RESEARCH ARTICLE

Comparison between non-orographic gravity-wave parameterizations used in QBOi models and Strateole 2 constant-level balloons

F. Lott[1] | **R. Rani[1]** | **C. McLandress[2]** | **A. Podglajen[1]** | **A. Bushell[3]** |
M. Bramberger[4] | **H.-K. Lee[5]** | **J. Alexander[4]** | **J. Anstey[2]** | **H.-Y. Chun[5]** |
A. Hertzog[6] | **N. Butchart[3]** | **Y.-H. Kim[7]** | **Y. Kawatani[8,9]** | **B. Legras[1]** | **E. Manzini[10]** |
H. Naoe[11] | **S. Osprey[12]** | **R. Plougonven[13]** | **H. Pohlmann[10]** | **J. H. Richter[14]** |
J. Scinocca[2] | **J. García-Serrano[15]** | **F. Serva[16]** | **T. Stockdale[17]** | **S. Versick[18]** |
S. Watanabe[8] | **K. Yoshida[11]**

Correspondence
F. Lott, Laboratoire de Météorologie
Dynamique, Ecole Normale Supérieure,
24 rue Lhomond, 75231 Paris, France.
Email:
francois.lott@sorbonne-universite.fr

Abstract

Gravity-wave (GW) parameterizations from 12 general circulation models (GCMs) participating in the Quasi-Biennial Oscillation initiative (QBOi) are compared with Strateole 2 balloon observations made in the tropical lower stratosphere from November 2019–February 2020 (phase 1) and from October 2021–January 2022 (phase 2). The parameterizations employ the three standard techniques used in GCMs to represent subgrid-scale non-orographic GWs, namely the two globally spectral techniques developed by Warner and McIntyre (1999) and Hines (1997), as well as the "multiwaves" approaches following the work of Lindzen (1981). The input meteorological fields necessary to run the parameterizations offline are extracted from the ERA5 reanalysis and correspond to the meteorological conditions found underneath the balloons. In general, there is fair agreement between amplitudes derived from measurements for waves with periods less than 1 h and parameterizations. The correlation between the daily observations and the corresponding results of the parameterization can be around 0.4, which is 99% significant, since 1200 days of observations are used. Given that the parameterizations have only been tuned to produce a quasi-biennial oscillation (QBO) in the models, the 0.4 correlation coefficient of the GW momentum fluxes is surprisingly good. These correlations nevertheless vary between schemes and depend little on their formulation (globally spectral versus multiwaves for instance). We therefore attribute these correlations to dynamical filtering, which all schemes take into account, whereas only a few relate the gravity waves to their sources. Statistically significant correlations are mostly found for eastward-propagating waves, which may be due to the fact that during both Strateole 2 phases the QBO is easterly at the altitude of the balloon flights.

We also found that the probability density functions (pdfs) of the momentum fluxes are represented better in spectral schemes with constant sources than in schemes ("spectral" or "multiwaves") that relate GWs only to their convective sources.

1 | INTRODUCTION

It is well known that the large-scale circulation in the middle atmosphere is in large part driven by gravity waves (GWs) that propagate upward in the stratosphere and mesosphere (Andrews *et al.*, 1987). These waves carry horizontal momentum vertically and interact with the large-scale flow when they break. Since the horizontal scale of these waves can be quite short, much shorter than the 1° to 2° horizontal resolution of the atmospheric general circulation models (GCMs) used in most Earth System models, they need to be parameterized (Alexander & Dunkerton, 1999). In the Tropics, GWs generated by convection are believed largely to dominate (Alexander *et al.*, 2000; Fovell *et al.*, 1992; Lane & Moncrieff, 2008). These waves also contribute significantly to the forcing of the quasi-biennial oscillation (QBO), a near-28-month oscillation of the zonal mean zonal winds that occurs in the lower part of the equatorial stratosphere (Baldwin *et al.*, 2001). For these reasons, convectively generated GWs need to be parameterized in order to simulate a QBO in most GCMs.

Although gravity-wave parameterizations are now used in many models with success, including in the Tropics (Anstey *et al.*, 2016; Beres *et al.*, 2005; Bushell *et al.*, 2015; Christiansen *et al.*, 2016; Lott & Guez, 2013; Orr *et al.*, 2010; Scinocca, 2003; Serva *et al.*, 2018; Song & Chun, 2005), their validation using direct in situ observations remains a challenge. Large horizontal-scale GWs can be obtained from global satellite observations of temperature (Geller *et al.*, 2013) and the corresponding momentum flux computed using polarization relations (Alexander *et al.*, 2010; Ern *et al.*, 2014). However, in order to observe the shorter horizontal scales that also contribute to the QBO forcing and to have a direct measurement of the corresponding momentum flux, in situ observations are required. The most precise measurements are provided by constant-level long-duration balloons, like those made in the Antarctic region during Strateole Vorcore (Hertzog, 2007) and Concordiasi (Rabier *et al.*, 2010), or in the deep Tropics during PreConcordiasi (Jewtoukoff *et al.*, 2013) and Strateole 2 (Haase *et al.*, 2018). Among many important results, these balloon observations have

shown that the momentum flux entering the stratosphere is extremely intermittent (Hertzog *et al.*, 2012). This intermittency implies that the mean momentum flux is mostly transported by a few large-amplitude GWs that potentially break at lower altitudes, rather than by many uniformly distributed GWs. This intermittent character, when reproduced by a parameterization (Alexander *et al.*, 2021; de la Cámara *et al.*, 2014; Kang *et al.*, 2017), can help reduce systematic errors in the midlatitudes, such as the timing of the final warming in the Southern Hemisphere polar stratosphere (de la Cámara *et al.*, 2016), or in simulation of the QBO (Lott *et al.*, 2012a). Balloon observations have also been used to characterize dynamical filtering by large-scale winds (Plougonven *et al.*, 2017) and to validate the average statistical properties of the GW momentum flux simulated offline using reanalysis data (Alexander *et al.*, 2021; Kang *et al.*, 2017). For completeness, note that here "dynamical filtering" is the process by which waves with smaller amplitude intrinsic phase speed break for smaller amplitude momentum fluxes (MFs) than waves with larger amplitude intrinsic phase speed (see eq. (3) and the associated discussion in Lott *et al.*, 2023).

However, previous evaluations of parameterizations using balloon observations compared global statistical behaviour (Alexander *et al.*, 2021; Jewtoukoff *et al.*, 2015; Kang *et al.*, 2017) rather than the ability to reproduce instantaneous values of momentum fluxes. One good reason to consider global statistical properties of momentum flux, rather than daily values, is that parameterizations are based on simplified quasi-linear wave theory, assume spectral distributions that are loosely constrained, and ignore lateral propagation almost entirely (some attempt to include it can be found in Amemiya & Sato, 2016; Kim *et al.*, 2024; see also the underlying theory in Achatz *et al.*, 2023). Nevertheless, some factors could mitigate these weaknesses. One factor is that in all parameterizations the wave amplitude is systematically limited by a breaking criterion that encapsulates nonlinear effects. Another is that some parameterizations relate launched waves to sources explicitly, and there is a continuing effort to improve the realism of convective sources (Liu *et al.*, 2022). Finally, observations suggest

systematically that dynamical filtering by the large-scale wind is extremely important for upward-propagating GWs (Plougonven *et al.*, 2017), and this central property is represented in all GW parameterizations. For all these reasons, it may well be that GW parameterizations using the large-scale flow found at a given place and time produce momentum fluxes that can be compared directly with those measured by a balloon at the same place.

Based on the relative success of previous offline calculations using reanalysis data (Alexander *et al.*, 2021; Jewtoukoff *et al.*, 2015; Kang *et al.*, 2017), Lott *et al.* (2023) have shown that such a direct comparison gives results of interest. The first is that the state-of-the-art convective gravity-wave drag scheme of Lott and Guez (2013) predicts momentum fluxes in the lower equatorial stratosphere, the amplitudes of which can be compared directly with those measured during phase 1 of the Strateole 2 balloon campaign. This gives a direct in situ observational confirmation that the theories and modelling of QBOs developed over the last 50 years are largely correct about the importance of GWs for driving the QBO. Moreover, the comparison showed a good level of correlation between the day-to-day variability in momentum fluxes between measured and parameterized values, a correlation that is much higher for waves carrying momentum fluxes in the eastward direction than in the westward direction. Such a good correlation is consistent with the fact that the Lott and Guez (2013) scheme relates gravity waves to their convective sources (not all schemes do) and that the GWs experience strong dynamical filtering in the middle troposphere and lower stratosphere. However, Lott *et al.* (2023) also show that a scheme that relates gravity waves to only convection failed to predict the right statistical behaviour of the momentum fluxes. More precisely, the probability density functions (pdfs) of the predicted momentum-flux amplitudes have long tails for low values, which are more pronounced than in observations. This suggests that the parameterization misses processes like lateral propagation or the presence of a background of waves, the source and nature of which need to be understood.

The purpose of this article is to extend the direct comparison used in Lott *et al.* (2023) by including more recent Strateole 2 observations and different gravity-wave parameterizations. Here we use nearly all the parameterizations used by the modelling groups participating in the Quasi-Biennial Oscillation initiative (QBOi: Butchart *et al.*, 2018). We then follow Lott *et al.* (2023) and use the eight balloons of the first phase of the Strateole 2 campaign that flew in the lower tropical stratosphere between November 2019 and February 2020, and add the 15 balloons that flew more than one day during the second phase of the Strateole 2 campaign, between October 2021 and January 2022. In those flights and at each time in those flights, we have identified the horizontal grid point in the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) reanalysis (Hersbach *et al.*, 2020) that is nearest to the balloon location. At those times and places, we extract from ERA5 the vertical profiles of wind and temperature, as well as the surface value of precipitation, to calculate the parameterized GW momentum fluxes. As some parameterizations need them, we also extract from ERA5 analysis and 3-h forecasts the diabatic heating rates and the cloud-base and cloud-top altitudes.

