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

[1]Laboratoire de Météorologie Dynamique (LMD)/IPSL, PSL Research Institute, Ecole Normale Supérieure, Paris, France.

[2]LMD/IPSL, Sorbone Université, Paris, France.

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

[4]Canadian Centre for Climate Modelling and Analysis (CCCma), Victoria, Canada

[5]Met Office, FitzRoy Road, Exeter, UK

[6]Yonsei University, Seoul, South Korea

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

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

[9]NorthWest Research Associates, Boulder Office, Boulder, CO, USA

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

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

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

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

[14]Institute of Marine Sciences, National

Gravity Waves (GWs) parameterizations from 12 General Circulation Models (GCMs) participating in the Quasi-Biennial Oscillation initiative (QBOi) are directly compared to Strateole-2 balloon observations made in the tropical lower stratosphere from November 2019 to February 2020 (phase 1) and from October 2021 and January 2022 (phase 2). The parameterizations used employ the 3 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 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 the waves with periods less than 1 hr and the 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 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 pdfs of the momentum fluxes are better represented in spectral schemes with constant sources than in schemes ("spectral" or "multiwaves") that relate GWs to their convective sources.

KEYWORDS

Gravity Waves, Balloon Observations, Quasi-Biennial Oscillation, Global Climate Models

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 horizontal resolution of conventional atmospheric General Circulation Models (GCMs) they need to be parameterized (Alexander and Dunkerton, 1999). In the tropics, the GWs generated by convection are believed to largely dominate (Fovell et al., 1992; Alexander et al., 2000; Lane and 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 (Scinocca, 2003; Song and Chun, 2005; Beres et al., 2005; Orr et al., 2010; Lott and Guez, 2013; Bushell et al., 2015; Anstey et al., 2016; Christiansen et al., 2016; Serva et al., 2018), 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 force the QBO 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 few large-amplitude waves that potentially break at lower altitudes rather than by a GW field that is more temporally uniform. This intermittent character, when reproduced by a parameterization (de la Cámara et al., 2014; Kang et al., 2017; Alexander et al., 2021), 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 on the simulation of the QBO (Lott et al., 2012). Balloon observations have also been used to characterize the dynamical filtering by the large scale winds (Plougonven et al., 2017), and to validate the average statistical properties of the GW momentum flux simulated offline using reanalysis data (Kang et al., 2017; Alexander et al., 2021). Note that here and in the following we refer to dynamical filtering as the process by which waves with smaller amplitude intrinsic phase speed saturate or start to break for smaller values of the MF than the waves with larger amplitude intrisic phase speed (for multiwaves parameterisations see the Eq. (3) for the saturated stress and the associated discussion in Lott et al. (2023)).

However, previous evaluations of parameterizations using balloon observations were often quite indirect and related more to their statistical behaviours (Jewtoukoff et al., 2015; Kang et al., 2017; Alexander et al., 2021) rather than their ability to directly 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 and Sato (2016), 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 explicitly relate launched waves to sources, and there is a continuing effort to improve the realism of these convective sources (Liu et al., 2022). Finally, observations systematically suggest 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 directly compared to those measured by a balloon at the same place.

Based on the relative success of previous offline calculations using reanalysis data (Jewtoukoff et al., 2015; Kang et al., 2017; Alexander et al., 2021), 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 whose amplitudes can be directly compared 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 the 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 the 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 function of the predicted momentum flux amplitudes have long tails for low values which are absent in observations. This suggests that the parameterization misses processes like lateral propagation or the presence of a background of waves whose origin remains a challenge to predict.

The purpose of this paper 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 8 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 ERA5 reanalysis (Hersbach et al., 2020) that is nearest to the balloon location. At those times and places we use the vertical profiles of wind and temperature, as well as the surface value of precipitation to calculate the parameterized GW momentum fluxes using the parameterizations used in the GCMs that participated to QBOi. We also extract from the analysis and from the associated 3hr forecast the analysis uses, the diabatic heatings rates and the cloud base and cloud top altitudes needed in some GW parameterization schemes.

The outline of the paper is as follows. Section 2 describes the data and the parameterization schemes used. Section 3 discusses the results in terms of daily correlations, as well as global averages and statistics. Section 4 summarizes the results. As we shall see the performances of each parameterization can be contrasted regarding that we use one type of result rather than other, but our purpose is not to promote one scheme in front of the others. Adapting other groups parameterization to a testbed that have been intensively used for LMDz (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 1980's 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; Song and Chun, 2005; Richter et al., 2010a; Lott and Guez, 2013; de la Cámara and Lott, 2015). In the following we will refer to such schemes as "multiwave". These schemes are expensive because they request a large number of harmonics to well represent a

	p_l	F_{LT}	$2\pi/m_*$	c_{min}
CMAM	100hPa	1.3mPa	1km	0.25 m/s
IFS	450hPa	5mPa	3km	0.5 m/s
ECEarth	450hPa	3.75mPa	2km	0.25 m/s
UMGA7gws	580hPa	$\sqrt{\text{Precip}}$	4.3km	not used

TABLE 1 WMI Parameters changing between CMAM, IFS, ECEarth, and UMGA7gws. UMGA7gws is shown distinctly because it is based on (Warner and McIntyre, 1999) simplified version of WMI rather than on (Scinocca, 2003)'s and realte launched MF to precipitations.

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

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

realistic wave field, but this limit can easily be circumvented by using stochastic approaches (Eckermann, 2011; Lott et al., 2012).

