Can we improve the realism of gravity wave parameterizations by imposing sources at all altitudes in the atmosphere?

B. Ribstein¹, C. Millet¹,², F. Lott³, and A. de la Cámara⁴

¹CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France
²CEA, DAM, DIF, F-91297 Arpajon, France
³LMD, PSL Research Institute, Ecole Normale Supérieure, Paris, France
⁴Dept. Earth Physics and Astrophysics, Universidad Complutense de Madrid (UCM), Spain

Key Points:

- Non-orographic gravity wave parameterization is improved by including inertial waves and waves sources at all model levels.
- Parameterized energy spectrum becomes much closer to observations.
- Global model with the new parameterization performs well, some model biases are modestly alleviated.

Corresponding author: Alvaro de la Cámara, acamarai@ucm.es
Abstract

A multiwave non-orographic gravity waves (GWs) 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 middle atmosphere 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 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), using radars (Love & Murphy, 2016; Shibuya & Sato, 2019), lidars (Baumgarten –2–
et al., 2016; Khaykin et al., 2015), and satellites (Alexander, 2015). Their amplitude grows
as they propagate vertically and they impact the middle atmosphere circulation when
they break, even if they have relatively small amplitudes at the source level. The spa-
tial 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 pa-
parameterized for GCMs to produce a reasonable circulation. Such parameterizations were
first introduced in the 1980’s 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, most models resolve the middle atmosphere requiring the param-
eterization 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 that over a large number of realizations the GW fluctuations of vertical wind
and temperature in the middle atmosphere follow a “universal” spectra, which shape is
derived from radiosondes and satellite data (e.g., Cot, 2001; Zhao et al., 2017). These
spectra are numerically robust (Lindborg, 2006; Brethouwer et al., 2007), and various
theories have been developed to explain them. Some involve wave breaking (e.g., Dewan
& Good, 1986), whereas other include nonlinear effects like triade interactions, Doppler
spreading, and inverse cascades (e.g., Broutman & Young, 1986; Lilly, 1983; C. Hines,
1996; Métais et al., 1996). Beyond the theoretical debate, a practical result is that the
existence of a saturated spectra allows semi-theoretical integrations that ease the param-
eterization of non-orographic GWs (C. O. Hines, 1997; Warner & McIntyre, 1996; Manzini
et al., 1997). To a certain extent, this approach is challenged by the recent balloon ob-
servations showing that the GW field is very intermittent, and is often dominated by rather
well-defined GWs packets (Hertzog et al., 2008; Wright et al., 2013; Alexander, 2015).
This intermittent nature makes that in each model gridbox and at each time step, the
number of GWs packets is not large enough to fulfill the law of large numbers underly-
ing the construction of spectra out of individual realisations. More generally, this con-
cern is related to that of statistical equilibrium and is central in the recent development
of stochastic parameterizations (Berner et al., 2017). In the case of the GWs this im-
plies that the “universal” spectral shape should be checked a posteriori and over a large
number of days, rather than being realised every time. This leads to an alternative ap-
proach 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” versus ”multiwave”) and the two types of observations (stationary universal ”vertical spectra” versus 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 Cámara et al., 2014). In route to realize this objective, the study of (Souprayen et al., 2001) is encouraging since it shows that the wave filtering by the large scale flow and the breaking of individual GWs packets can yield realistic spectra in the upper stratosphere and mesosphere.

In some modeling centers the amplitude of the parameterized GWs is related to their non-orographic sources, i.e. 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, and Polichtchouk and Scott (2020) discuss the generation from 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 hours 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 saturate more easily 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 present paper analyses if the “multiwave” non-orographic GW parameterization due to fronts and jets of de la Cámara and Lott (2015) can be adapted to include sources from all levels, small intrinsic phase speed waves (including near inertial waves),
and if these modifications can help to 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 few characteristics that makes 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 (Polichtchouk & Scott, 2020), and which is not limited to the troposphere. The second is that it points to 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 Institute Pierre Simon Laplace (IPSL) Earth system model (Hourdin et al., 2013), so it is routinely tested. The fourth is about observational constraints, in the sense that it qualitatively produces the observed intermittency of the nonorographic GW field (Hertzog et al., 2012; Wright et al., 2013; de la Cámara et al., 2014; Alexander, 2015). Indeed, de la Cámara, 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 40ms$^{-1}$. 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 in the troposphere only. Finally, it is worth noting that the shortcomings we try to deal with are today considered as 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., Muraschko et al., 2015; Ribstein & Achatz, 2016;
Amemiya & Sato, 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 goal of this paper is to present modifications so that the frontal GWs parameterization can include slow intrinsic phase speed waves, rotation and GWs sources at all levels in the atmosphere. Section 2 describes the modifications we propose to the frontal GWs parameterization used in LMDz. In section 3, offline tests are performed to test if the scheme predicts realistic vertical energy spectra. Section 4 presents online results obtained with the LMDz GCM, 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 \((u', T')\) due to GWs are represented by a stochastic Fourier series of \(J\) monochromatic waves,

