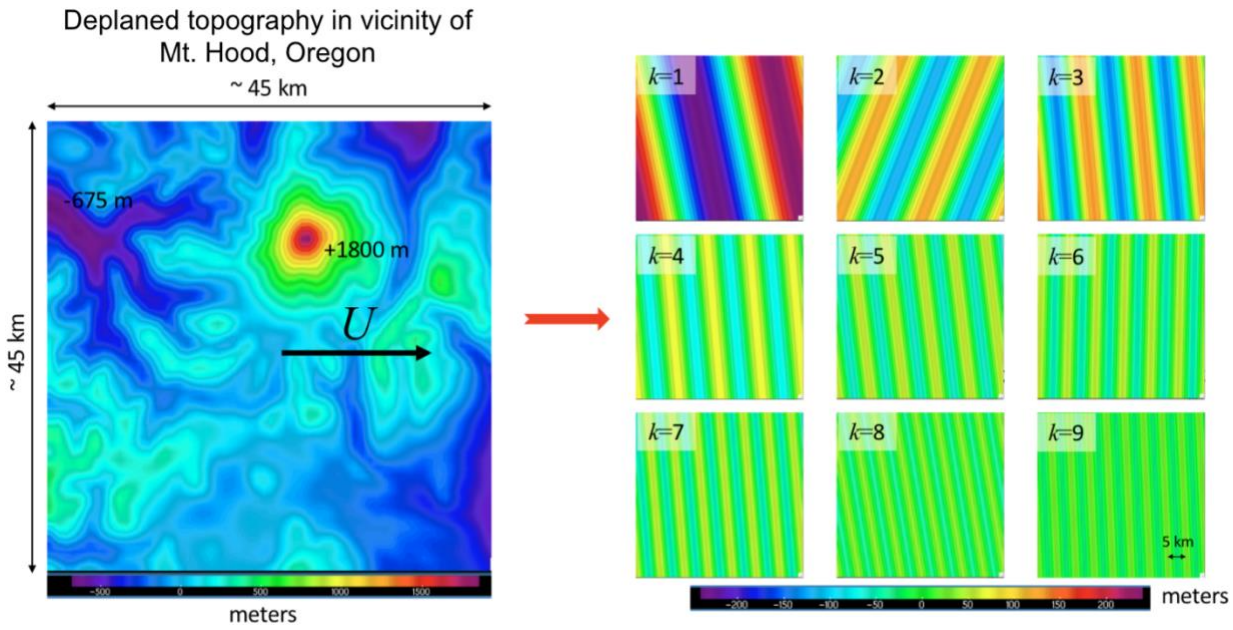


Figure 13. The left panel shows the deplaned topography (total topography minus the "best-fit" plane across the grid cell) which is decomposed by wavenumber (right panel) illustrates how we move from a single-wavenumber representation of subgrid topography to a Fourier series of 2D ridges. Each wavenumber can produce an associated gravity wave that may propagate in the vertical, depending on the atmospheric conditions, but only the most prominent wave numbers will dominate the total drag effects.

Appendix: Tuning the orographic blocking and mesoscale gravity wave drag schemes

The presence of gravity waves is readily detectible in the real atmosphere and in the output of atmospheric models, however, the ability to accurately quantify the wave momentum and energy fluxes in either case is a considerable challenge. The parameterizations presented in this document provide a starting point for determining the momentum fluxes of subgrid-scale GWs, however, additional effort is required to tune the schemes to provide physically reasonable results. The recent study of van Niekerk et al. (2020) provides benchmark orographic GWD and blocking data obtained from high-resolution atmospheric models for the purposes of analyzing and tuning drag parameterizations. Various international modeling centers participated in the benchmark (COORDE) testing, which consisted of 14 consecutive 24-hour global forecasts initiated at 00UTC from 1-14 January 2015. The subgrid blocking and OGWD drag stresses and momentum tendencies from the parameterizations of the various models were provided and were comparable to the corresponding values obtained from high-resolution (~1.8 – 10 km horizontal grid spacing) model runs, which are taken to be the "true" solution.