The outline of the article is as follows. Section 2 describes the data and the parameterization schemes used. Section 3 compares the MF values calculated with the parameterizations with the observed values by evaluating daily correlations, averages over each balloon flight, averages over all the Strateole 2 flights, and distributions (pdfs). Section 4 summarizes the results. As we shall see, the performance of each parameterization can be contrasted when we use one type of result rather than another, but our purpose is not to promote one scheme over the others. Adapting other groups' parameterizations to a testbed that has been developed by Lott *et al.* (2023) for the Laboratoire de Météorologie Dynamique with "zoom" (LMDz) GCM see Lott *et al.*, 2023 can give an unfair advantage to the corresponding scheme, which is absolutely not the objective of the present work. We return to this point in Section 4.

2 | DATA AND METHOD

2.1 | Parameterizations of non-orographic gravity-wave schemes

The parameterization schemes used in GCMs to calculate non-orographic gravity waves belong to two distinct families, dating back to the 1980s, when it became evident that a simulation of the middle atmosphere by global atmospheric models could not be done without including subgrid-scale GWs.

The first family is based on the formulation of Lindzen (1981), where the gravity-wave field is represented by waves that are monochromatic in the horizontal and time. Lindzen's scheme was first extended to treat a large ensemble of waves by Alexander and Dunkerton (1999), making the assumption that the breaking of each wave could be made independent from the others. An advantage of such schemes is that they are based on linear theories, where sources like convection and/or fronts can be introduced using closed-form solutions (Beres *et al.*, 2005; de la Cámara & Lott, 2015; Lott *et al.*, 2012b; Lott & Guez, 2013;

Richter *et al.*, 2010a; Song & Chun, 2005). In the following we will refer to such schemes as "multiwave". These schemes are expensive because they request a large number of harmonics to represent a realistic wave field well, but this limit can easily be circumvented by using stochastic approaches (Eckermann, 2011; Lott *et al.*, 2012a).

As an alternative, but also to represent the effect of wave breaking better, globally spectral schemes have been developed and used with success. These schemes use the observational fact that GWs produce kinetic energy spectra that have a quite universal shape when expressed as a function of vertical wavenumber. In the early 1990s, Hines (1991) developed a theory where GW breaking is represented by imposing an upper limit on the range of vertical wavenumbers, the limit being calculated according to the large-scale wind and including Doppler spreading by the other gravity waves (see also Hines, 1997). The scheme has been implemented in various GCMs (see for instance Manzini *et al.*, 1997), and will be referred to as "HDS" for Hines Doppler Spread. As an alternative, the theory in Warner and McIntyre (1996) imposes gravity-wave saturation according to an empirical spectrum, but treats vertical changes in the spectrum following the propagation-invariant character of GWs. The theory has been simplified and/or optimized to permit implementation, for instance in the UK Met Office (UKMO) model (Scaife *et al.*, 2002; Warner & McIntyre, 1999) and Canadian Middle Atmosphere Model (CMAM, Scinocca, 2003), and will be referred to henceforth as "WMI" for Warner and McIntyre. To a certain extent, spectral schemes can also take into account the relation with sources. For instance, the HDS scheme has been related to fronts in Charron and Manzini (2002), and the UKMO version of the WMI scheme to precipitation in Bushell *et al.* (2015).

In the present article, we compare the GW schemes used in 12 of the models that participate in QBOi, these models having horizontal grid resolutions between 50 and 300 km, that is, much too short to resolve the shortest-period gravity waves explicitly. In these models, the GW parameterizations all belong to one of the three types of scheme described above (WMI, HDS, and Multiwave). Since all the multiwave schemes analysed relate GWs to their convective sources and since only one of the spectral schemes does so (i.e., the UMGA7gws WMI scheme in Bushell *et al.*, 2015), the results from the spectral scheme in Bushell *et al.* (2015) will often be presented together with the results from source-related multiwave schemes.

Among the 12 models, three (CMAM, IFS, and ECEarth) use the Scinocca (2003) version of WMI. The specific versions of these schemes used for QBOi are detailed further in Anstey *et al.* (2016); Orr *et al.* (2010), and Davini *et al.* (2017) respectively. These schemes essentially differ by four parameters: the launch-level pressure p_l, the launched momentum flux F_{LT}, the characteristic vertical wavenumber m_*, and a minimum intrinsic phase speed in the launched spectra, the values of each being given here in Table 1. Note that, for EC-Earth, the exact values of the parameters in Table 1 are from J. García-Serrano (private communication).

Five of the 12 models use the HDS parameterization discussed in Manzini *et al.* (1997): ECham5, MIROC, MPIM, MRI-ESM, and EMAC. Their versions for QBOi are described in Serva *et al.* (2018); Watanabe *et al.* (2011); Pohlmann *et al.* (2013); Naoe and Yoshida (2019), and Jöckel *et al.* (2010) (see also Roeckner *et al.*, 2006). They differ mainly by three different parameters: the launch-level pressure p_l, the root-mean-square of the horizontal wind variability due to GWs at launch level σ, and the effective horizontal wavenumber K^* (see Table 2). There are also more numerical parameters of secondary importance that differ between models: a minimum value for the cutoff vertical wavenumber m_{min}, and two parameters that control smoothing in the vertical of the GW root-mean-square variance, the coefficient C_{smo} and the number of times the smoothing is applied N_{smo}. It is important to note that in ECham5 the variability parameter σ is chosen randomly, with a normal distribution centred at 1 m/s with standard deviation 0.2 m/s. The usefulness of such a stochastic ingredient was initially proposed by Piani *et al.* (2004), who found that it can help stabilize the QBO variability in large-scale models and over decades.

TABLE 1 WMI parameters changing between CMAM, IFS, ECEarth, and UMGA7gws.

	p_l	F_{LT}	$2\pi/m_*$	C_{min}
CMAM	100 hPa	1.3 mPa	1 km	0.25 m/s
IFS	450 hPa	5 mPa	3 km	0.5 m/s
ECEarth	450 hPa	3.75 mPa	2 km	0.25 m/s
UMGA7gws	1000 hPa	$\sqrt{Precip}$	4.3 km	Not used

Note: That CMAM, IFS, and ECEarth parameterizations are very near each other and described in Scinocca (2003), whereas UMGA7gws follows Warner and McIntyre (1999).

TABLE 2   HDS parameters changing between ECHam5, MIROC, MPIM, MRI-ESM, and EMAC.

	p_1	σ_s	$2\pi/K^*$	$2\pi/m_{min}$	C_{smo}	N_{smo}
ECham5	600 hPa	$1. \pm 0.2\,m/s$	125 km	0	2	5
MIROC	650 hPa	0.95 m/s	250 km	94 km	2	2
MPIM	650 hPa	1.2 m/s	125 km	0	2	2
MRI-ESM	700 hPa	1.9 m/s	1250 km	190 km	4	2
EMAC	650 hPa	1 m/s	125 km	0	2	2

TABLE 3   Correlation between observed and measured fluxes, Strateole phases 1 and 2.

	Days −DoF	CM AM	IFS	ECE ARTH	Ech am5	MI ROC	MPI M	MRI ESM	EM AC	LMD z	UMG A7gws	HadG EM2	WAC CM
East													
Phase 1	670–216	ns	**0.53**	**0.52**	**0.43**	**0.48**	**0.49**	**0.44**	**0.48**	**0.49**	**0.34**	**0.31**	ns
Phase 2	621–322	**-0.19**	**0.41**	**0.38**	**0.29**	**0.33**	**0.34**	**0.30**	**0.33**	**0.40**	**0.34**	**0.20**	**0.26**
1 + 2	1291–538	-0.11	**0.49**	**0.47**	**0.35**	**0.41**	**0.41**	**0.36**	**0.40**	**0.46**	**0.34**	**0.26**	ns
West													
Phase 1	670–216	0.14	ns	ns	ns	ns	ns	ns	ns	**0.30**	ns	ns	ns
Phase 2	621–322	**0.21**	**0.18**	**0.16**	ns	ns	ns	ns	ns	**0.40**	ns	0.14	ns
1 + 2	1291–538	**0.17**	ns	ns	ns	ns	ns	ns	ns	**0.34**	ns	0.11	ns

Note: 1% significant values according to two-sided Pearson test are in bold, 5% are in italic, "ns" stands for non-significant. To evaluate the number of degrees of freedom, we proceed as in Lott *et al.* (2023) and evaluate for each flight the time lag for which the autocorrelations of the daily averaged fluxes fall below 0.1 and divide the number of days by that lag.

Finally, the last four schemes we consider all link GWs to sources (convection or precipitation). Three are multiwave schemes that have been developed independently from each other: LMDz, HadGEM2, and WACCM. Their versions used in QBOi are described in Lott and Guez (2013); Song and Chun (2005), and Richter *et al.* (2010b). One of these schemes uses the ultra-simple version of the WMI scheme presented in Bushell *et al.* (2015) rather than the Scinocca (2003) version. Note that, for both HadGEM2 and WACCM, we do not use the exact version used in the QBOi models, but rather the offline versions developed by Kang *et al.* (2017) and Alexander *et al.* (2021), which were adapted by these authors to interpret observations. Since the differences between the three multiwave schemes are too numerous, the reader is referred to the above-mentionned articles. However, important differences can be outlined in the source term, the launching levels, and the intrinsic phase speed of the launched waves (for the multiwave schemes see also Table 3). More specifically, in LMDz the choice is to relate the launched momentum flux to square precipitation P_r^2, consistent with linear theory before

breaking (Lott & Guez, 2013), whereas in (Bushell *et al.*, 2015) it is related to $\sqrt{P_r}$ (see Table 1). Furthermore, in LMDz the waves are launched in the mid-troposphere whereas in the UMGA7gws they are launched in the lower troposphere near the surface. In the HadGEM2 scheme (Choi & Chun, 2011; Song & Chun, 2005), the launched momentum flux is directly related to convective heating distributed in the vertical between the cloud bottom and cloud top, the launch altitude being at the cloud top. In this case the launch level can typically vary between 2 and 15 km and the depth of the heating between 1 and 15 km. We will take the same inputs used for the HadGEM2 scheme to run the WACCM scheme, using the version in Alexander *et al.* (2021). Note that in WACCM the heating depth is one quarter of the cloud depth, and typically ranges between 1 and 4 km. Final important differences are that in LMDz the intrinsic phase speeds are chosen randomly according to a Gaussian distribution with 0 mean value and 30 m/s standard deviation, whereas in both UMGA7gws and WACCM absolute phase speed is used, with values uniformly distributed in the range $-100\,m/s < C_{abs} < 100\,m/s$.