As an alternative, but also to better represent the effect of wave breaking, globally spectral schemes have been developed and used with success. These schemes use the observational fact that GWs produce kinetic energy spectra which have a quite universal shape when expressed as a function of vertical wavenumber. In the early 1990's Hines (1991) developed a theory where GW breaking is represented by imposing an upper limit to the range of vertical wavenumbers, the limit being calculated according to the large-scale wind and including a Doppler spreading by the other gravity waves (see also Hines (1997)). The scheme has been implemented with success 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 treat vertical changes in the spectrum following the propagation invariant characters of GWs. The theory has been simplified and/or optimized to permit implementation, for instance in the UKMO model (Warner and McIntyre, 1999; Scaife et al., 2002) and in the CMAM model (Scinocca, 2003), and will be referred henceforth as "WMI" for "Warner and McIntyre". To a certain extent, the 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 paper, we compare the GWs schemes used in 12 of the models that participate in QBOi, which all belong to one of the three types of schemes described above (WMI, HDS, and Multiwave). Since all the multiwave schemes used here relate GWs to their convective sources and since only one of the spectral scheme does so (i.e., the UMGA7gws WMI scheme in Bushell et al. (2015)), the spectral scheme in Bushell et al. (2015) will be discussed with the source-related multiwave schemes.

Among the 12 models, three (CMAM, IFS and ECEarth) use the Scinocca (2003) version of WMI. The specific

	p_l	Phase Speed	Δz	Source
LMDz	500hPa	Intrinsic=Gauss(0m/s, 30m/s)	1km	Precip^2
HadGEM2	850hPa-100hPa	-100m/s<Absolute<100m/s	1km-15km	(Convective Heating)^2
WACCM	1000hPa-100hPa	-100m/s<Absolute<100m/s	1km-4km	(Convective Heating)^2

TABLE 3 Some parameters changing between LMDz, HadGEM2 and WACCM, for information only the schemes being extremely distinct one from the other

versions of these schemes used for QBOi are further detailed 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 value of the parameters in Table 1 are from J. García-Serrano (private communication).

Five of the 12 models, uses the HDS parameterization discussed in Manzini et al. (1997): ECham5, MIROC, MPIM, MRI-ESM, and EMAC. Their version 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 mainly differ by three different parameters: the launching level 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 $m_{\rm min}$, and two parameters that control smoothing in the vertical of the GWs root mean square variance, the coefficient $C_{\rm smo}$ and the number of time the smoothing is applied $N_{\rm smo}$. It is important to note that in ECham5 the variability parameter σ is chosen randomly, with a normal distribution centered at 1m/s with standard deviation 0.2m/s. The usefulness of such a stochastic ingredient was initially proposed by Piani et al. (2004) who found that it can help stabilizing the QBO variability in large scale models and over decades.

Finally the last 4 schemes we consider all link GWs to sources (convection or precipitation). Three are multiwaves 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)'s 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), and which were adapted by these authors to interpret observations. Since the differences between the 3 multiwave schemes are too numerous, the reader is referred to the above mentionned papers. However, important differences can be outlined in the source term, the launching levels and the intrinsic phase speed of the launched waves. 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 and Guez, 2013) whereas in (Bushell et al., 2015) it is related to $\sqrt{P_r}$ (see Table 1). Furthemore in LMDz the waves are launched in the mid troposphere and in the UMGA7gws model whereas they are launched in the lower troposphere near below 4km altitude. In the HadGEM2 scheme (Song and Chun, 2005; Choi and Chun, 2011), 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 vary between $2km$ and 15km typically and the depth of the heating between 1km and 15km. 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 this paper the WACCM scheme was adapted and partly re-written to use direct satellite observations of convective heating. Also note that in WACCM, the heating depth is one quarter of the cloud depth, and ranges between 1km and 4km typically. Final important differences are that in LMDz the harmonics are chosen randomly according to a Gaussian distribution with 0 mean value and 30m/s standard deviation, whereas in both UMGA7gws and WACCM absolute phase speed is used, with values uniformly distributed in the range -100m/s $< C_{abs} <$ 100m/s.

2.2 | Offline parameterization runs

To run the schemes in offline mode we use ERA-5 hourly data of precipitation and 3-hourly winds, surface pressure, temperature, cloud liquid and ice water content on a $1^o \times 1^o$ horizontal grid to mimic a large-scale climate model of fairly high horizontal esolution. Winds, surface pressure, temperature, and water contents are then linearly interpolated on 1hr interval so that they are synchronised 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.

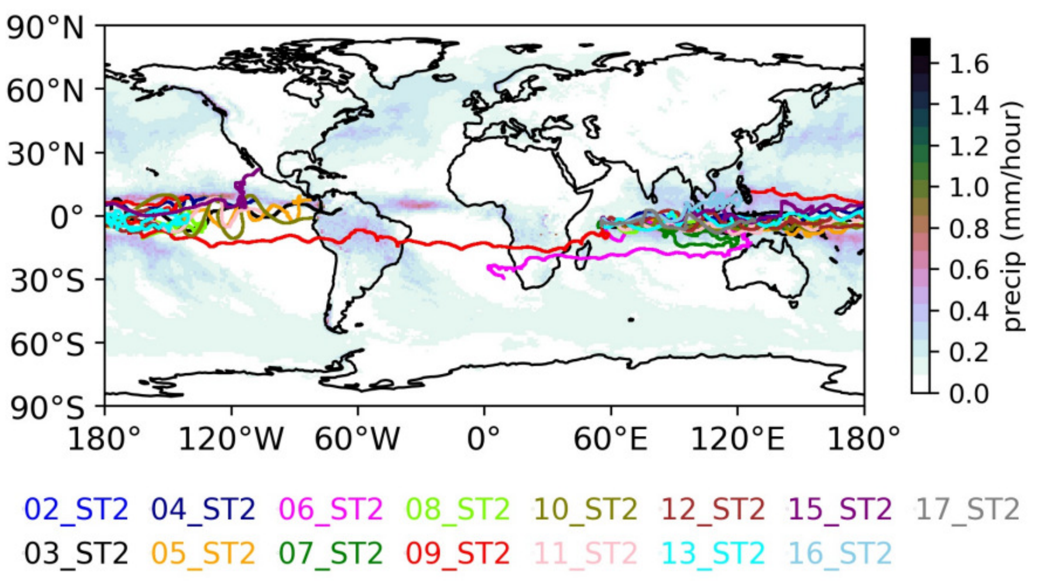

FIGURE 1 Strateole 2, Phase 2 balloon trajectories taking place between October 2021 and January 2022. Shading presents the precipitation field from ERA5 averaged over the period.