Figure 14d shows the zonal mean parameterized zonal surface stress of blocking and OGWD from the participating models run at ~80 – 100 km horizontal resolution, and Figs. 15c and 15d show the area-averaged parameterized OGWD zonal momentum tendency from OGWD over the Southern and Northern Hemispheres, respectively, from the models run at ~40 km resolution. These figures served as our guide for matching the corresponding fields produced by our parameterizations. Our starting point was to diagnose the blocking surfaces stresses using the Lott and Miller (1997) scheme implemented in the FV3GFS. We then matched these stresses using the current blocking scheme described in Subsection 2.2 by adjusting the *cdmbgwd*(1) tuning parameter. The resulting zonally averaged surface stress is shown in Fig. 14a for the C384 (~26 km resolution) grid. For the OGWD tuning, we adjusted the *cdmbgwd*(2) parameter until the sum of the blocking and OGWD surface stress matched as closely as possible those shown in Fig. 14d. The result for the C384 grid is shown in Fig. 14c. There is close agreement with our scheme and those of the COORDE participants except over the latitudes ranging from ~ -55° – -35° latitude, where our UGWP OGWD scheme is excessively strong. We are working to check on the validity of our results and to correct the parameterization as needed. At C384 resolution, the results of the tuning exercise provided the recommended values of *cdmbgwd*(1:2) = {5,5}. We also tested at C768 (~13 km grid spacing) (results not shown) to arrive at values of *cdmbgwd*(1:2) = {2.5,7.5}. We extrapolated the *cdmbgwd*(1:2) values for the remaining grid configurations as shown in Table 2.

Figs. 15a and 15b show the vertical profiles of the area-averaged zonal momentum tendencies in the stratosphere. There is close agreement to the corresponding profiles of the COORDE participants shown in Figs. 15c and 15d, which indicates that the tuning efforts based on the surface OGWD stress described above provides realistic zonal flow tendencies at upper levels due to drag.