2.2 | Offline parameterization runs

To run the schemes in offline mode, we use ERA-5 hourly data of precipitation and three-hourly winds, surface pressure, temperature, and cloud liquid and ice water content on a $1° \times 1°$ horizontal grid to mimic a large-scale climate model of fairly high horizontal resolution. Winds, surface pressure, temperature, and water contents are then linearly interpolated in a 1-h interval so that they are synchronized with the precipitation. In the vertical we use data at 67 model levels, taking every second ERA5 level, again to mimic a typical model's vertical resolution but also to speed up calculations. To estimate the vertical profiles of convective heating rates, we follow Fueglistaler et al. (2009) and evaluate diabatic heating using ERA5 hourly data from the short-range forecasts, computing it as the residual between the parameterized temperature tendency and the radiative heating rates (longwave plus shortwave). When needed, we also evaluate the cloud-bottom and cloud-top altitudes using the cloud water content (liquid+ice) given in ERA5.

2.3 | Strateole 2 balloon observations

The in situ observations we use are from the eight balloons of the first phase of the Strateole 2 campaign that flew in the tropical lower stratosphere between November 2019 and February 2020 and from the 15 balloons that flew for more than one day during the second phase of the Strateole 2 campaign, between October 2021 and January 2022. The trajectories during phase 2 are shown in Figure 1, superimposed upon which is the averaged precipitation (the same figure, but for phase 1, is in Lott et al., 2023). For the MFs calculated from observations, Corcos et al. (2021) distinguish the eastward-travelling waves with positive MFs in the zonal direction from the westward-travelling waves with negative MFs. They also distinguish waves with short periods (1 h 15 min) from those with periods up to 1 day (1 day 15 min). In the following, we will follow Lott et al. (2023) and keep the shortest periods, because for these periods the MFs measured compare well with the GW momentum fluxes needed in a GCM to produce a QBO. This leaves open the issue that waves with longer periods are certainly in the "grey" zone of the models and not realistically represented either.

To characterize the phase of the QBO during the balloon flights, Figure 2 shows a time versus altitude cross-section of the equatorial zonal mean zonal winds and gravity-wave drag (GWD) computed in offline mode using the LMDz scheme for 2018–2023 and averaged over the Tropics. The gravity-wave drag is negative (positive) where the vertical wind shear is negative (positive), consistent
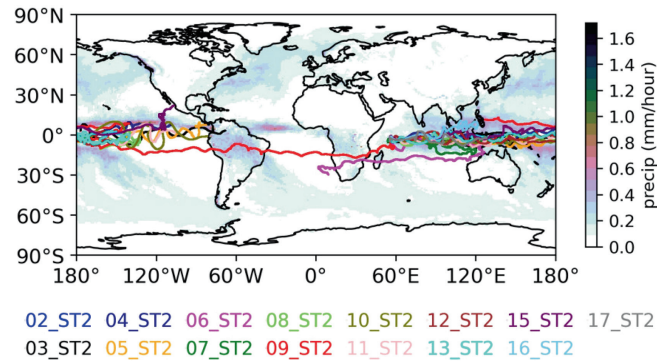

02_ST2 04_ST2 06_ST2 08_ST2 10_ST2 12_ST2 15_ST2 17_ST2
03_ST2 05_ST2 07_ST2 09_ST2 11_ST2 13_ST2 16_ST2

FIGURE 1 Strateole 2, Phase 2 balloon trajectories taking place between October 2021 and January 2022. Shading presents the precipitation field from ERA5 averaged over the period. [Colour figure can be viewed at wileyonlinelibrary.com]
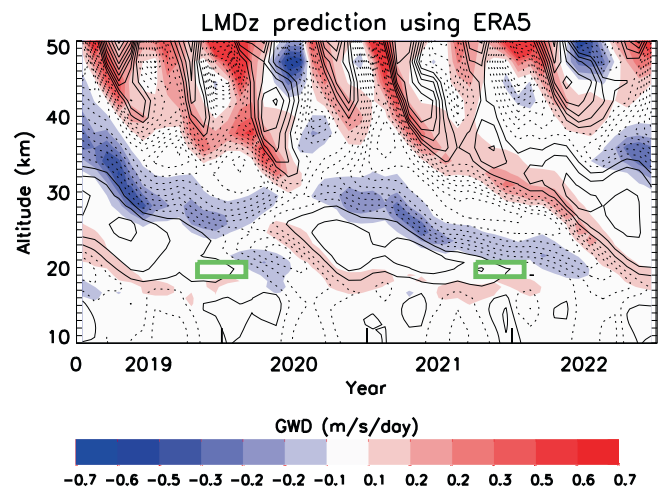

FIGURE 2 Time vertical sections of the zonal mean zonal wind (CI = 10 m/s, black contours with negative values dashed) and of the non-orographic gravity wave tendency (color shading). All zonal mean fields are also averaged over the equatorial band ($-6°$S to $+6°$N)). Input data are from ERA5 reanalysis and GW predictions from the LMDz scheme. The two green boxes indicate schematically the altitude and time ranges of the Strateole 2 phase 1 and 2 flights considered in this study. [Colour figure can be viewed at wileyonlinelibrary.com]

with the fact that it contributes to the QBO descent. We also note that the amplitudes vary between ± 0.5 m/s/day, a range characteristic of the parameterized GW drag tendency used in GCMs that produce a QBO-like oscillation (Butchart et al., 2018). The figure also indicates with green rectangles the regions and periods during which the balloons operated, typically during the end of the easterly QBO phase for both phases 1 and 2. As we shall see, this yields quite comparable results during the two phases, despite the fact that during phase 1 and above the altitude the second documented QBO disruption started (Anstey et al., 2021).

Our analysis compares the momentum fluxes derived from the balloon data for waves with intrinsic periods below 1 h and considers the ERA5 data at the points that are nearest to the balloon. The calculation is then made every hour and averaged over the day, partly because it is the time-scale needed for some of the schemes to sample a GW field realistically, and also because it takes about one day for a balloon flight to cover a model grid scale. Note that some of the sensitivities to these choices are discussed in Lott *et al.* (2023)'s conclusion.

3 | RESULTS

Figure 3 shows time series of daily values of momentum fluxes calculated by the parameterizations and measured during balloon flight 2 from strateole 2 phase 1. This is also the flight shown in fig. 3 of Lott *et al.* (2023), in which the time series of daily precipitation and zonal wind at flight altitude was shown. The top panel is for the WMI-based schemes, the middle panel for the HDS schemes, and the bottom panels for the schemes relating the GW fluxes to their sources (3 multiwave, 1 WMI). In all panels the black curves are for the daily observations. For clarity we present results for the eastward and westward MFs only. Overall, one sees that the parameterized MFs agree somewhat with the observed ones, at least in terms of amplitude. There are nevertheless significant differences in behaviour. For instance, the IFS schemes exhibit substantial peaks in eastward flux during the second half of the flight. This is a period during which the zonal wind at flight altitude becomes westward, potentially favouring eastward waves consistent with dynamical filtering. Note that in Lott *et al.* (2023) it was shown that the three peaks in measured fluxes around days 60, 75, and 83 also correspond to dates when there was precipitation near the balloon's horizontal location. These correspondences made us believe that a relation to convective sources is essential. However, we see here that dynamical filtering alone may well be the main cause. Although having smaller amplitudes, Figure 3 also shows that in EC-Earth, the momentum fluxes behave almost as in IFS. However, the results for CMAM are quite different. In this model we chose to place the launch altitude near the tropopause. As a consequence, the daily time series fluctuate less and exhibit long lasting "plateaus". Clearly, in this model, the distance between the launch level (100 hPa: see Table 1) and the balloon altitude is too small for dynamical filtering to be efficient. The second panel of Figure 3 for the HDS schemes is not fundamentally different from what was discussed above. The amplitude and fluctuations are comparable to those observed, some schemes predicting values that look either larger or smaller, but staying within the range of
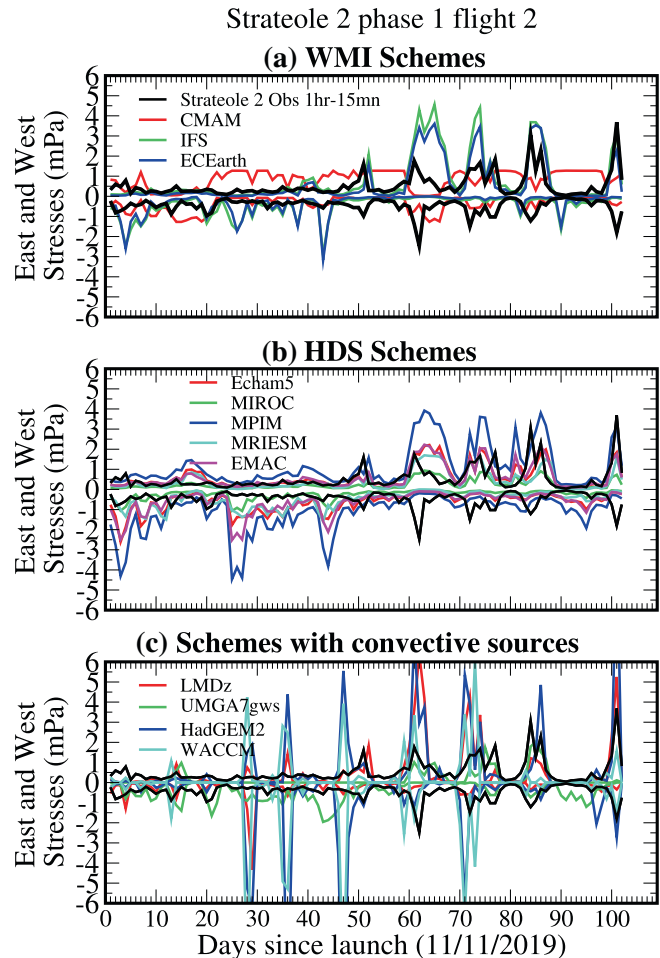
Strateole 2 phase 1 flight 2