2.3 | Strateole 2 balloon observations

The in situ observations we use are from the 8 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

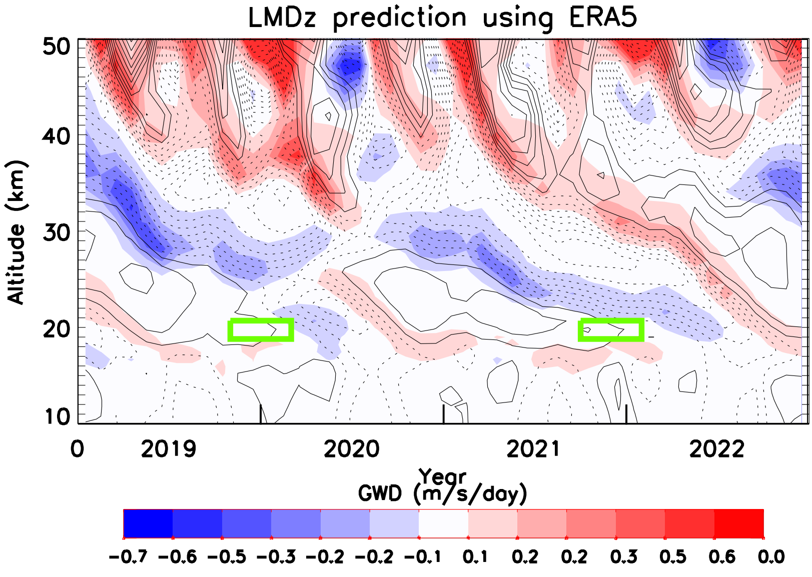

FIGURE 2 Time vertical sections of the zonal mean zonal wind (CI=10m/s, negative values are dashed and non-orographic gravity wave tendency averaged over the Equatorial band ($-6^o S - +6^o N$). Input data are from ERA5 reanalysis and GWs prediction from the LMDz scheme. The 2 green boxes indicate schematically the altitude and time ranges of the Strateole 2 phase 1 and 2 flights considered in this study.

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 momentum fluxes (MFs) calculated from observations, Corcos et al. (2021) distinguish the waves with short periods (1hr-15mn) from those with periods up to one day (1d-15mn). They also distinguish the eastward travelling waves with positive MFs in the zonal direction from the westward travelling waves with negative MFs.

To characterize the phase of the QBO during the balloon flights, Fig. 2 shows a time versus altitude cross section of the equatorial zonal mean zonal winds and 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 with the fact that it contributes to the QBO descent. We also note that the amplitudes vary between ±0.5m/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 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 flights altitude the 2nd 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 1hr and consider the ERA5 data at the points that are the 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 realistically sample a GW field, and also because it takes about one day for a balloon flight to cover a model gridscale. Note that some of the sensitivities to these choices are discussed in Lott et al. (2023)'s conclusion.

3 | RESULTS

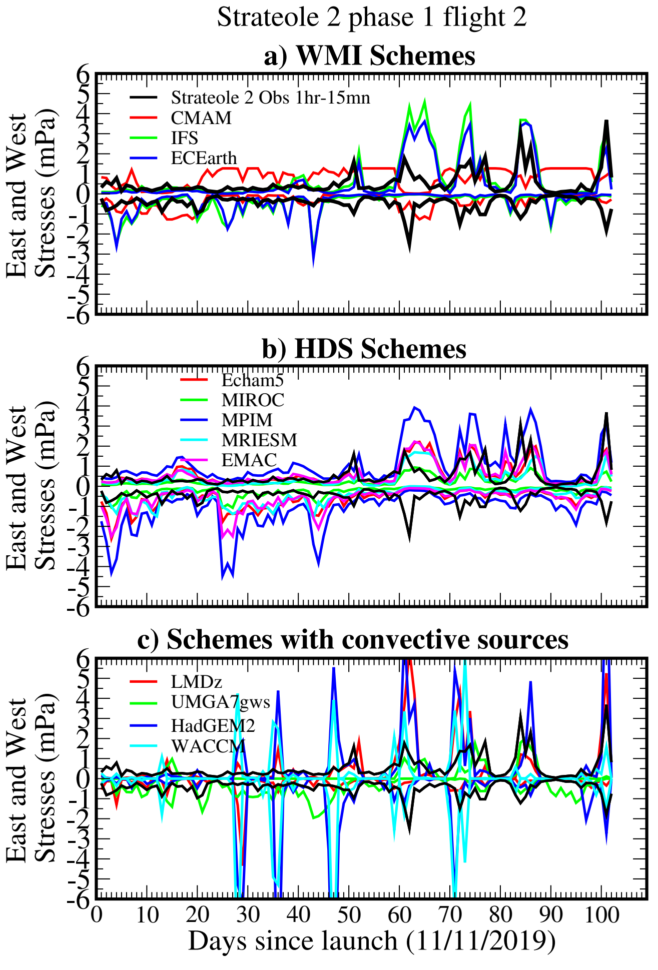

FIGURE 3 Comparison between daily averaged values of the eastward and westward MFs measured by the balloons during Strateole 2 phase 1 Flight 2 and estimated by the GW schemes at the balloon location and altitude. Colored curves are for the GW schemes using ERA5, black curves are for the observed MFs due to the 15mn-1hr GWs. a) WMI schemes; b) HDS Schemes; c) Schemes relating launched MFs with convective sources or precipitations: all multiwaves except UMGA7gws.

