Can We Improve the Realism of Gravity Wave Parameterizations by Imposing Sources at All Altitudes in the Atmosphere?

B. Ribstein¹, C. Millet¹,², F. Lott¹, and A. de la Cámara⁴

¹CMLA, ENS Cachan, CNRS, Université Paris-Saclay, Cachan, France, ²CEA, DAM, DIF, Arpajon, France, ³LMD, PSL Research Institute, Paris, France, ⁴Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid, Madrid, Spain

Abstract A multiwave non-orographic gravity wave (GW) scheme is adapted to represent waves of small intrinsic phase speed, inertial waves, and wave emission from all altitudes. This last change removes the launching altitude parameter, an arbitrary parameter systematically used in GW schemes. In offline calculations using reanalysis fields, these changes impose larger amplitude, saturated waves everywhere in the middle atmosphere, which produces more realistic GW vertical spectra than in previous configurations. The same scheme, tested online in the Laboratoire de Météorologie Dynamique Zoom (LMDz) general circulation model, performs at least as well as the operational non-orographic GW scheme. Some modest benefits are seen, for instance, in the equatorial tilt with altitude of the winter jets in the middle atmosphere. Although the scheme includes the effects of inertial waves, which are detected in the mesosphere by different observational platforms, the configuration that gives a reasonable climatology in LMDz hinders their vertical propagation and limits their presence at mesospheric altitudes.

Plain Language Summary Gravity waves are fluctuations in the atmosphere (seen in the temperature, wind velocity, and pressure fields) that transport energy and momentum from their sources in the troposphere and lower stratosphere to their sinks in the middle atmosphere. This way they exert a profound influence on the global circulation. Due to their relative small spatial scales, atmospheric general circulation models do not explicitly resolve these waves, and their effects on the circulation resolved by the model need to be parameterized. Parameterizations of gravity waves generated by fronts and flow imbalances typically assume that wave sources are located at a certain vertical level in the troposphere, which is easy to implement but neglects the fact that these processes can occur at all altitudes in the atmosphere. In this study, we explore to which extent parameterizations of gravity wave due to fronts and flow imbalances can be improved by allowing waves to be emitted from all model levels. Our results show evidence of modest corrections of some model biases, and a clear improvement in the parameterized gravity waves energy spectra.

1. Introduction

Atmospheric gravity waves (GWs) have long been observed with radio soundings (Tsuda et al., 1994), radar (Love & Murphy, 2016; Shibuya & Sato, 2019), lidar (Baumgarten et al., 2016; Khaykin et al., 2015), and satellite instruments (Alexander, 2015). Their amplitude grows as they propagate vertically and they impact the middle atmosphere circulation when they break and/or dissipate, even if they have relatively small amplitudes at the source level. The spatial scales of GWs are generally too small to be resolved by general circulation models (GCMs), and the generation, propagation and dissipation of these waves need to be parameterized for GCMs to produce a reasonable circulation. Such parameterizations were first introduced in the 1980s in models with a barely resolved stratosphere, and only high-amplitude orographically forced GWs were needed to be taken into account (Palmer et al., 1986). Nowadays many models resolve the middle atmosphere, requiring the parameterization of the effects of smaller-amplitude, non-orographic gravity waves (Manzini et al., 1997).

One way to parameterize non-orographic GWs consists in using the observational evidence of “universal” spectra built over a large number of realizations of GW fluctuations of vertical wind and temperature in the middle atmosphere, which shape is derived from radiosondes and satellite data (e.g., Cot, 2001; Zhao et al., 2017). These spectra are numerically robust (Brethouwer et al., 2007; Lindborg, 2006), and various theories have been developed to explain them. Some involve wave breaking (e.g., Dewan & Good, 1986), other include nonlinear...
effects such as triade interactions, Doppler spreading, or inverse cascades (e.g., Broutman & Young, 1986; C. Hines, 1996; Lilly, 1983; Métai et al., 1996). Beyond the theoretical debate about the reason of the observed spectral shape, a practical result is that the existence of a saturated spectra allows semi-theoretical integrations that ease the parameterization of non-orographic GWs (C. O. Hines, 1997; Manzini et al., 1997; Warner & McIntyre, 1996). To a certain extent, this approach is challenged by recent balloon observations showing that the GW field is very intermittent and often dominated by rather well-defined GW packets (Alexander, 2015; Hertzog et al., 2008; Wright et al., 2013). This intermittent nature makes that in each model gridbox and at each time step, the number of GW packets should not be large enough to fulfill the law of large numbers underlying the construction of spectra out of individual realizations. More generally, this concern is related to that of statistical equilibrium and is central in the recent development of stochastic parameterizations (Berner et al., 2017). For GWs, this implies that the “universal” spectral shape should be checked a posteriori and over a large number of days, rather than being realized every time. This leads to an alternative approach to parameterize non-orographic GWs, which is based on representing the GW field with a Fourier series in the horizontal and temporal directions (e.g., Alexander & Dunkerton, 1999). In this approach the individual harmonics are a crude representation of the individual wave packets, and the intermittency is taken into account by launching stochastically a few harmonics each timestep (Eckermann, 2011; Lott, Guez, & Maury, 2012). The challenge is then to reconcile the two types of schemes (“spectral” vs. “multiwave”) and the two types of observations (stationary universal vertical spectra vs. intermittent wave packets). As we shall see, this can be done by showing that the ensemble average of the periodograms associated with superposed harmonics can reproduce the observational “universal” spectra (de la Camara et al., 2014). In route to realize this objective, the study of Souprayen et al. (2001) is encouraging. They use ray-tracing reconstructions of the GW field and show that the wave filtering by the large scale flow and the breaking of individual GW packets can yield realistic spectra in the upper stratosphere and mesosphere.