FIGURE 3 Comparison between daily averaged values of the eastward and westward MFs measured by balloons during Strateole 2 phase 1 Flight 2 and estimated by the GW schemes at the balloon location and altitude. Coloured curves are for the GW schemes using ERA5, black curves are for the observed MFs due to 15 min–1 h GWs. (a) WMI schemes; (b) HDS schemes; (c) schemes relating launched MFs to convective sources or precipitation: all multiwaves except UMGA7gws. [Colour figure can be viewed at wileyonlinelibrary.com]

observations. The behaviour of the source-related schemes in the third panel of Figure 3 (multiwave for LMDz and HadGEM2, WMI for UMGA7gws) is more contrasted. As expected, there are long periods during which the schemes produce small and null momentum fluxes, which are interrupted by short-lasting strong peaks. These peaks sometimes exceed ±5mPa, which are values never reached by any of the spectral schemes in Figure 3a,b. In contrast to LMDz and HadGEM2, the UMGA7gws scheme exhibits smaller amplitude MFs and broader peaks. We attribute this to the fact that the UMGA7gws scheme relates the launch flux to $\sqrt{P_r}$ rather than P_r^2 as is done in LMDz, or to the square of heating as is done in both HadGEM2 and WACCM.
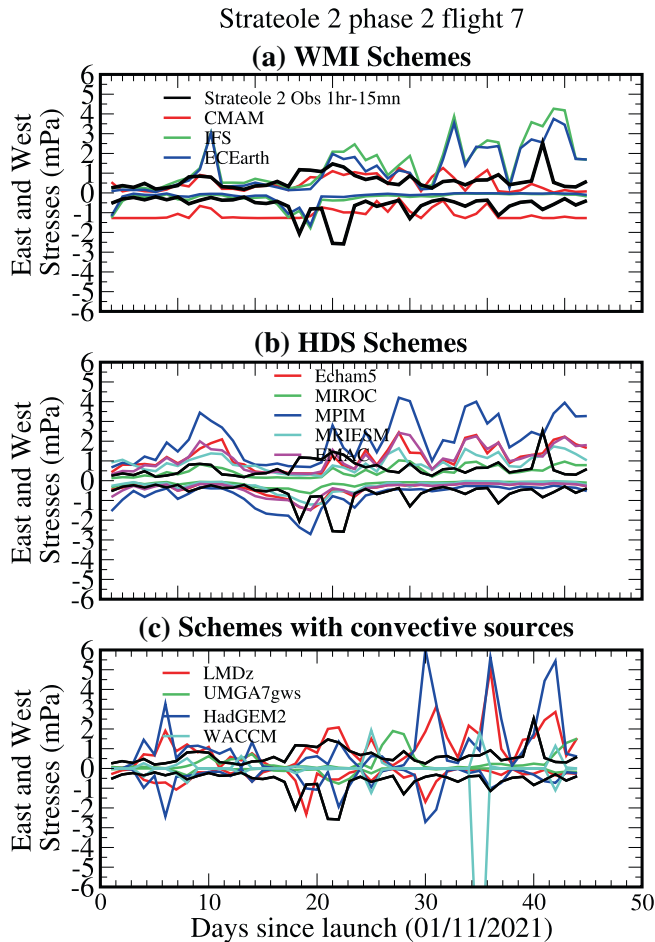
Strateole 2 phase 2 flight 7

FIGURE 4 Same as Figure 3 but for Strateole 2 Phase 2 Flight 7. [Colour figure can be viewed at wileyonlinelibrary.com]

The MF time series for a flight during the second phase of Strateole 2 is shown in Figure 4. Beyond the fact that the flight is shorter than in Figure 3, a difference in duration that characterizes most of the flights during phase 2 compared with phase 1, the overall behaviour stays about the same: the spectral schemes exhibit fluctuations with broader peaks, except maybe CMAM, as a result of the higher launch altitude, which results in dynamical filtering not yet being efficient at the balloon flight altitude. The last panel in Figure 4 also shows that UMGA7gws exhibits long periods with almost no fluxes. For this scheme, the launch height is low in the troposphere (see Table 1), which results in much more critical level filtering during the propagation through the troposphere. Finally, in the version of WACCM used here, there is one extreme outlier at day 33, with values below −10 mPa. We only found few of them over the entire campaign, and because WACCM has been intentionally tuned to produce such extreme values in MFs from time to time.

The fact that the different schemes estimate momentum fluxes of about the right amplitude is summarized in Figure 5, where the average of the fluxes over the 18 flights that last more than a month (eight during phase 1, 10 during phase 2) is shown. In this figure we see that the predicted values align quite well with the observed ones, though some schemes have a tendency to underestimate the fluxes slightly (MIROC, LMDz), and others to overestimate them (CMAM, HadGEM2). The WACCM scheme has a quite distinct behaviour: most balloons measure rather lower fluxes than parameterized on average, and few much larger ones. On average over all flights, these large values average out with smaller ones. However, we have to keep in mind that this behaviour is intentional: the version of the WACCM scheme we use has been tuned to produce a very intermittent behaviour and sometimes very strong fluxes (Alexander *et al.*, 2021). The numbers in each panel are the correlation coefficients between the 18 observed and parameterized values of MFs averaged over each flight. They show that the correlations are quite strong in some models, at least in the eastward direction. Interestingly, some models also have significant medium to high correlations in the westward direction (CMAM, LMDz, HadGEM2). This means that parameterizations can capture the low-frequency variability of the MFs (the changes with period larger than a month) quite well. Thus, it is tempting to say that it is good enough for simulation of the QBO.

Figure 6 compares the observed and parameterized eastward and westward fluxes averaged over all the balloon flights, confirming again that the parameterizations fall around the observed values. Although there are differences between the models, there is no systematic tendency for them to overestimate or underestimate the observed MF flux amplitude. This is elucidated by the green curve, which represents the average over all models and all balloon flights. As can be seen, the average amplitude of the eastward flux is very near that of the observed one (10% overestimate: 0.45 mPa for the parameterizations compared with 0.40 mPa for the observed), whereas the westward flux is overestimated by the models by less than 20% (−0.65 mPa for the parameterizations compared with −0.55 mPa observed). The near 50% error seen in the cumulated flux results from the fact that this flux is the sum of a large positive flux and a large-amplitude negative flux, the two almost equilibrating each other.

The daily time series in Figures 3 and 4 also suggest that observations and offline estimations sometimes evolve similarly day after day. A possible reason for this could be that both observed and parameterized MFs are sensitive to dynamical filtering, noting that some schemes also take into account convective sources. In the two examples shown in Figures 3 and 4, the correspondence between the observed and parameterized fluxes is quite
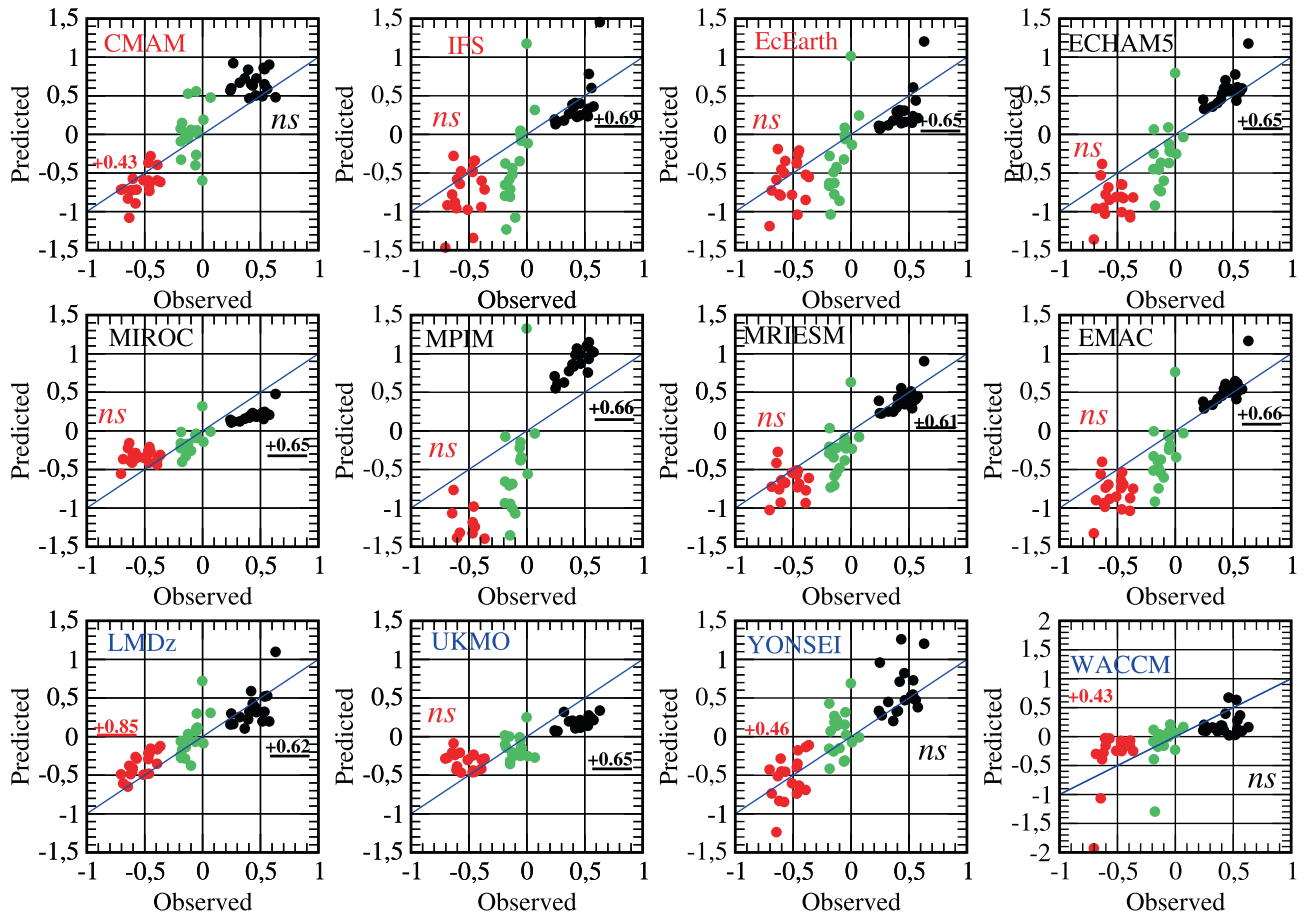
FIGURE 5 Scatter plot of the momentum fluxes measured by the balloon versus those parameterized using different models. We only consider here the 18 balloon flights that last more than a month (east: black; west: red; cumulated (east+west): green). Also shown are the correlations between observations and predictions; 99% significant levels are bold underlined, 95% are bold. Non-significant values are indicated by "*ns*". The number of DoFs for the Pearson test is 23, which is simply the number of balloon flights and is therefore very conservative, many balloons lasting more than few weeks, whereas the decorrelation time-scale of the daily series is well below a week. The names of the WMI, HDS, and convection-related GW schemes are in red, black, and blue respectively. Note the the change of vertical axis in the lower left panel. [Colour figure can be viewed at wileyonlinelibrary.com]