Figure 3 shows time series of daily values of momentum fluxes calculated by the parameterizations and measured during balloon flights 2 from strateole 2 phase 1. This is also the flight shown in Fig. 3 in Lott et al. (2023), in which was also shown the time series of daily precipitation and zonal wind at flight altitude. 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 somewhat agree with the observed ones, at least in term of amplitude. There are nevertheless significant differences in behaviour. For instance, the IFS scheme exhibits 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 favoring eastward waves consistent with dynamical filtering. Note that in Lott et al. (2023) it was shown that the 3 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 the relation with convective sources is essential. However, we see here that dynamical filtering alone may well be the main cause. Although having smaller amplitudes, the Fig. 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 it was chosen 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 (100hPa see Table 1) and the balloon altitude is too small for dynamical filtering to be efficient. The second panel of Fig. 3 for the HDS schemes is not fundamentally different from what was discussed above. The amplitude and fluctuations are comparable to observed, some schemes predicting values which look either larger or smaller but staying within the range of observations. The behaviour of the source related schemes in the third panel of Fig. 3 (multiwave for LMDz an HadGEM2, WMI for UMGA7gws) are 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 sometime exceed ±5mPa, which are values never reached by any of the spectral schemes in panels 3a) and 3b). In contrast to LMDz and HadGEM2, the UMGA7gws scheme exhibits MFs smaller amplitude 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.

The MF time series for a flight during the second phase of strateole 2 is shown in Fig. 4. Beyond the fact that the flight is shorter than in Fig. 3, a difference in duration that characterizes most of the flights during phase 2 compared to 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 balloon flight altitude. The last panel in Fig. 4 also shows that UMGA7gws exhibits long periods with almost no fluxes, which results from the launch height being low in the troposphere 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 outlayer at day 33, with values below -10mPa. We only found few of them over the entire campaign, and only in WACCM. It follows that WACCM has been tuned to produce sometimes and rarely extreme values in MFs, these extreme values significantly contribute to the averaged MFs.

The fact that the different schemes estimate momentum fluxes of about the right amplitude is summarized in Fig. 5 where the average of the fluxes over the 18 flights that last more than a month (8 during phase 1, 10 during phase 2) are shown. In this figure we see that the predicted values align quite well with the observed ones, though some schemes have a tendency to slightly underestimate the fluxes (MIROC, LMDz), and others to overestimate them (CMAM, HadGEM2). The WACCM scheme has a quite distinct behaviour, most balloons measure quite 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), and we cannot exclude that the WACCM model benefits from this. The numbers in each panel are the correlation coefficient 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 quite well the low frequency variability of the MFs (the changes with period

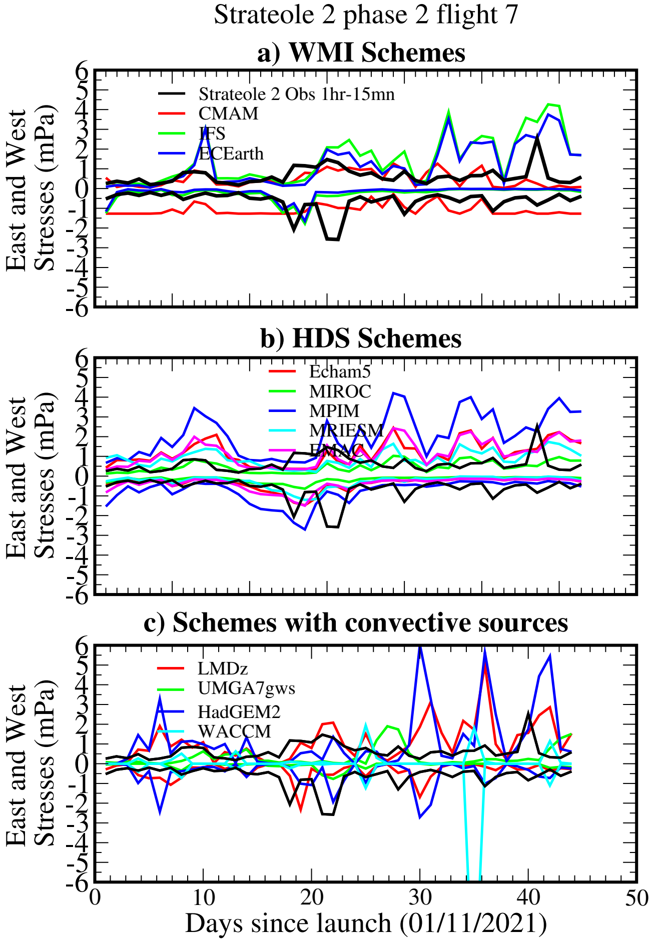

FIGURE 4 Same as Fig 3 but for Strateole 2 Phase 2 Flight 7.