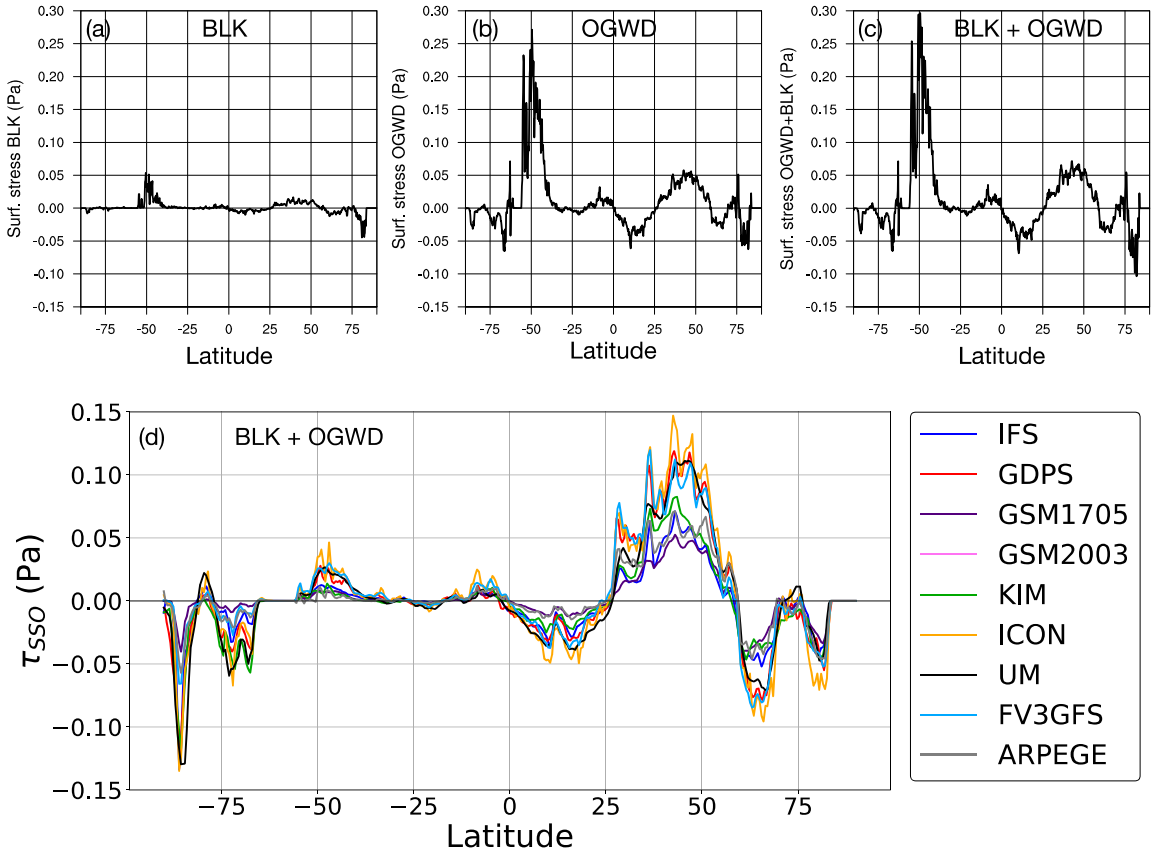


Figure 14. Zonal mean parameterized zonal surface stress (land only) from (a,b,c) the C384 (~26km horiz. resolution) FV3GFS model and from (d) the participating models of the LR CTL (~80-100km horiz. resolution) COORDE experiments of van Niekerk et al. (2020) (see their Fig. 1). The legend on the lower right of the figure lists the participating models. The blocking and OGWD stresses for the FV3GFS are shown in (a) and (b), respectively, and the sum of blocking and OGWD stresses for the FV3GFS and the COORDE experiment models are shown in (c) and (d), respectively.