In some GCMs the amplitude of the parameterized GWs is related to their non-orographic sources, that is, convection and fronts and/or flow imbalance, and the GWs are launched from a single source level in the troposphere. This last choice is not well justified for several reasons. One is that the presence of unbalanced flows that emit GWs is not restricted to the troposphere: Dörnbrack et al. (2018) and Sato and Yoshiki (2008) provide observational evidence for GW generation in the stratospheric polar vortex, while Polichtchouk and Scott (2020) discuss the GW generation at the critical layer of the stratospheric polar night jet using an idealized numerical model. A second reason is that observations often show that the GWs in the middle atmosphere have rather long periods (often around 6 hr and longer in Reichert et al. (2019), near the inertial period in Gelinas et al. (2012), Bellenger et al. (2017), Shibuya and Sato (2019), and Vincent and Alexander (2020)). The presence of these slow waves is difficult to justify if the GW sources are only in the troposphere, as waves with small intrinsic frequency have short vertical wavelengths and dissipate more easily as they propagate upwards than faster waves. This process is often referred to as dynamical filtering and occurs systematically when the waves approach a critical level. For GWs, this dynamical filtering is very effective, explaining most of the relation between the wind speed and GW amplitude in balloon measurements (Plougonven et al., 2017).

The goal of this paper is to present modifications in the “multiwave” non-orographic GW parameterization due to fronts and jets of de la Camara and Lott (2015) to include sources at all levels, small intrinsic phase speed waves (including near inertial waves), and analyze if these modifications can help reproduce the universal spectral shape systematically enforced in the “spectral schemes.” Although some of these concerns could be addressed with any other scheme, this one has some characteristics that make it suitable to treat all of them. The first and most important is that this scheme is based on a spontaneous emission theory (Lott, Plougonven, & Vanneste, 2012), which is a theory that partly explains the GW emission seen in quite sophisticated high resolution simulations (e.g., Polichtchouk & Scott, 2020), and which is not limited to the troposphere. The second is that it points to potential vorticity (PV) anomalies as a source of GWs, which is coherent with the fact that processes such as classical geostrophic adjustment or re-emission are associated with the presence of PV anomalies. The third is that it is operational in the Laboratoire de Météorologie Dynamique Zoom (LMDz) GCM, which is the atmospheric component of the Institute Pierre Simon Laplace (IPSL) Earth system model (Houardin et al., 2013, 2020), so it is routinely tested. The fourth is about observational constraints, in the sense that it qualitatively produces the observed intermittency of the non-orographic GW field (Alexander, 2015; de la Camara et al., 2014; Hertzog et al., 2012; Wright et al., 2013). Indeed, de la Camara, Lott, Jewtoukoff, et al. (2016) demonstrated that a good representation of the GW intermittency can be beneficial for models, helping the IPSL model to better simulate the timing of the Southern Hemisphere stratospheric final warming.
A final (and much less positive) reason is that in its operational version, de la Cámara and Lott (2015) use a Gaussian distribution of intrinsic phase speeds with standard deviation near 40 ms⁻¹. This large value is used because it helps the waves to propagate up to the upper mesosphere without attenuation, but it contradicts the fact that frontal waves resulting from spontaneous adjustment have small intrinsic phase speeds near their source (Lott, Plougonven, & Vanneste, 2012).

A central assumption made in this paper is that this bias toward larger than expected intrinsic phase speeds is common to other schemes, and that trying to correct it in one scheme could be indicative of what could be done in other schemes. In fact the corrections we test in this study have a general character: we include the Coriolis force because it can be significant at low intrinsic frequencies, and we place sources at all levels rather than at one fixed level in the lower troposphere. Finally, it is worth noticing that the shortcomings we try to deal with are considered priorities in the community (see discussion about low phase speed waves in Alexander et al. (2021)), but are not the only ones. Some authors place more emphasis on including three-dimensional propagation of gravity waves in parameterizations (e.g., Amemiya & Sato, 2016; Muraschko et al., 2015; Ribstein & Achatz, 2016). Including lateral propagation in highly parallelized code is extremely challenging computationally, so it is worthwhile to test if some improvements can be obtained through other routes. The remainder of the paper is structured as follows. Section 2 describes the modifications we propose to the frontal GW parameterization used in LMDz. In Section 3, offline tests are performed to estimate the characteristic altitudes of the sources and to test if the scheme produces realistic vertical energy spectra. Section 4 presents online results obtained with LMDz, using the standard parameterization and the updates. Section 5 gives the main conclusions.