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

whose horizontal wavevector \(k_j\) and absolute frequency \(\omega_j\) are chosen randomly, and the complex amplitudes \(\left( \hat{u}_j(z), \hat{T}_j(z) \right)\) 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

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

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

To evaluate the wave amplitude, we adapt 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, and consider that a given model level \(z_l\) of thickness \(dz_l\) emits
a vertical Eliassen-Palm (EP) flux with value,

\[
F(z_l) = G_0 \frac{\Delta z_d z_l}{f_0} \rho(z_l) N(z_l) \left[ f \tanh \left( \frac{\zeta(z_l)}{f} \right) \right]^2 e^{-z/H(z_l)},
\]

where \(U_z = |U_z|\) is the modulus of the vertical shear, \(N\) is the Brunt-Väisälä frequency, \(\rho = e^{-z/H}\), and \(H = 7\) km is the characteristic vertical scale of density decay. Still in equation (3), \(G_0\) is a tuning parameter, \(f_0 = 10^{-4} s^{-1}\) is a characteristic value of \(f\), \(\Delta z\) a characteristic depth of the subgrid scale PV anomalies. To derive equation (3) from theory we made the approximation that the subgrid scale vorticity equals the gridscale one \(\zeta\). The only novelty at this stage is the hyperbolic tangent term 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 moderates the emission intensity in the tropics.

de la Cámara and Lott (2015) assumed that the momentum fluxes essentially come from the troposphere, but integrated the contributions in equation (3) over all model levels \(L\), to calculate a total emitted flux,

\[
F = \sum_{l=1}^{L} F(z_l). \tag{4}
\]

It is then imposed that this momentum 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 \(F_j\) of the total EP-flux \(F\) is specified through the random horizontal wavevector \(k_j\), following the rule

\[
F_j(z_{La}) = -\frac{k_j}{|k_j|} F. \tag{5}
\]

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

\[
\hat{\omega}_j(z_{La}) = \omega_j - k_j \cdot U(z_{La}) > 0. \tag{6}
\]

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 (i) reduced by a small diffusivity \(\mu/\rho_0(z)\), (ii) limited by that of a saturated wave (e.g., Lindzen & Schoeberl, 1982), and (iii) set to zero immediately above an inertial level (Lott et al., 2015):

\[
F_j(z_{l+1}) = -\frac{k_j}{|k_j|} \Theta \left( \hat{\omega}_j(z_{l+1})^2 - f^2 \right) \times
\]

\[
\cdots
\]

\[
\cdots
\]

\[
\cdots
\]

\[
\cdots
\]
\[
\min \left\{ \| \mathbf{F}_j(z_l) \| \exp \left( 2 \frac{\mu m_j^2(z_l)}{\rho_0 \hat{\omega}_j(z_l)} dz_l \right), \rho_0 S_c^2 \| N(z_{l+1}) \|^2 \frac{k_{\min}^2}{|m_j(z_{l+1})|^2} \right\}, \quad (7)
\]

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

\[
m_j(z) = - \frac{N(z) \| \mathbf{k}_j \|}{\sqrt{\hat{\omega}_j(z)^2 - f^2}}, \quad \text{where} \quad \hat{\omega}_j(z) = \omega_j - \mathbf{k}_j \cdot \mathbf{U}(z), \quad (8)
\]

where the minus sign in the definition of \( m_j \) ensures upward propagation above the launching level. A novelty 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 much more reduced than for the waves in the direction opposed to it. In the scheme we also do exclude the emission of highly saturated waves, which also avoid 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 emissions 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 \( \mathbf{F}_j(z) \) 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 drag, a shortcoming that is justified by the rapid decrease of air density with altitude: the corrections to the fields and to the tendencies would presumably be of small amplitude around and below the launch level.