apparent, particularly in Figure 3, in regards to the peaks in the eastward direction discussed earlier. Correspondences are less apparent in Figure 4, where the observed MFs present less variation than the parameterized MFs. In Lott *et al.* (2023), where these daily variations were analysed flight by flight, in some of the flights the time series correlated well whereas in others they did not. This resulted in correlation coefficients C that are significant but "medium" in the eastward direction, $C \approx 0.5$, and "low" to "medium" in the westward direction, $C \approx 0.3$. Here and in the following, we refer to "medium" correlations when $0.3 < C < 0.5$ and "small" when $0.1 < C < 0.3$. As the latter values occured for the LMDz parameterization during Strateole 2 phase 1, the coefficients are given again in the ninth column of Table 4. Also given are the coefficients

for phase 2 and for phases 1 and 2 combined. Consistent with the results found for phase 1, during phase 2 we found medium correlation for the eastward MF ($C = 0.4$) and for the westward MF ($C = 0.40$), the values evaluated over the two phases being medium ($C = 0.46$ and $C = 0.34$, respectively). Here and for completeness, we follow the procedure used in Lott *et al.* (2023) to test the significance. We measure the number of degrees of freedom (DoF) for each dataset and calculate the decorrelation time-scale, which we take as the lag in days beyond which the lag autocorrelation of the time series falls below 0.2. As this time lag varies from one time series to the other, we give the DoF, which is the duration of the flight divided by the decorrelation time-scale, explicitly in column 2. Note that, for the decorrelation time, we use the daily averaged
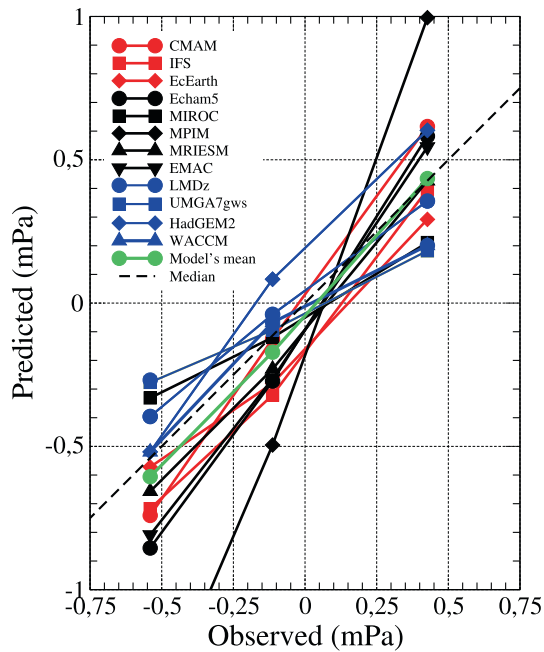
FIGURE 6 East, west, and cumulated zonal momentum fluxes averaged over the Strateole 2 phase 1 and 2 period and according to participating models. [Colour figure can be viewed at wileyonlinelibrary.com]

observations for simplicity, but found that it is not much different from that evaluated with the offline estimates (not shown).

When we calculate the daily correlations for the other parameterization schemes, there is also strong variability between the flights, and the global correlations shown in Table 3 summarize the averaged behaviour well. First, and as for LMDz, the correlations evaluated using Phase 2 data stay robust compared with correlations evaluated using phase 1, whatever the level of correlation ("medium", "low", or "non-significant"). Second, many schemes managed to have "medium" correlations ($0.3 < C < 0.5$) in the eastward direction. The schemes having no or small correlations in the eastward direction (CMAM, HadGEM2, and WACCM) are characterized by the fact that in them the launch level is quite high. For instance, in CMAM it is near the tropopause, which strongly mitigates dynamical filtering between the launching level and the balloon altitude. Also interesting, HadGEM2 and WACCM have low or no correlations; in those two models and in the case of deep convection, waves are launched from quite high levels in the troposphere (not shown), suggesting that in those models as well, and for waves with strong eastward flux, there is not enough distance between launch levels and balloon altitude for dynamical filtering to be efficient.

The results in the westward direction are more intriguing. Here the correlations are always small except for one

scheme (LMDz), and some "low" correlations are found for the two schemes that often launch waves from quite near the tropopause (CMAM and HadGEM2). We have difficulties in interpreting this last result. It may means that launching some waves from near the tropopause can improve the westward correlations (as CMAM and HadGEM2 do here), and that always launching waves from the same altitude well in the troposphere fails in most cases. But if this is true, the fact that LMDz westward MFs correlate better with observations than any other schemes is in contradiction. Maybe the skill of LMDz comes from elsewhere: for instance, from the fact that LMDz is a "multiwaves" scheme that explicitly launches waves according to their intrinsic frequency, a choice that directly affects dynamical filtering, whereas in the globally spectral schemes the dynamical filtering is more indirect and in the other two multiwave schemes (HadGEM2 and WACCM) the waves are launched according to their absolute frequency. These are merely speculations given here to emphasize the differences that are dynamically significant in our opinion; what is maybe more interesting to notice is that there is room to improve GW parameterizations to obtain better fits between predicted and measured fluxes in both directions of propagation, as illustrated by the case of LMDz.

For this latter model, with moderate intrinsic phase-speed range (see Table 4) and therefore strong explicit dynamical filtering, the fact that the westward correlations are significant but lower than the eastward correlations could suggest that the two phases of the campaign are during comparable eastward phases of the QBO, that is, when eastward waves are larger in amplitude and more easy to detect (here we assume that waves near saturation have larger amplitude than unsaturated waves and are therefore more easy to detect). It would be instructive to document whether westward correlations could be larger when the QBO is westward at the balloon level.

As stated in the Introduction, more than predicting the right fluxes at the right time, it is often believed that parameterizations should more importantly be validated against their statistical behaviour. A example is that observed GW MFs are strongly intermittent, a statistical character that deeply impacts the effect of the waves on the climate in the middle atmosphere (de la Cámara *et al.*, 2016). In a recent article, Green *et al.* (2024) showed that this intermittent behaviour is captured well when the GW momentum fluxes have pdfs following a log-normal distribution. These authors even concluded that, in all directions of propagation, momentum-flux characteristics could be summarized in terms of the mean and variance of log-normal distributions. As seen in Figure 7, such log-normal distributions describe accurately the daily average of the

TABLE 4 Some parameters changing between LMDz, HadGEM2, and WACCM, for information only, the schemes being extremely distinct from one another.