2. Non-Orographic Gravity Waves Due To Fronts and Flow Imbalance

2.1. General Formalism

We next summarize the formalism of the stochastic parameterization used in LMDz (de la Cámara & Lott, 2015), and emphasize the modifications introduced in this study. The horizontal wind and temperature disturbances (\(\mathbf{u}'\), \(T'\)) due to GWs are represented by a stochastic Fourier series of \(J\) monochromatic waves,

\[
(\mathbf{u}', T') (x, z, t) = \sum_{j=1}^{J} C_j (\hat{u}_j(z), \hat{T}_j(z)) e^{i(k_j \cdot x - \omega_j z)}
\]  

whose horizontal wavevector \(k_j\) and absolute frequency \(\omega_j\) are chosen randomly, and the complex amplitudes \((\hat{u}_j(z), \hat{T}_j(z))\) vary in the vertical direction measured by the log-pressure coordinate \(z\). The intermittency coefficient \(C_j\) measures the probability of the presence of the corresponding wave at a given horizontal location within the gridbox. In previous versions of the parameterization we had always assumed equiprobability for simplicity and taken

\[
\sum_{j=1}^{J} C_j^2 = 1, \quad \text{i.e.} \quad C_j = 1/\sqrt{J}.
\]  

as we shall see, the value of this parameter can be changed to represent sources located at different vertical levels.

To evaluate the wave amplitude, de la Cámara and Lott (2015) adapts Lott, Plougonven, and Vanneste (2012)'s analytical estimate of the GW momentum flux emitted by a potential vorticity (PV) anomaly in a vertically sheared flow. More specifically, they consider that at each model level \(z_l\), the PV disturbance, \(pq'\) in Lott, Plougonven, and Vanneste (2012)'s notation, compares in amplitude with the gridscale vertical stratification times the gridscale relative vorticity, \(\theta_l \zeta\), where \(\theta\) is potential temperature and \(\zeta = \partial \nu - \partial \theta u\) is the relative vorticity. They also consider that the thickness of the PV disturbance has an intermediate value between the model level thickness \(dz\) and a fixed characteristic vertical scale \(\Delta z\). With these parameters, linear f-plane theory can be used to predict the emitted Eliassen-Palm (EP) flux at \(z_l\) (see Equation (3) in (de la Cámara & Lott, 2015)), which we modify into
\[ F(z_i) = G_0 \rho(z_i) \frac{\Delta z d z_i}{4 f_0} N(z_i) \left[ f \tanh \left( \frac{\zeta(z_i)}{f} \right) \right]^2 e^{-\frac{N(z_i)}{N(t)}}. \]  

where \( U_z = |\mathbf{U}| \) is the modulus of the vertical shear, \( N \) is the Brunt-Väisälä frequency, \( \rho = \rho e^{-z/H} \) is air density, with \( \rho_1 = 1 \text{ kg m}^{-3} \) and \( H = 7 \text{ km} \) being the characteristic vertical scale of atmospheric density decay. Still in Equation 3, \( G_0 \) is a tuning parameter, \( f_0 = 10^{-3} \text{s}^{-1} \) is a characteristic value of \( f \), and \( \Delta z \) a characteristic depth of the subgrid-scale PV anomaly. Compared to de la Cámara and Lott (2015), the only change at this stage is the hyperbolic tangent function that is used to limit the relative vorticity to values below \( |f| \). This is a reasonable assumption since flows with relative vorticity larger than \( |f| \) are likely to be strongly unstable, and it also helps moderate the emission intensity in the tropics. For completeness, we recall that the introduction of a characteristic value \( f_0 \) at the numerator in 3, rather than dividing by \( |f| \), was also made in de la Cámara and Lott (2015) to limit emission in the tropics.

de la Cámara and Lott (2015) assume that the GW momentum flux are essentially generated in the troposphere, but integrate Equation 3 over all model levels \( L \) to calculate a total emitted flux,

\[ F = \sum_{i=1}^{L} F(z_i). \]  

it is then imposed that this momentum flux radiates from a specified launching level \( z_{la} \) in the troposphere, distributed over the ensemble of \( J \) monochromatic waves in (1). Among the randomly chosen parameters, the direction of each harmonic \( \mathbf{F}_j \) of the total EP flux \( F \) is specified through the random horizontal wavevector \( \mathbf{k}_j \), following the rule

\[ \mathbf{F}_j(z_{la}) = -\frac{\mathbf{k}_j}{||\mathbf{k}_j||} F. \]  

(5)

to derive (5) from (4) we have chosen by convention that the sign of the intrinsic phase speed at the launch level

\[ \hat{\omega}_j(z_{la}) = \omega_j - \mathbf{k}_j \cdot \mathbf{U}(z_{la}) > 0. \]  

(6)

to evaluate the vertical profile of the flux above the launching level we consider that from one vertical level to the next above, the flux is (a) reduced by a small diffusivity \( \mu \rho \hat{\omega}_j(z) \), (b) limited by that of a saturated wave (e.g., Lindzen & Schoeberl, 1982), and (c) set to zero immediately above an inertial level (Lott et al., 2015):

\[ \mathbf{F}_j(z_{ls}) = -\frac{\mathbf{k}_j}{||\mathbf{k}_j||} \Theta \left( \hat{\omega}_j(z_{ls})^2 - f^2 \right) \times \min \left\{ ||\mathbf{F}_j(z)|| \exp \left( \frac{2 \mu m_j^2(z_i)}{\rho(z_i) \hat{\omega}_j(z_i)} d z_i \right), \rho(z_{ls}) S_j z_{ls}^2 \right\} \frac{k_{min}^2}{||\mathbf{k}_j||}. \]  

(7)

In Equation 7, \( \Theta \) is the Heaviside function to handle inertial levels, \( k_{min} \) is the horizontal wave number associated with the largest resolved GW in the model (it is related to the model’s horizontal resolution), and \( S_j \) is a tunable parameter that controls the saturation amplitude. Still in Equation 7, the vertical wavenumber \( m_j \) and intrinsic frequency \( \omega_j \) are given by

\[ m_j(z) = -\frac{N(z) ||\mathbf{k}_j||}{\sqrt{\hat{\omega}_j(z)^2 - f^2}}, \text{ where } \hat{\omega}_j(z) = \omega_j - \mathbf{k}_j \cdot \mathbf{U}(z), \]  

(8)

where the minus sign in the definition of \( m_j \) ensures upward propagation above the launching level. A novelty of the present study is that we have included the Coriolis force in Equations 7 and 8.