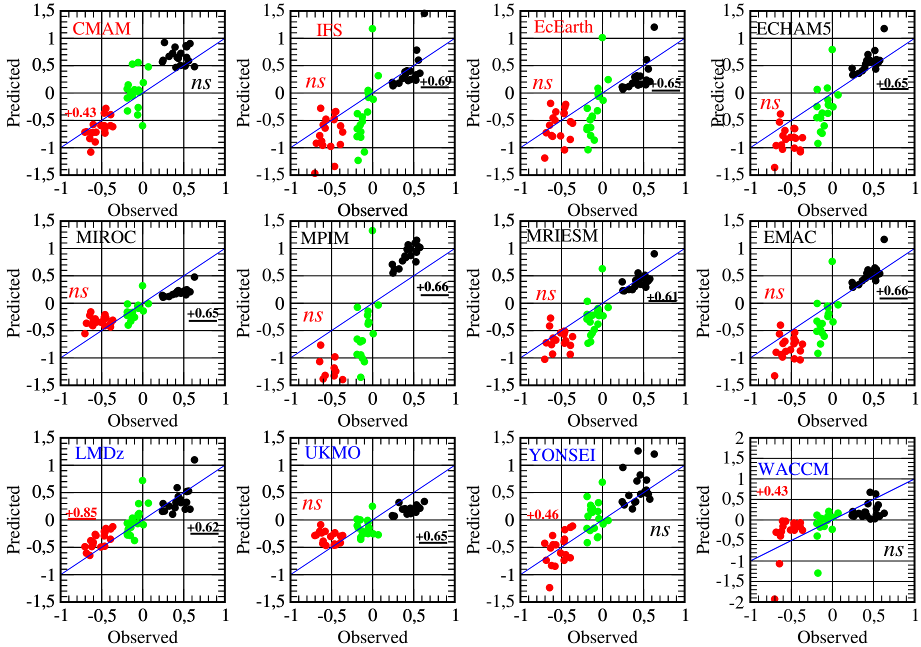

FIGURE 5 Scatter plot of the momentum fluxes measured by the balloon versus parameterized using different models. Only considered here the 18 balloon flights that last more than a month (East: black; West: red; Cumulated (East+West): green). Also shown are the correlations between observations and predictions, 99% significant levels are bold underlined, 95% are bold. Non significant values indicated by *"ns"*. The number of DoF for Pearson test is 23, which is simply the number of balloon flights and which is therefore very conservative, many balloons lasting more than few weeks, whereas the decorrelation time scale of the daily series being well below a week. Color of the names of the WMI, HDS, and convection-related GWs schemes are in red, black and blue respectively. Note the the change of vertical axis in lower left panel.

larger than a month). Thus, it is tempting to say that it is good enough for the 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 over all balloon flights. As can be seen the average amplitude of the eastward flux is very near that of the observed (10% overestimate: 0.45mPa for the parameterizations compared to 0.40mPa for the observed), whereas the westward flux is overestimated by the models by less than 20% (-0.65mPa for the parameterizations compare to -0.55mPa observed). This 10%-20% error explains the quite large relative error (50%) in the cumulated (i.e., east plus west) flux but for it the large relative error is in good due to the fact that large positive and negative fluxes oppose each other.

The daily time series in Figs 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 Figs. 3 and 4, the correspondence between the observed and parameterized fluxes is quite apparent, particularly

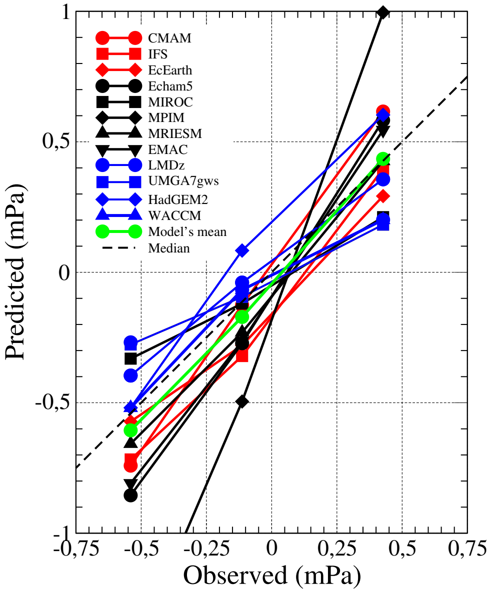

FIGURE 6 East, West and cumulated zonal momentum fluxes averaged over the Strateole 2 phase 1 and 2 period and according to participating models.

in the first (Figure 3) in regards to the peaks in the eastward direction discussed earlier. Correspondences are less apparent in the second case (Figure 4) where the observed MFs present less variations 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 computed using all of the flights that were smaller than for an individual flight like the one shown in Fig. 3. This resulted in correlations 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 9th column of Table 4. In this table are also given the coefficients for Phase 2 and for the phase 1 and 2 combined. Consistent with the results found for phase 1, we found during phase 2 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 day 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 explicitly the DoF in column 2, which is the duration of the flight divided by the decorrelation time scale. Note that for the decorrelation time, we use for simplicity the daily averaged observations, but found that it is not much different from that evaluated with the offline estimates (not shown).

If we now look at the results for the other parameterization schemes, the results are contrasted but quite in agreement. A lot a variations between flights (not shown) the overall behaviour being well summarized by the correlation coefficients shown in Table 4. First, and as for LMDz, the correlations evaluated using Phase 2 data stay robust when

East	Day Dof	CM AM	IFS	ECE ARTH	Ech am5	MI ROC	MPI M	MRI ESM	EM AC	LMD z	UMG A7gws	HadG EM2	WAC CM
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	Day Dof	CM AM	IFS	ECE ARTH	Ech am5	MI ROC	MPI M	MRI ESM	EM AC	LMD z	UMG A7gws	HadG EM2	WAC CM
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**	0.00	0.11	ns

TABLE 4 Correlation between observed and measured fluxes, strateole phases 1 and 2. 1% significant values according to 2-sided Pearson test are in bold, 5% are in italic, ’ns’ stands for non-significant. To evaluate the number of degree of freedom, we proceed as in Lott et al. (2023) and evaluate for each flight the time lag for which the auto correlations of the daily averaged fluxes fall below 0.1 and divide the number of days by that lag.

compared to correlations evaluated using phase 1, and whatever is the level of correlation ("medium", "low", or "non significant"). Second, is that 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, the HadGEM2 and WACCM also 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 balloons 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 for two schemes that launch waves quite near the tropopause (CMAM and HadGEM2). We have difficulties in interpreting this last result. It may mean that the approaches where some waves are launched from near the tropopause should not be disregarded, and that launching from a fixed altitude well in the troposphere fails in some cases. But if this is the case, the performance of LMDz is in contradiction since the launching level is in the mid troposphere, as many other schemes according to tables 3-2-1. Maybe its skill is because LMDz explicitly launch 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 HadGEM2 and WACCM scheme the waves are launched according to their absolute frequency. These are more speculations given here to emphasize the differences that are dynamically significant in our opinion, what is maybe more interesting to notice that there is room to improve GWs parameterizations to obtain better fits between predicted and measured fluxes in both directions of propagation, as illustrates the case of LMDz.