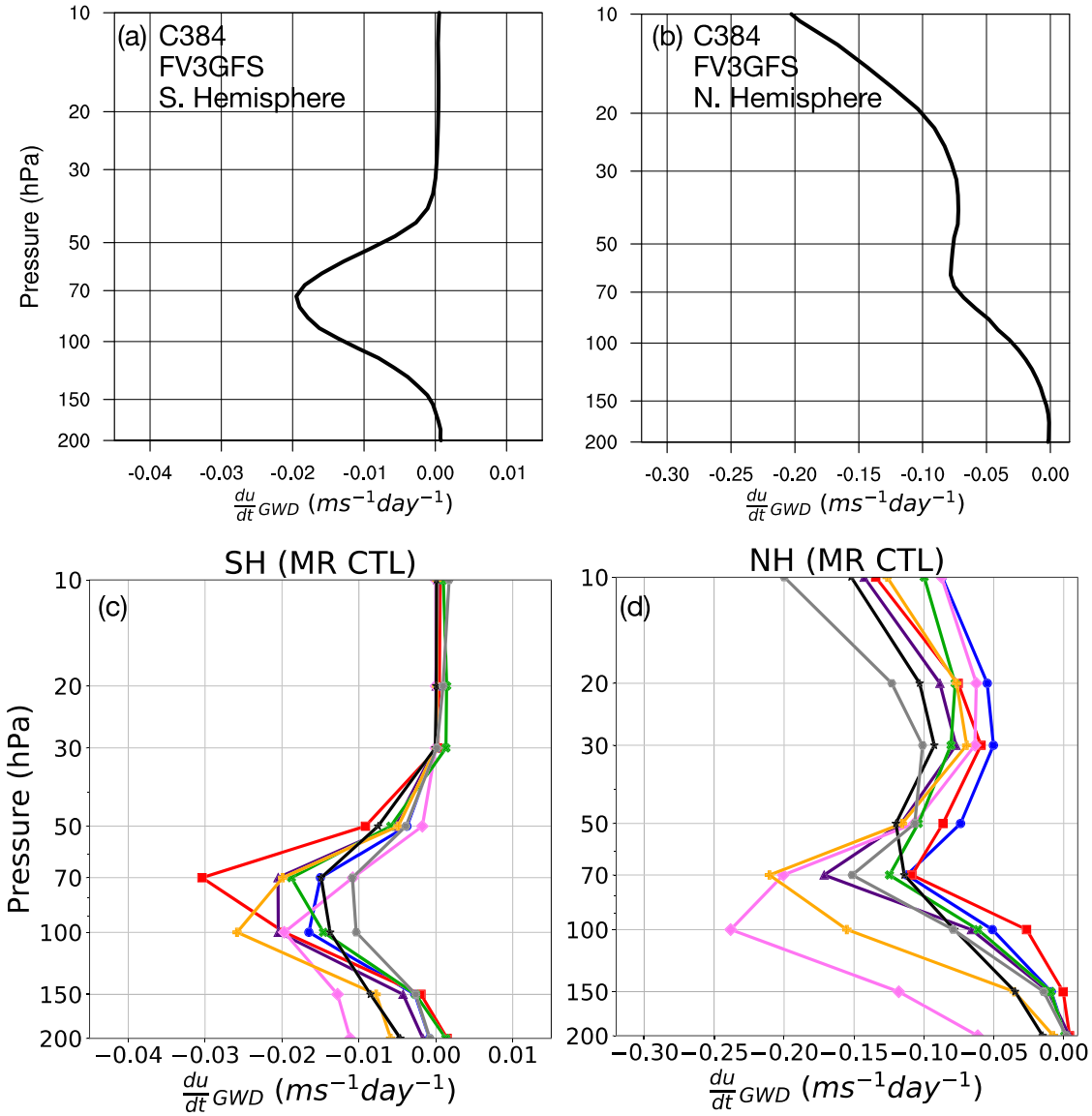


Figure 15. Area-averaged parameterized orographic gravity wave zonal wind drag over (a,c) the Southern Hemisphere and (b,d) the Northern Hemisphere stratosphere from (a,b) the C384 FV3GFS (~26km horiz. resolution) and (c,d) the participating models of the MR CTL (~40km horiz. resolution) COORDE experiments of van Niekerk et al. (2020) (see their Fig. 3). The list of participating models in the COORDE study are the same as in Fig. 14.

References

Alpert, J. C., 2004: Sub-grid scale mountain blocking at NCEP. *16th Conf. on Num. Weather Prediction,* Amer. Meteor. Soc. https://www.researchgate.net/publication/228688611_Sub-grid_scale_mountain_blocking_at_NCEP.

Becker, E., and C. McLandress, 2009: Consistent scale interaction of gravity waves in Doppler spread parameterization. *J. Atmos. Sci.,* **66,** 1434-1449. https://doi.org/10.1175/2008JAS2810.1.

Beljaars, A. C. M., A. R. Brown, and N. Wood, 2004: A new parametrization of turbulent orographic form drag. *Quart. J. Roy. Meteor. Soc.,* **130,** 1327-1347, doi:10.21957/6c5u3bbpk.

Benjamin, S. G., S. S. Weygandt, J. M. Brown, M. Hu, C. R. Alexander, T. G. Smirnova, J. B. Olson, E. P. James, D. C. Dowell, G. A. Grell, H. Lin, S. E. Peckham, T. L. Smith, W. R. Moninger, J. S. Kenyon, and G. S. Manikin, 2016: A North American hourly assimilation and model forecast cycle: the Rapid Refresh. *Mon. Wea. Rev.,* **144**, 1669–1694, doi:10.1175/MWR-D-15-0242.1.

Choi, H.-J., and S.-Y. Hong, 2015: An updated subgrid orographic parameterization for global atmospheric forecast models. *J. Geophys. Res. Atmos*., **120**, 12445-12457, doi:10.1002/2015JD024230.

Danielson, J. J., and D. B. Gesch, 2011: Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011-107, 26 pp, doi:10.3133/ofr20111073.

Fritts, D. C., and T. E. Vanzandt, 1993: Spectral estimates of gravity wave energy and momentum fluxes. Part I: Energy dissipation, acceleration, and constraints. *J. Atmos. Sci.,* **50,** 3685-3694. https://doi.org/10.1175/1520-0469(1993)050%3C3685:SEOGWE%3E2.0.CO;2.

Heinzeller, D., L. R. Bernardet, G. J. Firl, M. Zhang, X. Sun, and M. B. Ek, 2022: The Common Community Physics Package (CCPP) Framework v6, EGUsphere[preprint], https://doi.org/10.5194/egusphere-2022-855.

Holton, J. R., 2004: *An Introduction to Dynamic Meteorology.* 4th ed. Academic Press, 535 pp.

Hong, S.-Y., Choi, H.-J., Chang, E., Park, H., & Kim, Y., 2008: Lower-tropospheric enhancement of gravity wave drag in a global spectral atmospheric forecast model, *Weather and Forecasting*, ***23***(3), 523-531. https://doi.org/10.1175/2007WAF2007030.1

Hong, S.-Y., J.-W. Bao, S. Michelson, E. Grell, M. Toy, J. Olson, J. Han, F. Yang, M.-S. Koo and H.-J. Choi, 2025: Revisions to the subgrid orographic parameterization of the NCEP Global Forecasting System (GFS). *Weather and Forecasting,* (accepted for publication).

Kim, Y.-J., and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J.Atmos. Sci.,* **52,** 1875-1902.

Kim, Y.-J., S. D. Eckermann, and H.-Y. Chun, 2003: An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather prediction models. *Atmos-Ocean,* **41:1,** 65-98, doi:10.3137/ao.410105.