At this stage, the emitted flux is equidistributed among all possible horizontal directions, which somehow contradicts the theory saying that the preferential emission is in the direction opposite to the wind shear. In practice, however, waves emitted with phase speeds in the direction of the shear get their intrinsic phase speeds decreased and their vertical wave numbers increased as they are evaluated at the next vertical level above. According to (7)
the fluxes for these waves are more strongly reduced than those for the waves in the direction opposed to the wind shear. In the scheme, we exclude the emission of highly saturated waves, which avoids imposing a huge drag just above the launching level, by systematically replacing the launching value of the flux by its value at the level above (in $z_{La+1} = z_{La} + dz_{La}$). This naturally tends to reduce the emission of waves with phase speeds in the direction of the shear. Finally, below the launching altitude the fluxes are kept constant, which technically allows to define $F_j(z)$ at all model levels. This last choice is consistent with the fact that in the model we do not extract momentum from the source's surroundings region to balance the emitted wave stress, but it is a shortcoming that may be tested in the future. Nevertheless, in the next section we will see that this choice is justified, to a certain extent, by the fact that GWs are mainly produced in the troposphere and lower stratosphere, and induce accelerations of the order of few tens of m/s/day in the mesosphere. Applied to the troposphere, the scaling by density makes that the opposed accelerations would be about 2–3 orders of magnitude smaller than those in the mesosphere. Since 10 m/s/day is also a characteristic scale for the divergence of the EP flux due to the large scale eddies in the troposphere (Gastineau et al., 2020), we see that the neglected terms are likely to be small.

Once $F_j(z)$ is evaluated at all vertical levels, we use the Wentzel-Kramers-Brillouin (WKB) formalism to relate the EP flux to the disturbance fields in (1):

$$\hat{u}_j(z) = -\frac{k_j\tilde{\omega}_j}{\omega_j^2 - f^2} \phi_j(z) e^{i\beta_j m_2(z)dz + i\xi_j},$$

$$\hat{T}_j(z) = i m_j \frac{H}{R} \phi_j(z) e^{i\beta_j m_2(z)dz + i\xi_j},$$

(9)

(10)

where the amplitude of the geopotential is

$$\phi_j(z) = \sqrt{\frac{\|F_j\|N^2}{\rho m_j k_j \|k_j\|}} e^{i\xi_j},$$

(11)

and $\xi_j$ is a phase with no effect on the EP flux, which is chosen randomly when computing the physical fields offline. For completeness, note that the GW drag is computed as

$$\rho \frac{\partial u}{\partial t} \big|_{GW} = \sum_{j=1}^{J} C_{ij} \frac{\partial u}{\partial t} \big|_{j} = \sum_{j=1}^{J} C_{ij} \frac{\partial F_j}{\partial z}.$$  

(12)

Thermal effects are taken into account by evaluating the work performed against the wind:

$$\frac{\partial T}{\partial t} \big|_{GW} = -u \cdot \frac{\partial u}{\partial t} \big|_{GW}/C_P,$$

(13)

where $C_P$ is the heat capacity of dry air.

At this point, we have applied the described equations to $J$ waves emitted at a single launching level, $z_{La}$. To adapt our formalism in order to work with $J$ waves emitted from different launching levels we choose the launching level randomly, $z_j$, together with the horizontal wavevector $k_j$ and intrinsic frequency $\tilde{\omega}_j$ (see Figure 1). Then the launching flux in 5 is replaced by

$$F_j(z_j) = -\frac{k_j}{\|k_j\|} F_j(z_j),$$

(14)
the emitted amplitude $F$ at the corresponding level being directly given by $3$. The various profiles are then evaluated above and below $z_j$ following Equations 7–11, but with $z_{j+1}$ replaced by $z_j$ for each waves. But now that we have $J$ different launching altitudes, with the possibility that $J > L$, we need to take for intermittency parameter:

$$C_j^2 = L/J.$$  \hspace{1cm} (15)

### 2.2. Model and Reanalysis

In offline mode we feed the parameterization with daily fields of temperature and horizontal wind velocity from the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA2; Gelaro et al., 2017) at $1^\circ \times 1^\circ$ resolution and from the ground to 0.01 hPa. The simulations are forced with the observed seasonal cycle of sea surface temperatures and sea-ice from the CMIP5 database for the period 1980–1995, and the ozone climatology is built from the ACC/SPARC ozone database. All runs have the same settings of the orographic and convective GW parameterizations (Lott & Guez, 2013; Lott & Miller, 1997), and the different settings of the parameterization of GWs due to fronts and jet imbalances are detailed in the next subsection. The reader is referred to Hourdin et al. (2020) for a comprehensive description of the LMDz model equations and specifications of its grid.