As stated in the introduction, more than predicting the right fluxes at the right time, it is often believed that parameterizations should better be validated against their statistical behaviour. A example is that observed gravity waves MFs are strongly intermittent, a statiscal 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 paper, Green et al. (2023) showed that this intermittent behaviour is well captured when the GWs 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 Fig. 7, such lognormal distributions accurately

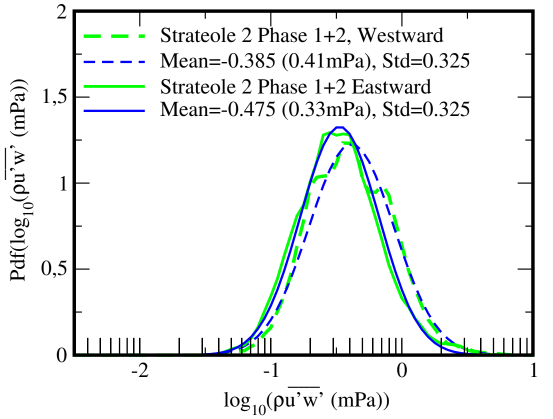

FIGURE 7 PDFs of daily values of Momentum flux distribution evaluated from Strateole Phases 1 and 2. The PDFs are calculated from histograms of 1291 MFs daily value within intervals of $\Delta\left(\log_{10}\rho\overline{u'w'}(\mathrm{mPa})\right) = 0.05$, thereafter smoothed by a 5 point non-recursive 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 the mean and standard deviations given in caption.

describe the Strateole-2 data, where 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 0.1mPa to 10mPa. Furthermore, the pdfs of the westward fluxes are seen to be shifted toward higher values compared to 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.

To analyse the QBOi schemes in this framework , Figure 8 presents PDFs of the distributions of the parameterized daily values of the 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 smuch 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 to 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 the MFs. These tails are not realistic which suggests that in these parameterizations miss a background of wave activity that exist 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 conclusions of Green et al. (2023) that differences in GW momentum fluxes between direction of propagations 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 (10mPa) than the other schemes, this tail is consistent with the fact that some balloons have very large fluxes on average (see Fig. 6).

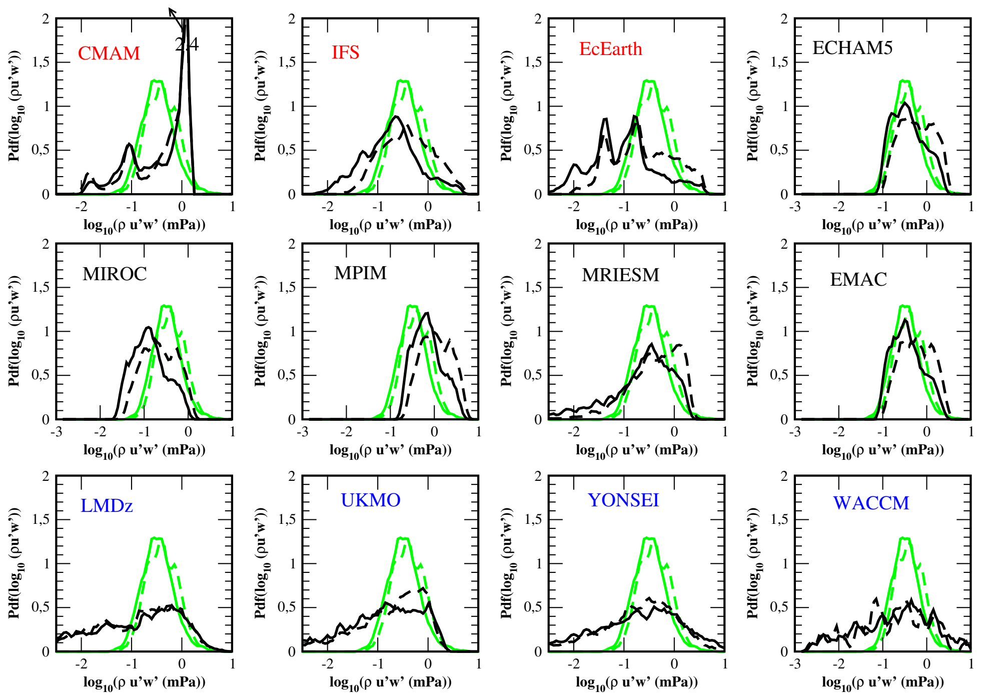

FIGURE 8 PDFs of daily values of Momentum flux distribution, same method as in Fig. 7. Measured values are in green, estimations using ERA5 data and the parameterizations are in black. Solid lines are for Eastward, dashed lines are for Westward.

4 | CONCLUSION

The main result of this paper is that state-of-the-art parameterizations of GWs reproduce reasonably well the momentum flux due to the high-frequency waves (periods between 15mn and 1hr) deduced from in situ measurements made onboard constant-level balloons. The parameterizations represent well the eastward and westward values of the momentum fluxes and in some cases their variations from day to day. Although the various schemes performed differently regarding the day-to-day correlations, our results show that improvement can be done in this regard. The fact that some schemes present "medium" correlations in the eastward direction, means that such a correlation level can be reached. In the westward direction, the day-to-day correlations are "low" at best and in 1 model,(the other models did not have a statistically significant correlation for the westward fluxes), we can only say that such a level can be reached.