Once \( \mathbf{F}_j(z) \) 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):

\[
\hat{u}_j(z) = -\frac{k_j \hat{\omega}_j - if e_x \times k_j}{\hat{\omega}_j^2 - f^2} \phi_j(z) e^{i \int_0^z m_j(z') dz' + i \xi_j}, \quad (9)
\]

\[
\hat{T}_j(z) = im_j \frac{H}{R} \phi_j(z) e^{i \int_0^z m_j(z') dz' + i \xi_j}, \quad (10)
\]
where the amplitude of the geopotential is

$$\phi_j(z) = \sqrt{\|F_j\|N/(\rho m_j\|k_j\|)}$$  \hspace{1cm} (11)

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

$$\frac{\partial u}{\partial t}|_{GW} = \sum_{j=1}^{J} C_j^2 \frac{\partial F_j}{\partial z}.$$  \hspace{1cm} (12)

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

$$\frac{\partial T}{\partial t}|_{GW} = - u \cdot \frac{\partial u}{\partial t}|_{GW} / C_p.$$  \hspace{1cm} (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 $\hat{\omega}_j$ (see Fig. 1). Then the launching flux in (5) is replaced by

$$F_j(z_j) = - \frac{k_j}{\|k_j\|} F(z_j).$$  \hspace{1cm} (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) to (11), but with $z_{La}$ replaced by $z_j$ for each waves. But now that we have $J$ different launching altitudes, with the possibility that $J \neq L$, we need to take for intermittency parameter:

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

2.2 Model and reanalysis

In offline mode we use daily fields of temperature and horizontal winds from the Modern-ERa Retrospective Analysis for Research and Applications version 2 (MERRA2) (Gelaro et al., 2017) at 1° × 1° resolution and from the ground to 0.01hPa.

In online mode we use the stratospheric version of the Laboratoire de Meteorologie Dynamique Zoom (LMDz) model with 142×144 uniform latitude longitude grid and $L = 80$ vertical levels, the model top being at 0.01hPa. 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 GWs parameterizations (Lott & Miller, 1997; Lott & Guez, 2013), the different setting of the parameterization of the 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 $k_j = |\mathbf{k}_j|$ with uniform probability in the interval $k_{\text{min}} \leq k_j \leq k_{\text{max}}$, with $2\pi/k_{\text{max}} = 6.3 \text{km}$ and $2\pi/k_{\text{min}} = 315 \text{km}$ 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 $\theta_j (\mathbf{k}_j = k_j (\cos \theta_j \mathbf{i} + \sin \theta_j \mathbf{j}))$ 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 $z_l$ from a half-normal distribution with a standard deviation of $c_\phi$, and in the direction given by $\mathbf{k}_j$. The parameter $c_\phi$ is key, and tuning it in different experimental setups requires adjustments in the launched momentum flux amplitude and saturation parameters $G_0$ and $S_c$ in (3) and (7) respectively.

In this study we present results from three different setups of the nonorographic GWs parameterizations. The strategy adopted consists of (1) decreasing the phase speed drastically and (2) introduce multiple level sources adapting the other parameters to give performances that are quite comparable to 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 $z_{\text{La}} = 500 \text{ m}$, and take $c_\phi = 50 \text{ m} \cdot \text{s}^{-1}$, $G = 4$, and $S_c = 0.6$. These values stay reasonably close to those used in previous studies (e.g., de la Cámara, Lott, & Abalos, 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 $2\pi/m \approx 2\pi c_\phi/N \approx 20 \text{km}$,
we see that we are essentially taking into account long waves (since $c_\phi$ is large), and we
will refer to this setup as LW.

In the second experimental setup we keep a single launching level but decreases the
intrinsic phase speed down to $c_\phi = 10\text{ms}^{-1}$. This requires a slight increase in $G$ up to
$G = 5$, but since the characteristic vertical scale is now much shorter, i.e. $2\pi c_\phi/N \approx 4\text{km}$, the saturation parameters need to be increased substantially up to $S_c = 2.5$ 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). Finding this level out of $J$ cases
is equivalent to average over $J$ levels, so there is no need to change substantially the tuning
parameters $G$ and $S_c$ compared to the previous case, and we take for the new values $G = 4.2$ and $S_C = 3$. This configuration will be referred to as SW-ML.