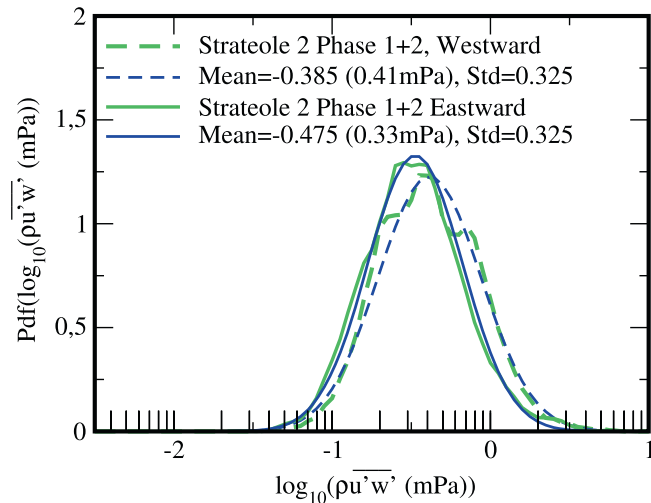|	p_1	Phase speed	Δz	Source
LMDz	500 hPa	Intrinsic = Gauss (0 m/s, 30 m/s)	1 km	Precip[2]
HadGEM2	850–100 hPa	−100 m/s < Absolute < 100 m/s	1–15 km	(Convective heating)[2]
WACCM	1000–100 hPa	−100 m/s < Absolute < 100 m/s	1–4 km	(Convective heating)[2]

FIGURE 7 Pdfs of daily values of momentum flux distribution evaluated from Strateole phases 1 and 2. The pdfs are calculated from histograms of 1291 MF daily values within intervals of $\Delta\left(\log_{10}\overline{\rho u'w'}(\text{mPa})\right) = 0.05$, thereafter smoothed by a five-point non-recursive filter with weight (0.1, 0.2, 0.4, 0.2, 0.1). Measured values are in green, log-normal fits are in blue. Solid lines are for eastward, dashed lines are for westward. Here the log-normal probability density function is defined as $P(X) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(X-M)^2/(2S^2)}$, where $X = \log_{10}\rho|\overline{u'w'}|$, and M and S are the mean and standard deviations given in the legend. [Colour figure can be viewed at wileyonlinelibrary.com]

MFs due to waves with periods between 15 min and 1 h deduced from Strateole 2 data. In this figure, the pdfs of the observed fluxes and the log-normal fits are shown in green and blue, respectively. The fluxes are seen to range in amplitude from 0.1–10 mPa. Furthermore, the pdfs of the westward fluxes are seen to be shifted toward higher values compared with those for the eastward fluxes, with little change to the shapes of the curves. The figure also shows that the shifts in the pdfs between eastward and westward fluxes are also well described by shifts in means and variances of log-normal distributions. For completeness, note nevertheless that, because we use daily averages and also because our dataset is quite short, we miss departures from log-normality that occurs for the lower values of MFs documented in Ern *et al.* (2022).

To analyse the QBOi schemes in this framework, Figure 8 presents pdfs of the distributions of the parameterized daily values of momentum fluxes. We see that for the WMI schemes (model names in red) the pdfs are much broader than the observed pdfs (green curves), and often far from log-normal. CMAM and EC-earth, for instance, exhibit peaks in the pdfs not located in the middle of the distribution. Quite remarkably, the HDS schemes (model names in black) are more realistic: the pdfs are narrower and much closer to log-normal distributions. It is important to note that, in all the globally spectral schemes without convective sources (WMI and HDS), the shift of the westward pdfs toward higher values compared with the eastward pdfs is reproduced (except for CMAM). Finally, the schemes that relate GWs to convection (names in blue) all have much broader pdfs, with long tails toward small values of MFs. These tails are not realistic, which suggests that these parameterizations miss a background of wave activity that exists even in the absence of convection nearby. In addition, the shift of the westward pdfs toward higher values than the eastward pdfs is not apparent. Instead, larger westward fluxes eventually occur as a result of changes in pdf rather than through translations (see for instance UMGA7gws and HadGEM2). If we now return to the conclusion of Green *et al.* (2024) that differences in GW momentum fluxes between directions of propagation could essentially be summarized by log-normal pdfs shifted by differences in mean values, one sees that including sources in single-column parameterizations is not necessarily skilful to achieve this objective. Finally, note that the WACCM scheme has a larger tail toward higher values (10 mPa) than the other schemes; this tail is consistent with the fact that some balloons have very large fluxes on average (see Figure 6).

4 CONCLUSION

The main result of this article is that state-of-the-art parameterizations of GWs reproduce reasonably well the eastward and westward values of the momentum fluxes due to high-frequency waves (periods between 15 min and
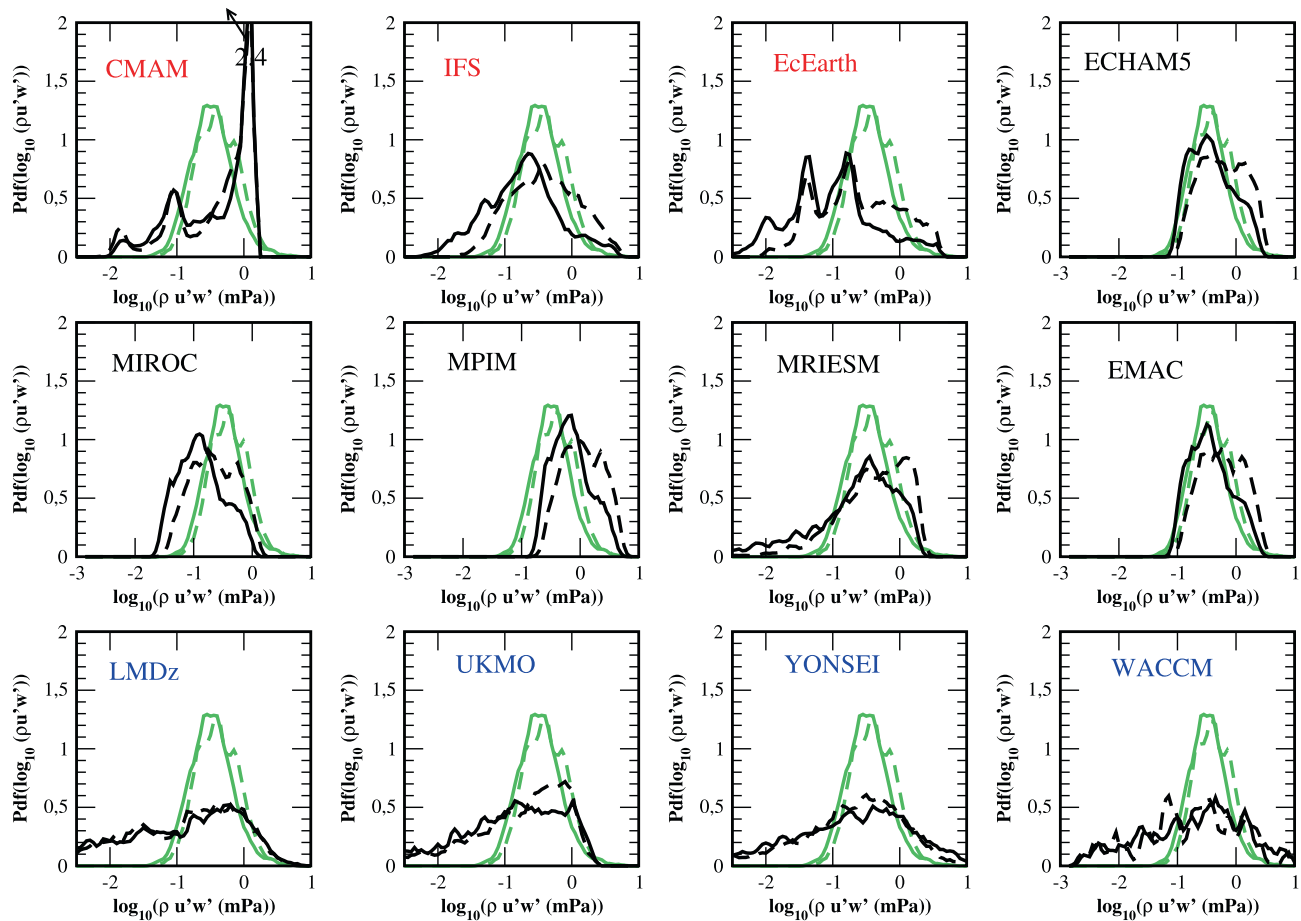
FIGURE 8 Pdfs of daily values of momentum flux distribution, same method as in Figure 7. Measured values are in green, estimations using ERA5 data and the parameterizations are in black. Solid lines are for eastward, dashed lines are for westward. [Colour figure can be viewed at wileyonlinelibrary.com]

1 h) deduced from in situ measurements made on board constant-level balloons. In terms of day-to-day variations, our results are more mitigated: in the eastward direction and without prior tuning, many schemes (multiwave HDS or WMI) present "medium" correlations with observations; in the westward direction only two schemes produce significant ("low" and "medium") correlations. As these results were obtained without any tuning, they suggest that reaching a medium level of correlation with daily observations is a possible objective, even in the westward direction. We also found that for our dataset it is easier to find significant correlations for eastward waves than for westward waves; this is probably related to the phase of the QBO at the balloons' altitudes. It is therefore important to plan another campaign in the opposite phase of the QBO.

Due to the low to medium level of the day-to-day correlations we found, we could ask ourselves whether it is mandatory to improve GW schemes according to such criteria. After all, when the momentum fluxes are averaged over periods near a month (here we rather consider averages over balloon flights), the correlations become "medium" to "strong" in the eastward direction (see

Figure 5) and sometimes medium in the westward direction. Such a level of correlation is probably enough in the context of QBO forcing, since the QBO is evolving over time-scales much longer than a month.

Substantial differences are also found when we compare the pdfs of the parameterized momentum fluxes with the pdfs of the measured fluxes. The spectral schemes following the Hines Doppler Spread parameterization (HDS) behave the most realistically in this respect. The pdfs for the HDS schemes exhibit one isolated maximum and extend broadly along a log-normal curve of about the right width. The HDS schemes also reproduce the shift of the pdfs toward larger values for the westward MFs, something that the Warner and McIntyre schemes (WMI) also do. The fact that both the HDS and WMI spectral schemes reproduce these characteristics is an interesting result. In these schemes, the source amplitude is constant and they are supposed to represent a broad ensemble of waves, two factors that could make them much less intermittent than the multiwave schemes including sources explicitly. It happens that for these schemes the dynamical filtering is efficient enough to reproduce a log-normal pdf

shifted according to the wave directions. This is important, since log-normal behaviours are significant to the model climate; they capture in good part the intermittency (Green *et al.*, 2024) needed in some models to represent well the final warming in the Southern Hemisphere (de la Cámara *et al.*, 2014) or the fluctuations of QBO periodicity (Lott & Guez, 2013). As dynamical filtering is also important to produce log-normal pdfs (Hertzog *et al.*, 2012), it is also not surprising that CMAM fails in reproducing such a distribution: it launches waves from too near the balloon height for them to be influenced by the vertical variations in wind during their propagation.