Due to the low to medium level of the day-to-day correlations we found, we could ask ourselves if it is mandatory to improve GW schemes according to such a 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 Fig. 5) and sometime medium in the westward direction. Such a level of correlation is probably enough in the context of the QBO forcing, since the QBO evolving over time scales much longer than a month. Also, it is important to recall that the offline testbed we have used to test the different schemes has been initially designed to evaluate the LMDz scheme against the strateole 2 data. For this scheme we have taken great care over the years that the offline setup stay close from the online one. In other words, the offline setup used here is not necessarily optimal for the other parameterizations. One should therefore only conclude that significant daily correlation can be obtained offline, as illustrates here one scheme in both propagation directions. One can also conclude that it is easier to find significant correlations for eastward waves than for westward waves, as many schemes show. This is probably related to the phase of the QBO at the balloons altitudes it would be important to plane an other campaign in an other phase of the QBO.

Substantial differences are also found when we compare the pdfs of the parameterized momentum fluxes to the pdfs of the measured fluxes. The spectral schemes following the Hines Doppler Spread parameterization (HDS) behave the more realistically in this respect. The pdfs for the HDS schemes exhibit one isolated maxima 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 them the source amplitude is constant and they are supposed to represent a broad ensemble of waves, two factors that could make them much less intermitent than the multiwave schemes including sources explicitely. It happens that for these schemes the dynamical filtering is efficient enough to reproducing 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., 2023) needed in some models to represent well the final warmings in the southern hemisphere (de la Cámara et al., 2014) or the fluctuations of the QBO peridiodicity (Lott and Guez, 2013). Consistent with dynamical filtering, it is also not surprising that CMAM fails in capturing a log-normal distribution since it launches waves from quite near the balloon height.

The schemes that relate the GWs to convection also have broad momentum flux pdfs, much broader than the spectral schemes. So in this sense they can be viewed as being even more intermittent then the spectral schemes. Furthemore they are also characterized by long tails toward small values which seem unrealistic. For these schemes it therefore seems important to introduce a background in wave launching amplitude. This problem could also be in part corrected out by introducing lateral propagation (Amemiya and Sato, 2016), 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 have not much data though, but could use the Loon data post-processed in a comparable way as Strateole 2 by Green et al. (2023), which would permit coverage of much wider regions. We could also complement these observations with the convection permitting global which outcomes look promising (Stephan et al., 2017; Köhler et al., 2023; Sun et al., 2023). We should also test if improving the schemes parameters to improve the fit with observations improves or does not degrade the models 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 (Lott et al., 2014; Holt et al., 2022). Also, we could also 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. Unfortunately, our results are not too promising in this regard: a common believe is that such an error could well be reduced by launching waves from near the tropopause, but the parameterizations which do so here are not much realistic when it comes to predict MFs variabilities (over days or months).

5 | OPEN RESEARCH

All data and routines needed to run the parameterizations in offline mode and to compare the results with the strateole 2 flghts are available on a dedicated web site, see details here:

`https://web.lmd.jussieu.fr/~flott/DATA/Documentation.pdf`

references

Achatz, U., Kim, Y.-H. and 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. URL: https://doi.org/10.1063/5.0165180.

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

Alexander, M. J. and Dunkerton, T. J. (1999) A Spectral Parameterization of Mean-Flow Forcing due to Breaking Gravity Waves. *J. Atmos. Sci.*, **56**, 4167–4182.

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., Sato, K., Eckermann, S., Ern, M., Hertzog, A., Kawatani, Y., Pulido, M., Shaw, T. A., Sigmond, M., Vincent, R. and Watanabe, S. (2010) Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Q. J. R. Meteorol. Soc.*, **136**, 1103–1124.

Alexander, M. J., Liu, C. C., Bacmeister, J., Bramberger, M., Hertzog, A. and 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. E2020JD033954 2020JD033954.

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

Andrews, F. G., Holton, J. and Leovy, C. (1987) *Middle Atmosphere Dynamics*. Academic Press.

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

Anstey, J. A., Scinocca, J. F. and 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.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K. and Takahashi, M. (2001) The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179–229.

Beres, J. H., Garcia, R. R., Boville, B. A. and 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**.

Bushell, A. C., Butchart, N., Derbyshire, S. H., Jackson, D. R., Shutts, G. J., Vosper, S. B. and Webster, S. (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., Kawatani, Y., Kim, Y.-H., Lott, F., Scinocca, J., Stockdale, T. N., Andrews, M., Bellprat, O., Braesicke, P., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Dobrynin, M., Garcia, R. R., Garcia-Serrano, J., Gray, L. J., Holt, L., Kerzenmacher, T., Naoe, H., Pohlmann, H., Richter, J. H., Scaife, A. A., Schenzinger, V., Serva, F., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (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. and 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. and 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. and Madsen, M. S. (2016) Do strong warm enso events control the phase of the stratospheric qbo? *Geophysical Research Letters*, **43**, 10,489–10,495.

Corcos, M., Hertzog, A., Plougonven, R. and 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., Watson, P. A. G., Weisheimer, A. and Palmer, T. N. (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. URL: `https://gmd.copernicus.org/articles/10/1383/2017/`.

de la Cámara, A. and Lott, F. (2015) A parameterization of gravity waves emitted by fronts and jets. *Geophys. Res. Lett.*, **42**, 2071–2078.

de la Cámara, A., Lott, F. and Hertzog, A. (2014) Intermittency in a stochastic parameterization of nonorographic gravity waves. *J. Geophys. Res.: Atmospheres*, **119**, 11905–11919.

de la Cámara, A., Lott, F., Jewtoukoff, V., Plougonven, R. and Hertzog, A. (2016) On the gravity wave forcing during the southern stratospheric final warming in LMDZ. *J. Atmos. Sci.*, **73**, 3213–3226.

Eckermann, S. D. (2011) Explicitly Stochastic Parameterization of Nonorographic Gravity Wave Drag. *J. Atmos. Sci.*, **68**, 1749–1765.