3 Parameterized gravity wave spectra

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-b show
the spatial distribution on a given day (22 January 2012) of the daily average of vertical EP flux amplitude in SW-ML at the 500 hPa level. 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 towards the poles. In general, we see that SW-ML represents slightly larger fluxes
at all latitudes than the other two setups, and that SW-1L consistently works with 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 qual-
itatively follows the observed log-normal distribution in the lower stratosphere (not shown but see de la Cámara and Lott (2015)).

Next we analyze the ability of LW, SW-1L and SW-ML to reproduce the empirical GW energy spectra. The vertical profiles of the wind and temperature are obtained from the parameterized profile of vertical EP flux as in equations (9), (10), and (11). More specifically, and to build spectra out of individual realisations we first construct 1000 monochromatic waves, each corresponding to randomly chosen values of phase speed and horizontal wavenumber. 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 grid, together with the large scale fields of wind and Temperature needed in Eqs (9) and (10). To construct 1 realization out of these 1000 monochromatic waves, we pick \( J = 48 \) of them randomly, choose the phase \( \xi_i \) of each randomly and then sum over the \( J \) waves. Let \( u'(z) \) be one such realization of the horizontal wind disturbance, we use here the same notation as in (1), we then perform a Fourier analysis of this realisation which gives the periodogram \( \hat{u} \hat{u}^* \), \( \hat{u} \) being the Fourier coefficients and the stars indicating conjugation. To avoid numerical artifacts in the boundaries, a tapered cosine window is used together before the fast Fourier transform. The spectra are then estimated at each horizontal places by doing an ensemble average of individual periodograms \( \langle \hat{u} \hat{u}^* \rangle \) constructing an ensemble of 20 independent realisations. Although we found our results about spectral shape to be little sensitive to the procedure, we could have averaged at a given place over different days, or average the same day at different places, we adopt this one because it permits to compare without ambiguity regions with presumably large GW emission to regions with small GW emission (for instance the locations A and B in Figures 2a-b and the same day).

In the following we restrict the discussion to the horizontal wind spectra, but we verified that the temperature spectra exhibit a similar shape (not shown). The spectra for altitudes lower than 25 km and 65 km are displayed in Figure 3 at the two specific locations A and B and still for the 22 january 2012. For both locations the energy spectra obtained with the configurations LW and SW-1L are proportional to \( m^{-4.5} \), 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 (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 4a-b 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 4a-b 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 semianual 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 4b), 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 4a), 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 4 show the corresponding zonal wind profiles (black contours) for SW-1L-LMDz (Figures 4c-d) and SW-ML-LMDz (Figures 4e-f), 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 4a-b), but SW-ML-LMDz performs qualitatively better. This is particularly evident in the case of the polar night jet in the SH in JJA (Figure 4f), which tilts towards the tropics with altitude and the wind differences with LW-LMDz are similar to those between MERRA2 and LW-LMDz (Figure 4b).

To test if these changes can be associated with changes in GW drag, Fig. 5 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. For SW-1L and SW-ML the differences of GW drag and zonal mean zonal wind with respect to LW are generally consistent with this view since regions of slower (larger) mesospheric winds roughly correspond to regions of smaller (larger) GW drag in Figure 5 (see the summer mesosphere between 60km < z < 80km and 80°S-60°S in Figs. 4c and 5c). As we have modified the phase speeds substantially, this correspondence between GW drag and zonal wind differences could be simply related to changes in dynamical filtering at high altitude. In the regions mentioned before, this does not seem to be the case: weaker winds in summer should allow the propagation of more waves with positive intrinsic phase speeds that would strengthen the net positive drag, which is not what is found in summer in the upper mesosphere. Therefore, it would seem that the new sets of parameters are simply producing less GW drag and acceleration.

Dynamical filtering is more evident in the differences between the SW-ML run and the LW run and at the places where the differences in zonal wind are the more pronounced -i.e. at altitudes between 60km and 80km and around 30° in each hemisphere (Figures. 4(e)-(f)). The difference in GW drag at those places (Figures. 5(e)-(f)) shows a weaker drag in SW-ML below the jet shear zone, and a stronger drag above. The difference in sign between the two sides of the shear zone is essential, it reveals that we do not put globally more or less drag, as was the case in summer for SW-1L, but that we distribute it differently according to the wind speed. Another important thing to notice is that the
latitudes around 30° also correspond to the latitudes of the winter tropospheric jet center at $z \approx 15$km. We can therefore speculate that since the wind shear is negative on the upper flank of the jet, positive phase speed waves are produced, they create acceleration right above at mesospheric levels, but are not much efficient further above because they have small intrinsic phase speeds and are easily filtered out by the wind they produces. Our explanation is therefore that the stratospheric jet tilt in the SH, which is reproduced in SW-ML-LMDz, is in part supported by GWs with small intrinsic phase speeds coming from the upper troposphere and lower stratosphere. This can be realized by a GW scheme with sources that are not confined to the troposphere.