### 2.3. Gravity Waves Parameterization Setup

In all experiments discussed in this paper, we consider $J = 48$ waves each physical time step in the model as well as in the offline reconstructions. We also choose the horizontal wavenumber amplitude $k_j = ||k_j||$ with uniform probability in the interval $k_{min} \leq k_j \leq k_{max}$, with $2\pi k_{min} = 6.3$ km and $2\pi k_{max} = 315$ km crudely bounding the smallest horizontal wavelengths that can be attributed to GWs, and the largest horizontal wavelength that cannot be represented in the model gridbox. The direction of horizontal propagation $\theta_j (k_j = k_j (\cos \theta_j i + \sin \theta_j j))$ is chosen uniformly within the interval $0 \leq \theta_j \leq 2\pi$. The attribution of frequency is indirect since we first select the wave intrinsic phase speed at the launch level $z_j$ from a half-normal distribution with a standard deviation of $c_v$, and in the direction given by $k_j$. The parameter $c_v$ is key, and tuning it in different experimental setups requires adjustments in the launching momentum flux amplitude and saturation parameters $G_0$ and $S_j$ in Equations 3 and 7 respectively.

In this study we present results from three different setups of the non-orographic GW parameterization. The strategy adopted consists of 1) decreasing the phase speed drastically, and 2) introduce multiple level sources adapting the other parameters so the overall model performance compares well the existing operational scheme. We therefore target rather neutral effects on the model climate, which is in itself a task that demands a substantial amount of trial simulations.

In the first experimental setup, we proceed as in de la Cámara and Lott (2015) and choose the launching altitude at $z_j = 500$ m, and take $c_v = 50$ m s$^{-1}$, $G_0 = 4$, and $S_j = 0.6$. These values stay reasonably close to those used in previous studies (e.g., de la Cámara et al., 2016), considering that we now include the Coriolis force and bound the disturbance vorticity to values below $f$. If we consider that the characteristic vertical scale of the waves produced with this setup is $2\pi m \approx 2\pi c_v/N \approx 20$ km, we see that we are essentially taking into account long waves (since $c_v$ is large), and we will refer to this setup as LW.

In the second experimental setup, we keep a single launching level but decrease the intrinsic phase speed down to $c_v = 10$ ms$^{-1}$. This requires a slight increase in $G_0$ up to $G_0 = 5$, but since the characteristic vertical scale is now much shorter, that is, $2\pi c_v/N \approx 4$ km, the saturation parameters need to be increased substantially up to $S_j = 2.5$ to keep the saturated flux of about the same amplitude within the middle atmosphere. Note that the characteristic vertical scale is closer to the wavenumber $m*$ introduced by Warner and McIntyre (1996). These changes in the wave parameters imply parameterized waves of shorter vertical scales, so this setup will be referred to as SW-1L.

In the third experimental setup, we consider waves emitted from multiple model levels. Nevertheless we realized that as the amplitude of emission in Equation 3 is very sensitive to the Richardson number and to the relative
vorticity $\zeta$, there is often one level that dominates the sum of EP fluxes in Equation 5. More precisely, and in a statistical sense, the momentum flux $A(z_l)$ at the level where it is maximum typically represents a third of the total sum over the column in 4. Finding this vertical level out of $J$ cases is comparable in amplitude to average over $J$ levels. This, combined with fact that launching waves at high levels reduces dynamical filtering, allows us to work with values for $G_0$ and $S_c$ that are close to the previous case. We take $G_0 = 4.2$ and $S_c = 3$ in this setup, which will be referred to as SW-ML.

3. Offline Results

The performance of the three configurations LW, SW-1L, and SW-ML is addressed in off-line runs using the meteorological fields from MERRA2 as input. Figures 2a and 2b show the spatial distribution on a given day (22 January 2012) of the daily average of the characteristic amplitude of the emitted EP flux in SW-ML at the 500 hPa level. Specifically, we show the daily average of

$$\sum_{j=1}^{J} C_j^2 |F_j(z_{j+1})|,$$

where the level index $j + 1$ again indicates that the emitted flux has been corrected to stay below its saturated value just above the launching level (see Section 2.1). Peak values of around 50 mPa are found in the vicinity of the subtropical jets in both hemispheres, reflecting the direct relation between the emitted EP flux and the grid-scale relative vorticity, as indicated by Equation 3. Figure 2c compares the latitudinal distribution of the zonal-mean EP flux for LW, SW-1L, and SW-ML. The three curves present very similar features with very low values in the tropics, a maximum around the subtropical jets, and a gradual descent toward the poles. In general, we see that SW-ML produces slightly larger fluxes at all latitudes than the other two setups, and that SW-1L consistently produces smaller fluxes. It is important to emphasize that although the amplitude of the EP flux is calculated deterministically, the probability density function of the launched fluxes qualitatively follows the observed log-normal distribution in the lower stratosphere (not shown, see de la Cámara and Lott (2015)).

To further analyze which vertical levels emit the waves that contribute the most to the GW drag, Figure 3 presents regressions of $F(z_l)$ in Equation 3 on the extrema in gravity wave drag amplitude: $\max_{l:z_l \leq z} \| H(z_l) \|_{1}(z_l)$, the regression being averaged over each individual wave. The waves contributing most to the gravity wave drag are issued from a large variety of vertical levels spanning the entire troposphere up to the tropopause and lowermost stratosphere (Figure 3a). On the poleward side of the jet there are two sectors of preferential emission, the surface and the mid-troposphere. The mid-troposphere region of emission extends up to the tropopause and into the lowermost
stratosphere, that is, up to \( z = 12 \) km around 40°N. This emission from the lowermost stratosphere becomes predominant on the equatorial flank of the jets with extrema in sources located between 8 and 16 km typically.