The schemes that relate GWs to convection also have broad momentum-flux pdfs, much broader than the spectral schemes. In this sense, therefore, they can be viewed as being even more intermittent than the spectral schemes. Furthermore, they are also characterized by long tails toward small values, which seems unrealistic. For these schemes it therefore seems important to add a background of wave activity even in the absence of convection. This problem could also be corrected in part by introducing lateral propagation (Amemiya & Sato, 2016; Kim *et al.*, 2024), a process that is important in the balloon observations used here (Corcos *et al.*, 2021), but this will not be sufficient over quite large and dry regions.

We did not try to fit the parameters of the schemes we use in order to improve daily correlations or pdfs or both, but we plan to do it in the near future. We do not have much data, however, but we could use the Loon data post-processed in a comparable way to Strateole 2 by Green *et al.* (2023), which would permit coverage of much wider regions. We could also complement these observations with convection-permitting global models, an outcome that looks promising (Köhler *et al.*, 2023; Stephan *et al.*, 2016; Sun *et al.*, 2023). We should also test whether improving the schemes' parameters to improve the fit with observations improves or does not degrade the model climate. It may well be that parameterizations compensate for potentially resolved equatorial waves, for instance, the latter showing a lot of variability between CMIP5 and QBOi models (Holt *et al.*, 2022; Lott *et al.*, 2014). Also, we hope that a better fit with observed values would help reduce persistent systematic errors in the QBO simulations, one of them being that models underestimate the QBO amplitude in the low stratosphere (Bushell *et al.*, 2022). Unfortunately, our results are not too promising in this regard: a common belief is that such an error could well be reduced by launching waves from near the tropopause, but the parameterizations that do so here are not very realistic when it comes to predict observed MF variabilities (over days or months).

We have also tried to identify whether some characteristics or deficiencies in the quasi-biennial oscillations simulated in QBOi models and reported in QBOi articles could be related to one characteristic or other of the schemes we have analysed. We did not find any, in terms of either periodicities, amplitude, asymmetry as documented in Bushell *et al.* (2022), teleconnections with the midlatitudes (Anstey *et al.*, 2022), response to climate change Richter *et al.* (2022), equatorial wave amplitude (Holt *et al.*, 2022), or semi-annual oscillation (Smith *et al.*, 2022). Of course, in these simulations potential signatures of the parameterization characteristics can be hidden by the many other model differences. A more thorough test would consist of comparing the different parameterizations online within the same model. The challenge in this case would be to compare differences once a common target is achieved (for instance, the QBO period and variability at 70 hPa).

AFFILIATIONS

[1]LMD/IPSL, Sorbonne Université, PSL Research Institute, Ecole Normale Supérieure, Paris, France

[2]Canadian Centre for Climate Modelling and Analysis (CCCma), Victoria, British Columbia Canada

[3]Met Office, Exeter, UK

[4]NorthWest Research Associates, Boulder Office, Boulder, Colorado, USA

[5]Yonsei University, Seoul, South Korea

[6]LMD/IPSL, Sorbonne Université, Paris, France

[7]Institut für Atmosphäre und Umwelt, Goethe-Universität, Frankfurt am Main, Germany

[8]Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan

[9]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan

[10]Max Planck Institute for Meteorology, Hamburg, Germany

[11]Meteorological Research Institute (MRI), Tsukuba, Japan

[12]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK

[13]LMD/IPSL, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

[14]National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA

[15]Group of Meteorology Universitat de Barcelona, Barcelona, Spain

[16]Institute of Marine Sciences, National Research Council, Rome, Italy

[17]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

[18]Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany

DATA AVAILABILITY STATEMENT

All data and routines needed to run the parameterizations in offline mode and to compare the results with the Strateole 2 flights are available on a dedicated web site at https://web.lmd.jussieu.fr/~flott/DATA /Documentation.pdf.

ORCID

F. Lott https://orcid.org/0000-0003-2126-5510

H. Naoe https://orcid.org/0000-0002-6261-0854

S. Osprey https://orcid.org/0000-0002-8751-1211

J. H. Richter https://orcid.org/0000-0001-7048-0781

F. Serva https://orcid.org/0000-0002-7118-0817

T. Stockdale https://orcid.org/0000-0002-7901-0337

REFERENCES

Achatz, U., Kim, Y.-H. & Voelker, G.S. (2023) Multi-scale dynamics of the interaction between waves and mean flows: from nonlinear WKB theory to gravity-wave parameterizations in weather and climate models. *Journal of Mathematical Physics*, 64, 111101. Available from: https://doi.org/10.1063/5.0165180

Alexander, M.J., Beres, J.H. & Pfister, L. (2000) Tropical stratospheric gravity wave activity and relationships to clouds. *Journal of Geophysical Research: Atmospheres*, 105, 22299–22309.

Alexander, M.J. & Dunkerton, T.J. (1999) A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, 56, 4167–4182.

Alexander, M.J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F. et al. (2010) Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quarterly Journal of the Royal Meteorological Society*, 136, 1103–1124.

Alexander, M.J., Liu, C.C., Bacmeister, J., Bramberger, M., Hertzog, A. & Richter, J.H. (2021) Observational validation of parameterized gravity waves from tropical convection in the whole atmosphere community climate model. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033954.

Amemiya, A. & Sato, K. (2016) A new gravity wave parameterization including three-dimensional propagation. *Journal of the Meteorological Society of Japan Series II*, 94, 237–256.

Andrews, F.G., Holton, J. & Leovy, C. (1987) *Middle atmosphere dynamics*. SanDiego: Academic Press.

Anstey, J.A., Banyard, T.P., Butchart, N., Coy, L., Newman, P.A., Osprey, S. et al. (2021) Prospect of increased disruption to the QBO in a changing climate. *Geophysical Research Letters*, 48, e2021GL093058.

Anstey, J.A., Scinocca, J.F. & Keller, M. (2016) Simulating the qbo in an atmospheric general circulation model: sensitivity to resolved and parameterized forcing. *Journal of the Atmospheric Sciences*, 73, 1649–1665.

Anstey, J.A., Simpson, I.R., Richter, J.H., Naoe, H., Taguchi, M., Serva, F. et al. (2022) Teleconnections of the quasi-biennial oscillation in a multi-model ensemble of qbo-resolving models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1568–1592.

Baldwin, M.P., Gray, L.J., Dunkerton, T.J., Hamilton, K., Haynes, P.H., Randel, W.J. et al. (2001) The quasi-biennial oscillation. *Reviews of Geophysics*, 39, 179–229.

Beres, J.H., Garcia, R.R., Boville, B.A. & Sassi, F. (2005) Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the whole atmosphere community climate model (waccm). *Journal of Geophysical Research: Atmospheres*, 110, D10108.

Bushell, A.C., Anstey, J.A., Butchart, N., Kawatani, Y., Osprey, S.M., Richter, J.H. et al. (2022) Evaluation of the quasi-biennial oscillation in global climate models for the SPARC-QBO initiative. *Quarterly Journal of the Royal Meteorological Society*, 148, 1459–1489.

Bushell, A.C., Butchart, N., Derbyshire, S.H., Jackson, D.R., Shutts, G.J., Vosper, S.B. et al. (2015) Parameterized gravity wave momentum fluxes from sources related to convection and large-scale precipitation processes in a global atmosphere model. *Journal of the Atmospheric Sciences*, 72, 4349–4371.

Butchart, N., Anstey, J.A., Hamilton, K., Osprey, S., McLandress, C., Bushell, A.C. et al. (2018) Overview of experiment design and comparison of models participating in phase 1 of the sparc quasi-biennial oscillation initiative ("qboi"). *Geoscientific Model Development*, 11, 1009–1032.

Charron, M. & Manzini, E. (2002) Gravity waves from fronts: parameterization and middle atmosphere response in a general circulation model. *Journal of the Atmospheric Sciences*, 59, 923–941.

Choi, H.-J. & Chun, H.-Y. (2011) Momentum flux spectrum of convective gravity waves. Part i: an update of a parameterization using mesoscale simulations. *Journal of the Atmospheric Sciences*, 68, 739–759.

Christiansen, B., Yang, S. & Madsen, M.S. (2016) Do strong warm enso events control the phase of the stratospheric qbo? *Geophysical Research Letters*, 43, 10489–10495.

Corcos, M., Hertzog, A., Plougonven, R. & Podglajen, A. (2021) Observation of gravity waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical Research: Atmospheres*, 126, e2021JD035165.

Davini, P., von Hardenberg, J., Corti, S., Christensen, H.M., Juricke, S., Subramanian, A. et al. (2017) Climate sphinx: evaluating the impact of resolution and stochastic physics parameterisations in the ec-earth global climate model. *Geoscientific Model Development*, 10, 1383–1402.

de la Cámara, A. & Lott, F. (2015) A parameterization of gravity waves emitted by fronts and jets. *Geophysical Research Letters*, 42, 2071–2078.

de la Cámara, A., Lott, F. & Hertzog, A. (2014) Intermittency in a stochastic parameterization of nonorographic gravity waves. *Journal of Geophysical Research: Atmospheres*, 119, 11905–11919.

de la Cámara, A., Lott, F., Jewtoukoff, V., Plougonven, R. & Hertzog, A. (2016) On the gravity wave forcing during the southern stratospheric final warming in LMDZ. *Journal of the Atmospheric Sciences*, 73, 3213–3226.

Eckermann, S.D. (2011) Explicitly stochastic parameterization of nonorographic gravity wave drag. *Journal of the Atmospheric Sciences*, 68, 1749–1765.

Ern, M., Ploeger, F., Preusse, P., Gille, J., Gray, L.J., Kalisch, S. et al. (2014) Interaction of gravity waves with the QBO: a satellite perspective. *Journal of Geophysical Research: Atmospheres*, 119, 2329–2355.