Overall, the mean middle atmospheric circulation with the two new settings of the frontal GW parameterization (i.e. SW-1L and SW-ML) is reasonable, and presents specific improvements in the shape of the austral winter jet that look promising to correct long-standing biases. To gain further confidence into the performances of the model in these new settings, we next evaluate the interannual variability of the different model runs. Figure 6 shows the annual cycle of the zonal mean temperature at 80°N and the variability over the period 1996-2010 quantified through the 5th and 95th percentiles (shading). Starting with the control run LW-LMDz (red line and shading in Figure 6a), the model has a clear warm bias at 10 hPa as compared to MERRA2 (black line and shading in Figure 6a, and Fig. 6d), which is consistent with a weak wind bias (Figure 4a). This temperature bias is reduced using the alternative settings SW-1L and SW-ML (Figures 6b-c), both in the mean cycle and the variability given by the percentiles. However, the three model simulations clearly overestimate temperature fluctuations in early and late winter.

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) was based on a formalism that can be used to address these challenges. For this purpose we introduce the following modifications to this parameterization: 1) we reduce the horizontal intrinsic phase speeds of the launched waves following sugges-
tions from observations and high-resolution model runs (Plougonven et al., 2017) (SW-1L configuration); and 2) 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. From a more scientific point of view, some results are worth highlighting. The first is that decreasing the intrinsic phase speeds and launching GWs from all levels shifts the slope of the vertical energy spectra of parametrized GWs 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 (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 (Vadas et al., 2003; Lott, 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 at-
mospheric circulation in a climate model. These changes are much less computationally demanding than the development of schemes allowing lateral propagation (e.g., Büöni 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.

Acknowledgments

This work was supported by the Laboratoire de Recherche Conventionn Yves Rocard, a collaborative unit between CEA and Ecole Normale Supérieure. BR acknowledges funding from the European project ARISE 2 (Horizon 2020, GAN653980), FL from VESRI Schmidt Future project "DataWave", and AC from the Santander-UCM grant StratWars (PR87/19-22679).

References


Baumgarten, G., Fiedler, J., Hildebrand, J., & Luebken, F.-J. (2016). Inertia gravity wave in the stratosphere and mesosphere observed by doppler wind


Lott, F. (2003). Large-scale flow response to short gravity waves breaking in a rotat-
ing shear flow. *Journal of the atmospheric sciences, 60*(14), 1691–1704.


doi: 10.1256/smsqj.53703


Figure 1. Schematic representation of the scheme used in (de la Cámara & Lott, 2015) (left) and of its modification to include waves emitted from all model levels (right).

Figure 2. (a,b) Horizontal distribution of the total EP flux (mPa) at pressure level 500 hPa in the SW-ML setup. The geographical locations A and B are the locations where the energy spectra are constructed as shown in Figure 3. (c) Zonal mean EP flux at 500 hPa for the three GWs configurations.
Figure 3. Zonal wind spectra from realizations of the GWs fields obtained using parameterizations LW, SW-1L and SW-ML, respectively, for locations A (a,c) and B (b,d) (see figure 2) and altitude ranges 0-65 km (a,b) and 0-25 km (c,d).
Figure 4. 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
- f) JJA SW-ML-LMDz

The color shading corresponds to the differences between LW-LMDz minus MERRA 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\(^{-1}\).
Figure 5. Cross-sections of zonal mean GW drag (black contours, interval: 5\text{m\cdot s}^{-2}), 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.
Figure 6. Annual cycle of the simulated mean temperature over a period of 15 years and associated 5th and 95th percentiles at 10 hPa and latitude 80° N. (a) LW-LMDz, (b) SW-LMDz, and (c) SW-ML-LMDz. The Observed 1996 - 2010 mean temperatures from MERRA2 (d) are superimposed to simulated results.