One of the purposes of radiating waves at all altitudes was to reduce the distance between the source regions and the regions where the gravity waves contribute most to the large scale dynamics. However, the comparison of source regions and the gravity wave drag in Figure 3 illustrates that we have not succeeded; the source regions are in the troposphere and lower stratosphere, and the maxima of GW drag occurs in the mesosphere. This suggests that we could have limited our search for emission levels to the troposphere and lowermost stratosphere. Nevertheless, the large spread of emission levels, plus the fact that one level dominates at a given place, confort the idea that the emission level for GWs should be determined at each time and horizontal location rather than imposed.

Next we analyze the ability of LW, SW-IL and SW-ML to reproduce the empirical GW energy spectra. The vertical profiles of wind and temperature are obtained from the parameterized profile of vertical EP flux as in Equations 9–11. More specifically, in order to build the spectra out of individual realizations we first construct 1000 monochromatic waves, each corresponding to randomly chosen values of phase speed and horizontal wave-number. The wind and temperature profiles of each wave are then sampled every 100m, which is a much higher vertical resolution than that used in MERRA2. We therefore linearly interpolate the EP flux to the target vertical grid, together with the large scale fields of wind and temperature needed in Equations 9 and 10. To construct one realization out of these 1000 monochromatic waves, we pick \( J = 48 \) of them randomly, choose the phase \( \xi \) of each randomly and then sum over the \( J \) waves. Let \( u'(z) \) be one such realization of the horizontal wind disturbance. We perform a Fourier analysis of this realization, which gives the periodogram \( \hat{u}' \hat{u}'^* \), \( \hat{u} \) being the Fourier coefficients and the stars indicating conjugation. To minimize the Gibbs effect, a tapered cosine window is applied in the vertical to all data before the fast Fourier transform. The spectra are then estimated at each horizontal location by performing an ensemble average of individual periodograms \( \langle \hat{u}' \hat{u}'^* \rangle \), constructing an ensemble out of 20 independent realizations. We have found our results about the spectral shape not very sensitive to the specific procedure to compute the spectra. Other possibilities would have been to average the individual periodograms at a given place over different days, or average the periodograms at different places over the same day. The adopted procedure allows to compare without ambiguity locations with different intensity of GW emission.

Detailed results are presented for locations A and B on 22 January 2012 (Figures 2a and 2b), which are characterized by a marked contrast in the GW sources and the vertical propagation conditions: location A is a place where the emitted momentum flux is large and the middle atmospheric winds above are predominantly positive (see Figures 2a and 3a), while location B is a place of small emission and the middle atmosphere winds are predominantly negative (see Figures 2b and 3a). In the following we restrict the discussion to the horizontal wind spectra, but we have verified that the temperature spectra exhibit a similar shape (not shown). The spectra for altitudes

Figure 3. Wave by wave regression of the source term on the max of GWD (shading, color interval 0.5 mPa). (a) Position with respect to the troposphere and lower stratosphere zonal wind (in m s\(^{-1}\)). (b) Position with respect to the gravity wave drag (in m s\(^{-1}\) d\(^{-1}\)). Zonal and temporal average over January 2012.
lower than 65 km are displayed in Figure 4 at the two specific locations A and B. For both locations the energy spectra obtained with the configurations LW and SW-1L are proportional to $m^{-4.5}$, but the shape obtained with vertically distributed GW sources (i.e., SW-ML) is characterized by a $m^{-3}$ tail, which is suggestive that saturation occurs much more systematically under this configuration (e.g., Dewan & Good, 1986). Besides, the vertical spectra for SW-1L and SW-ML are shifted toward smaller wavelengths, which is consistent with the fact that GWs are launched with slower phase velocities.

To summarize, when smaller intrinsic phase speed are imposed (i.e., in SW-1L and SW-ML), there is more energy concentrated at shorter wavelengths, this energy corresponds more often to saturated waves, and the effect of the saturation on decreasing the wave amplitude is compensated in SW-ML by launching waves from all altitudes. These results demonstrate that multiwave schemes with small intrinsic phase speed can fairly reproduce the observed GW energy spectra, bridging a gap with the spectral schemes that include sources but prescribe a saturated spectrum (e.g., Bushell et al., 2015).

4. Impacts on the Simulation of the Stratosphere

We next evaluate the performance of the three configurations of the frontal GW parameterization in 15-year runs with the climate model LMDz. Figures 5a and 5b show latitude-height cross-sections of zonal-mean zonal wind in LW-LMDz (black contours), seasonally averaged for December-January-February (DJF) and June-July-August (JJA). Well-known features of the zonal mean structure of the troposphere and middle atmosphere are captured in LW-LMDz. The subtropical jets in the upper troposphere are stronger in the winter than in the summer hemisphere, and displaced further poleward in the summer hemisphere. In the middle atmosphere, there are westerly winds in winter (i.e., the polar night jet) and easterlies in summer. The color shading in Figures 5a and 5b represents the difference of zonal wind between MERRA2 (period 1996–2010) and LW-LMDz. In the tropics, there are differences above 40-km height in all seasons that are related to the representation of the amplitude of the semiannual oscillation (Lott & Guez, 2013; Smith et al., 2017). In the extratropics, the strongest bias takes place in the upper stratosphere and mesosphere in JJA in the SH (Figure 5b), where the weaker winds in LW-LMDz than in MERRA2 are related to a polar night jet in the model that does not tilt equatorward with height as compared to reanalysis. In the NH in DJF (Figure 5a), the westerly winds in LW-LMDz are also systematically weaker than in MERRA2, which implies a weaker polar vortex.