Ern, M., Preusse, P. & Riese, M. (2022) Intermittency of gravity wave potential energies and absolute momentum fluxes derived from infrared limb sounding satellite observations. *Atmospheric Chemistry and Physics*, 22, 15093–15133.

Fovell, R., Durran, D. & Holton, J.R. (1992) Numerical simulations of convectively generated stratospheric gravity waves. *Journal of Atmospheric Sciences*, 49, 1427–1442.

Fueglistaler, S., Legras, B., Beljaars, A., Morcrette, J.-J., Simmons, A., Tompkins, A.M. et al. (2009) The diabatic heat budget of the

upper troposphere and lower/mid stratosphere in ecmwf reanalyses. *Quarterly Journal of the Royal Meteorological Society*, 135, 21–37.

Geller, M.A., Alexander, M.J., Love, P.T., Bacmeister, J., Ern, M., Hertzog, A. et al. (2013) A comparison between gravity wave momentum fluxes in observations and climate models. *Journal of the Atmospheric Sciences*, 26, 6383–6405.

Green, B., Sheshadri, A., Alexander, M., Bramberger, M. & Lott, F. (2024) Gravity wave momentum fluxes estimated from project loon balloon data. *Journal of Geophysical Research: Atmospheres*, 129, e2023JD039927.

Haase, J.S., Alexander, M.J., Hertzog, A., Kalnajs, L.E., Deshler, T., Davis, S.M. et al. (2018) Around the world in 84 days [dataset]. *Eos*, 99. https://eos.org/science-updates/around-the-world-in-84-days

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J. et al. (2020) The ERA5 global reanalysis [dataset]. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Hertzog, A. (2007) The stratéole-vorcore long-duration balloon experiment: a personal perspective. *Space Research Today*, 169, 43–48.

Hertzog, A., Alexander, M.J. & Plougonven, R. (2012) On the intermittency of gravity wave momentum flux in the stratosphere. *Journal of the Atmospheric Sciences*, 69, 3433–3448.

Hines, C.O. (1991) The saturation of gravity waves in the middle atmosphere. Part ii: development of doppler-spread theory. *Journal of Atmospheric Sciences*, 48, 1361–1379.

Hines, C.O. (1997) Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 2: broad and quasi monochromatic spectra, and implementation. *Journal of Atmospheric and Solar Terresrial Physics*, 59, 387–400.

Holt, L.A., Lott, F., Garcia, R.R., Kiladis, G.N., Cheng, Y.-M., Anstey, J.A. et al. (2022) An evaluation of tropical waves and wave forcing of the QBO in the QBOi models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1541–1567.

Jewtoukoff, V., Hertzog, A., Plougonven, R., de la Cámara, A. & Lott, F. (2015) Comparison of gravity waves in the southern hemisphere derived from balloon observations and the ecmwf analyses. *Journal of the Atmospheric Sciences*, 72, 3449–3468.

Jewtoukoff, V., Plougonven, R. & Hertzog, A. (2013) Gravity waves generated by deep tropical convection: estimates from balloon observations and mesoscale simulations. *Journal of Geophysical Research: Atmospheres*, 118, 9690–9707.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H. et al. (2010) Development cycle 2 of the modular earth submodel system (messy2). *Geoscientific Model Development*, 3, 717–752.

Kang, M.-J., Chun, H.-Y. & Kim, Y.-H. (2017) Momentum flux of convective gravity waves derived from an offline gravity wave parameterization. Part i: spatiotemporal variations at source level. *Journal of the Atmospheric Sciences*, 74, 3167–3189.

Kim, Y.-H., Voelker, G.S., Bölöni, G., Zängl, G. & Achatz, U. (2024) Crucial role of obliquely propagating gravity waves in the quasi-biennial oscillation dynamics. *Atmospheric Chemistry and Physics*, 24, 3297–3308.

Köhler, L., Green, B. & Stephan, C.C. (2023) Comparing loon superpressure balloon observations of gravity waves in the tropics with global storm-resolving models. *Journal of Geophysical Research: Atmospheres*, 128, e2023JD038549.

Lane, T.P. & Moncrieff, M.W. (2008) Stratospheric gravity waves generated by multiscale tropical convection. *Journal of the Atmospheric Sciences*, 65, 2598–2614.

Lindzen, R.S. (1981) Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical Research*, 86, 9707–9714.

Liu, C., Alexander, J., Richter, J. & Bacmeister, J. (2022) Using trmm latent heat as a source to estimate convection induced gravity wave momentum flux in the lower stratosphere. *Journal of Geophysical Research: Atmospheres*, 127, e2021JD035785.

Lott, F., Denvil, S., Butchart, N., Cagnazzo, C., Giorgetta, M.A., Hardiman, S.C. et al. (2014) Kelvin and rossby-gravity wave packets in the lower stratosphere of some high-top cmip5 models. *Journal of Geophysical Research: Atmospheres*, 119, 2156–2173.

Lott, F. & Guez, L. (2013) A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *Journal of Geophysical Research*, 118, 8897–8909.

Lott, F., Guez, L. & Maury, P. (2012a) A stochastic parameterization of non-orographic gravity waves: formalism and impact on the equatorial stratosphere. *Geophysical Research Letters*, 39, L06807.

Lott, F., Plougonven, R. & Vanneste, J. (2012b) Gravity waves generated by sheared three-dimensional potential vorticity anomalies. *Journal of the Atmospheric Sciences*, 69, 2134–2151.

Lott, F., Rani, R., Podglajen, A., Codron, F., Guez, L., Hertzog, A. et al. (2023) Direct comparison between a non-orographic gravity wave drag scheme and constant level balloons. *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037585.

Manzini, E., McFarlane, N.A. & McLandress, C. (1997) Impact of the doppler spread parameterization on the simulation of the middle atmosphere circulation using the ma/echam4 general circulation model. *Journal of Geophysical Research: Atmospheres*, 102, 25751–25762.

Naoe, H. & Yoshida, K. (2019) Influence of quasi-biennial oscillation on the boreal winter extratropical stratosphere in qboi experiments. *Quarterly Journal of the Royal Meteorological Society*, 145, 2755–2771.

Orr, A., Bechtold, P., Scinocca, J., Ern, M. & Janiskova, M. (2010) Improved middle atmosphere climate and forecasts in the ecmwf model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, 23, 5905–5926.

Piani, C., Norton, W.A. & Stainforth, D.A. (2004) Equatorial stratospheric response to variations in deterministic and stochastic gravity wave parameterizations. *Journal of Geophysical Research: Atmospheres*, 109, D14101.

Plougonven, R., Jewtoukoff, V., de la Cámara, A., Lott, F. & Hertzog, A. (2017) On the relation between gravity waves and wind speed in the lower stratosphere over the southern ocean. *Journal of the Atmospheric Sciences*, 74, 1075–1093.

Pohlmann, H., Müller, W.A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F.S.E. et al. (2013) Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophysical Research Letters*, 40, 5798–5802.

Rabier, F., Bouchard, A., Brun, E., Doerenbecher, A., Guedj, S., Guidard, V. et al. (2010) The Concordiasi project in Antarctica. *Bulletin of the American Meteorological Society*, 91, 69–86.

Richter, J.H., Butchart, N., Kawatani, Y., Bushell, A.C., Holt, L., Serva, F. et al. (2022) Response of the quasi-biennial oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1490–1518.

Richter, J.H., Sassi, F. & Garcia, R.R. (2010a) Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, 67, 136–156.

Richter, J.H., Sassi, F. & Garcia, R.R. (2010b) Toward a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, 67, 136–156.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L. et al. (2006) Sensitivity of simulated climate to horizontal and vertical resolution in the echam5 atmosphere model. *Journal of Climate*, 19, 3771–3791.

Scaife, A.A., Butchart, N., Warner, C.D. & Swinbank, R. (2002) Impact of a spectral gravity wave parameterization on the stratosphere in the met office unified model. *Journal of the Atmospheric Sciences*, 59, 1473–1489.

Scinocca, J.F. (2003) An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *Journal of the Atmospheric Sciences*, 60, 667–682.

Serva, F., Cagnazzo, C., Riccio, A. & Manzini, E. (2018) Impact of a stochastic nonorographic gravity wave parameterization on the stratospheric dynamics of a general circulation model. *Journal of Advances in Modeling Earth Systems*, 10, 2147–2162.

Smith, A.K., Holt, L.A., Garcia, R.R., Anstey, J.A., Serva, F., Butchart, N. et al. (2022) The equatorial stratospheric semi-annual oscillation and time-mean winds in qboi models. *Quarterly Journal of the Royal Meteorological Society*, 148, 1593–1609.

Song, I.-S. & Chun, H.-Y. (2005) Momentum flux spectrum of convectively forced internal gravity waves and its application to gravity wave drag parameterization. Part i: theory. *Journal of the Atmospheric Sciences*, 62, 107–124.

Stephan, C., Alexander, M.J. & Richter, J.H. (2016) Characteristics of gravity waves from convection and implications for their parameterization in global circulation models. *Journal of the Atmospheric Sciences*, 73, 2729–2742.

Sun, Y.Q., Hassanzadeh, P., Alexander, M.J. & Kruse, C.G. (2023) Quantifying 3d gravity wave drag in a library of tropical convection-permitting simulations for data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003585.

Warner, C.D. & McIntyre, M.E. (1996) On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *Journal of the Atmospheric Sciences*, 53, 3213–3235.

Warner, C.D. & McIntyre, M.E. (1999) Toward an ultra-simple spectral gravity wave parameterization for general circulation models. *Earth, Planets and Space*, 51, 475–484.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H. et al. (2011) Miroc-esm 2010: model description and basic results of cmip5-20c3m experiments. *Geoscientific Model Development*, 4, 845–872.