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

Fovell, R., Durran, D. and 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. and Uppala, S. (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., Manzini, E., Preusse, P., Sato, K., Scaife, A. A. and Zhou, T. (2013) A comparison between gravity wave momentum fluxes in observations and climate models. *J. Atmos. Sci.*, **26**.

Green, B., Sheshadri, A., Alexander, M., Bramberger, M. and Lott, F. (2023) Gravity wave momentum fluxes estimated from project loon balloon data. *Journal of Geophysical Research: Atmospheres*, **Submitted**.

Haase, J. S., Alexander, M. J., Hertzog, A., Kalnajs, L. E., Deshler, T., Davis, S. M., Plougonven, R., Cocquerez, P. and Venel, S. (2018) Around the world in 84 days [Dataset]. *Eos*, **99**.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (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. URL: `https://www.sciencedirect.com/science/article/pii/S1752929807800478`.

Hertzog, A., Alexander, M. J. and Plougonven, R. (2012) On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere. *Journal of the Atmospheric Sciences*, 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.

— (1997) Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. part 2: Broad and quasi monochromatic spectra, and implementation. *J. Atmos. Solar Terr. Phys.*, **59**, 387–400.

Holt, L. A., Lott, F., Garcia, R. R., Kiladis, G. N., Cheng, Y.-M., Anstey, J. A., Braesicke, P., Bushell, A. C., Butchart, N., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Kawatani, Y., Kerzenmacher, T., Kim, Y.-H., McLandress, C., Naoe, H., Osprey, S., Richter, J. H., Scaife, A. A., Scinocca, J., Serva, F., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (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.

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., Jam, A., Musat, I., Idelkadi, A., Fairhead, L., Foujols, M.-A., Mellul, L., Traore, A.-K., Dufresne, J.-L., Boucher, O., Lefebvre, M.-P., Millour, E., Vignon, E., Jouhaud, J., Diallo, F. B., Lott, F., Gastineau, G., Caubel, A., Meurdesoif, Y. and Ghattas, J. (2020) LMDZ6A: The atmospheric component of the ipsl climate model with improved and better tuned physics [Software]. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001892.

Jewtoukoff, V., Hertzog, A., Plougonven, R., de la Cámara, A. and Lott, F. (2015) Comparison of gravity waves in the southern hemisphere derived from balloon observations and the ecmwf analyses. *J. Atmos. Sci.*, **72**.

Jewtoukoff, V., Plougonven, R. and 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., Baumgaertner, A., Gromov, S. and Kern, B. (2010) Development cycle 2 of the modular earth submodel system (messy2). *Geoscientific Model Development*, **3**, 717–752.

Kang, M.-J., Chun, H.-Y. and 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.

Köhler, L., Green, B. and 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**.

Lane, T. P. and Moncrieff, M. W. (2008) Stratospheric gravity waves generated by multiscale tropical convection. *J. Atmos. Sci.*, **65**, 2598–2614.

Lindzen, R. S. (1981) Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.*, **86**, 9707–9714.

Liu, C., Alexander, J., Richter, J. and 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. E2021JD035785 2021JD035785.

Lott, F., Denvil, S., Butchart, N., Cagnazzo, C., Giorgetta, M. A., Hardiman, S. C., Manzini, E., Krismer, T., Duvel, J.-P., Maury, P., Scinocca, J. F., Watanabe, S. and Yukimoto, S. (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. and Guez, L. (2013) A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *J. Geophys. Res.*, **118**, 8897–8909.

Lott, F., Guez, L. and Maury, P. (2012) A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere. *Geophys. Res. Lett.*, **39**, L06807.

Lott, F., Rani, R., Podglajen, A., Codron, F., Guez, L., Hertzog, A. and Plougonven, R. (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. and 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. and 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. and 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. and Stainforth, D. A. (2004) Equatorial stratospheric response to variations in deterministic and stochastic gravity wave parameterizations. *Journal of Geophysical Research: Atmospheres*, **109**.

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

Pohlmann, H., Müller, W. A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F. S. E., Kadow, C., Illing, S. and Marotzke, J. (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., Karbou, F., Peuch, V., El Amraoui, L., Puech, D., Genthon, C., Picard, G., Town, M., Hertzog, A., Vial, F., Cocquerez, P., Cohn, S. A., Hock, T., Fox, J., Cole, H., Parsons, D., Powers, J., Romberg, K., Vanandel, J., Deshler, T., Mercer, J., Haase, J. S., Avallone, L., Kalnajs, L., Mechoso, C. R., Tangborn, A., Pellegrini, A., Frenot, y., Thépaut, J.-N., Mcnally, A., Balsamo, G. and Steinle, P. (2010) The Concordiasi Project in Antarctica. *Bulletin of the American Meteorological Society*, **91**, 69–86. URL: `https://hal-insu.archives-ouvertes.fr/insu-00562459`.

Richter, J. H., Sassi, F. and 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.

— (2010b) Toward a Physically Based Gravity Wave Source Parameterization in a General Circulation Model. *J. Atmos. Sci.*, **67**, 136–156.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U. and Schulzweida, U. (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. and 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. and 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.

Song, I.-S. and Chun, H.-Y. (2005) Momentum flux spectrum of convectively forced internal gravity waves and its application to gravity wave drag parameterization. part i: Theory. *J. Atmos. Sci.*, **62**, 107–124.

Stephan, C., Alexander, M. J. and Richter, J. H. (2017) Characteristics of gravity waves from convection and implications for their parameterization in global circulation models. *J. Atmos. Sci.*, **73**, 2729–2742.

Sun, Y. Q., Hassanzadeh, P., Alexander, M. J. and 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. and McIntyre, M. E. (1996) On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *J. Atmos. Sci.*, **53**, 3213–3235.

— (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., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S. and Kawamiya, M. (2011) Miroc-esm 2010: model description and basic results of cmip5-20c3m experiments. *Geoscientific Model Development*, **4**, 845–872.