The middle and bottom panels of Figure 5 show the corresponding zonal wind profiles (black contours) for SW-1L-LMDz (Figures 5c and 5d) and SW-ML-LMDz (Figures 5e and 5f), with the color shading displaying differences with respect to LW-LMDz. Both SW-1L-LMDz and SW-ML-LMDz produce a stronger westerly
jet in DJF in the upper stratosphere and mesosphere than LW-LMDz, something consistent with MERRA2. In general, the differences between SW-1L-LMDz and LW-LMDz do not necessarily imply reduced biases (cf. Figures 5a and 5b), but SW-ML-LMDz performs qualitatively better. This is particularly evident in the case of the westerly jet in the SH mesosphere in JJA (Figure 5f), which tilts toward the tropics with altitude and the wind differences with LW-LMDz are similar to those between MERRA2 and LW-LMDz (Figure 5b).

Figure 5. Cross-sections of zonal mean zonal wind from the three experiments (black contours, interval: 10 m s⁻¹): (a) DJF LW-LMDz, (b) JJA LW-LMDz, (c) DJF SW-1L-LMDz, (d) JJA SW-1L-LMDz, (e) DJF SW-ML-LMDz, and (f) JJA SW-ML-LMDz. The color shading corresponds to the differences between LW-LMDz minus Modern-Era Retrospective Analysis for Research and Applications in (a) and (b), SW-1L-LMDz minus LW-LMDz in (c), and (d), and SW-ML-LMDz minus LW-LMDz in (e) and (f). Units are in m s⁻¹.
To test if these changes can be associated with changes in GW drag, Figure 6 shows the cross-sections of zonal mean non-orographic GW drag in SW-1L-LMDz and SW-ML-LMDz (black contours), and the corresponding differences with respect to the control run LW-LMDz (shading). The non-orographic GW drag is more effective at higher altitudes in the upper stratosphere and mesosphere, where we find positive drag in the summer hemisphere and negative drag in the winter hemisphere, contributing to decelerate the easterlies and westerlies, respectively. The differences of GW drag between SW-1L and LW are consistent, at least in summer, with the differences in zonal winds, since regions of slower (faster) mesospheric winds roughly correspond to regions of smaller (larger) GW drag in Figure 6 (in the SH, see between 60 km < z < 80 km and 80°S-30°S in Figures 5c and 6a, and in the NH between 60 km < z < 80 km and 30°N-80°N in Figures 5d and 6b).

On the other hand, the relations between GW drag and zonal wind are less direct in winter, particularly for SW-ML, precisely when the differences of zonal wind between SW-ML and LW are more pronounced (altitudes between 60 and 80 km and around 30° in each hemisphere in Figures 5e–5f). The difference in GW drag at those places (Figures 6c and 6d) shows a weaker drag in SW-ML below the jet shear zone, and a stronger drag above. The different sign at the two sides of the shear zone reveals that it is not the case that stronger or weaker drag is being exerted, as happened in summer for SW-1L, but that the available drag is distributed differently according to the background wind speed. Another important thing to notice is that the latitudes around and equatorward of 30° also correspond to the locations of the tropospheric jet center at z ≈ 12 km. We can therefore speculate that waves with positive intrinsic phase speed produced in these regions (see Figure 3) will easily propagate upward since the zonal mean wind decreases with altitude. These waves will then yield the difference in acceleration seen at mesospheric levels. Moreover, the hypothesis that the differences in SW-ML with respect to LW are

Figure 6. Cross-sections of zonal mean GW drag (black contours, interval: 5 m s⁻¹ d⁻¹), for (a) SW-1M-LMDz and DJF, (b) SW-1ML-LMDz and JJA, (c) DJF SW-1L-LMDz and DJF, and (d) SW-1L-LMDz and JJA. The color shading represents the difference with respect to LW-LMDz.
due to low phase speed waves is supported by the fact that the differences in GW drag in the winter mesosphere show pronounced dipoles tilted almost exactly along the equatorward flank of the westerly jet in the mesosphere. However, the verification of this hypothesis would need a more detailed analysis of the relation between the wave sources and drag in the model itself. Our explanation is therefore that the stratospheric jet tilt in the SH, which is reproduced in SW-ML-LMDz, is in part due to the forcing of GWs with small intrinsic phase speeds generated in the upper troposphere and lower stratosphere. This can be realized by a GW scheme with sources that are not confined to a single level in the troposphere.