Kim, Y.-J., and J. D. Doyle, 2005: Extension of an orographic-drag parameterization scheme to incorporate orographic anisotropy and flow blocking. *Q. J. R. Meteorol. Soc*., **131**, 1893-1921, doi:10.1256/qj.04.160.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.,* **86,** 9707-9714, doi:10.1029/JC086iC10p09707.

Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. Q. J. R. Meterol. Soc., 123, 101-127, https://doi.org/10.1002/qj.49712353704.

Palmer, T. N., Shutts, G. J. and Swinbank, R., 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity-wave drag parameterization. *Q. J. R. Meteorol. Soc.,* **112,** 1001-1029, doi: 10.1002/qj.49711247406.

Pierrehumbert, R. T., 1986: "An essay of the parameterization of orographic gravity-wave drag". *Seminar/Workshop on Observation, Thory and Modelling or Orographic effects.* Pp. 251-282, https://www.ecmwf.int/node/11673.

Sandu, I., A. van Niekerk, T. G. Shepherd, S. B. Vosper, A. Zadra, J. Bacmeister, A. Beljaars, A. R. Brown, A. Dörnbrack, N. McFarlane, F. Pithan, and G. Svensson, 2019: Impacts of orography on large-scale atmospheric circulation. *Climate and Atmos. Sci.,* doi:10.1038/s41612-019-0065-9.

Scinocca, J. F. and R. Ford, 2000: The nonlinear forcing of large-scale internal gravity waves by stratified shear instability. *J. Atmos. Sci.,* **57,** 653-672, https://doi.org/10.1175/1520-0469(2000)057<0653:TNFOLS>2.0.CO;2.

Scinocca, J. F., 2003: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.,* **60,** 667-682, https://doi.org/10.1175/1520-0469(2003)060%3C0667:AASNGW%3E2.0.CO;2.

Scinocca, J. F., and McFarlane, N. A., 2000: The parameterization of drag induced by stratified flow over anisotropic topography. *Q. J. R. Meteorol. Soc.,* **126,** 2353-2393. doi:10.1002/qj.49712656802.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang, 2019: A description of the Advanced Research WRF Model Version 4.1. NCAR Tech Note, NCAR/TN-556+STR, 145 pp, doi:10.5065/1dfh-6p97.

Smith, R. B., and Kruse, C. G., 2018: A gravity wave drag matrix for complex terrain, *J. Atmos. Sci*., *75*(8), 2599-2613. https://doi.org/10.1175/JAS-D-17-0380.1

Steeneveld, G. J., A. A. M. Holtslag, C. J. Nappo, B. J. H. van de Wiel, and L. Mahrt, 2008: Exploring the possible role of small-scale terrain drag on stable boundary layers over land. J. Appl. Meteor., 47, 2518-2530. doi:10.1175/2008JAMC1816.1.

Tsiringakis, A., G. J. Steeneveld, and A. A. M. Holtslag, 2017: Small-scale orographic gravity wave drag in stable boundary layers and its impact on synoptic systems and near-surface meteorology. *Q. J. R. Meteorol. Soc.,* **143,** 1504-1516. doi:10.1002/qj.3021.

van Niekerk, A., I. Sandu, A. Zadra, E. Bazile, T. Kanehama, M. Köhler, M.-S. Koo, H.-J. Choi, Y. Kuroki, M. D. Toy, S. B. Vosper, V. Yudin, 2020: COnstraining ORographic Drag Effects (COORDE): A model comparison of resolved and parametrized orographic drag. *J. Adv. in Model. Earth Syst. (JAMES),* https://doi.org/10.1029/2020MS002160.

Yudin, V. A., R. A. Akmaev, J. C. Alpert, T. J. Fuller-Rowell, and S. I. Karol, 2018: Gravity wave physics and dynamics in the FV3-based atmosphere models extended into the mesosphere. *25th Conf. on Num. Weather Prediction,* Amer. Meteor. Soc. https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345706.html.

Wedi, N., 2016: Experience in creating orography ancillary files (presentation). European Center for Medium-Range Weather Forecasts (ECMWF). https://www.ecmwf.int/node/16661.

Zhang, M., L. R. Bernardet, G. Firl, D. Heinzeller, and W. Li, 2022: CCPP v6.0.0 Physics Scientific Documentation, https://dtcenter.ucar.edu/GMTB/v6.0.0/sci_doc/index.html.