Overall, the two new settings of the frontal GW parameterization (i.e., SW-IL and SW-ML) produce a reasonable middle atmosphere mean circulation, and even contribute to specific improvements in the shape of the austral winter jet that look promising to correct long-standing biases. To gain further confidence in the performance of the model under these new settings, and to test whether the changes in the time-average zonal mean winds are not made at the expense of the variability, we next evaluate the zonal mean temperature at 80°N and 10hPa, which is a conventional region to identify the occurrence of sudden stratospheric warmings (Baldwin et al., 2021). The panels in Figure 7 show the annual evolution of temperature over the years 1996–2010 centered on NH winters and the variability is quantified through the 5th and 95th percentiles (shading). Starting with the control run LW-LMDz (red line and shading in Figure 7a), the model has a clear warm bias at 10 hPa as compared to MERRA2 (black line and shading in Figures 7a and 7d), which is consistent with a weak wind bias (Figure 5a). This temperature bias in the mean cycle is reduced in late winter and spring using the alternative settings SW-IL.
and SW-ML (Figures 7b and 7c). Interestingly, reducing the mean bias also results in larger and more realistic variability, which is quite common when decreasing the GW drag in a model (Lott et al., 2005). Although the results here are essentially shown to illustrate that the parameterizations used work reasonably well, it stays that the three model simulations have a tendency to produce a substantially warmer polar stratosphere in late winter than the reanalysis.

5. Conclusions

Specifying the emission of GWs based on the model grid-scale dynamics, and using our observational knowledge of the GW field to constrain tunable parameters in the process, are two major challenges for improving parameterizations of GW drag (Alexander et al., 2010; Plougonven et al., 2020), particularly for GWs generated by fronts and jet imbalance (Plougonven & Zhang, 2014). The parameterization introduced in de la Cámara and Lott (2015) is based on a formalism that can be used to address these challenges. For this purpose we introduce the following modifications to this parameterization, with the goal of improving the representation of the GW field in the scheme: (a) we reduce the horizontal intrinsic phase speeds of the launched waves following suggestions from observations and high-resolution model runs (Plougonven et al., 2017; SW-1L configuration); and (b) we launch GWs from all model levels (SW-ML configuration) instead of launching from only one tropospheric level. This can result in more saturated waves everywhere in the middle atmosphere.

A technical result of our study is that launching GWs at different model levels can be achieved at a moderate numerical cost, in our case the cost of the stochastic parameterization of non-orographic GWs used in LMDZ. To limit the cost, if needed, we could even use the fact that one launching level often dominates the GW source term, and take that level as that of emission. From a scientific point of view, some results are worth highlighting. The first is that decreasing the intrinsic phase speeds and launching GWs from all levels improves the slopes of the vertical energy spectra of parametrized GWs and shifts it toward the observed value of −3. Second, the middle atmospheric circulation in climate model simulations responds reasonably well to the applied changes. There are even some indications of a weak reduction of model biases, such as an improved equatorward tilt of the austral polar night jet and stratospheric polar temperatures over the Arctic. These bias reductions may be model dependent, or could have been obtained in the same model with further refined tuning of the initial scheme. However, we obtain these results using configurations with about the same amount of wave stress launched (see Figure 2c) and only moderate alterations of the parameters of the scheme. This may indicate that the improvements obtained are not simply a signature of refined tuning.

Despite the modifications performed to the GW parameterization, placing inertial gravity waves in the middle atmosphere suggested by observations (e.g., Gelinas et al., 2012) has proven difficult (not shown). This may call for substantial changes in the formalism, perhaps in the source amplitude specification, the latter only including spontaneous adjustment. Classical geostrophic adjustment of the kind arising after short gravity waves breaking (secondary emission) could also be considered (Lott, 2003; Lund et al., 2020; Vadas et al., 2003). In any case, the choice of including sources at all model levels stays a valid option. Secondary emission, for instance, generates waves at altitudes where primary GWs break, and this usually takes place well in the middle atmosphere.

Current efforts of GW parameterization development are generally focused on adding complexity in the way parameterizations treat wave propagation and dissipation (e.g., Kim et al., 2020; Plougonven et al., 2020, and references therein). Our results demonstrate that modest changes in source-specifications guided by GW observations can have desirable effects in both the realism of the energy spectra and the simulated middle atmospheric circulation in a climate model. These changes are less computationally demanding than the development of schemes allowing lateral propagation (e.g., Bölöni et al., 2016). Since an argument for including lateral propagation is that columnar propagation hinders GWs to propagate to specific regions where they are needed, our approach provides a stopgap to represent the desirable waves.

Data Availability Statement

The last version of the LMDZ source code can be downloaded freely from the LMDZ web site http://lmdz.lmd.jussieu.fr/. The version used for the specific simulations runs for this paper is the “svn” release 3404 from 2018/10/16 which can be downloaded and installed on a Linux computer by running the “install lmdz.sh” script.
available at http://www.lmd.jussieu.fr/lmdz/pub/install%20lmdz.sh%2020181016.trunk. In agreement with the FAIR Data standards, the data used in this article has been archived, published, and made available in the dedicated repository https://data.ipsl.fr/catalog/srv/eng/catalog/search/#/metadata/167f4879-0296-49cf-bfb1-c526e49f884. The MERRA reanalysis data is accessible at https://doi.org/10.5067/3Z173KIE2TPD.

Acknowledgments
We thank two anonymous reviewers for their detailed comments that substantially helped improve the manuscript. This study has been supported by the Laboratoire de Recherche Conventiоnnelle “Yves Rocard” and the VESRI Schmidt Future project “DataWave”. AC acknowledges funding from the Spanish Ministry of Science project DYNWARM (PID2019-109107GB-I00).

References


