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 Research Council, Rome, Italy
[15]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
[16]Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany
[17]Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japa

Key Points:

- The 3 standard non-orographic gravity waves (GWs) parameterizations tuned to produce a realistic tropical quasi-biennial oscillation in 12 global climate models are used to predict in-situ balloon observations.
- Parameterized GWs needed in large-scale models have realistic amplitudes in the tropical lower stratosphere.
- Balloon averaged and daily values of GWs momentum fluxes can correlate with observations when the parameterized GWs are coming from the lower and middle tropsphere.
- The probability density distributions can also be realistically reproduced, but problem arises for parameterizations that try to relate gravity waves to their convective sources.

Version 2, date: January 23, 2024

Corresponding author: Francois Lott, `flott@lmd.ens.fr`

Abstract

Gravity Waves (GWs) parameterizations from 12 General Circulation Models (GCMs) participating to the Quasi-Biennial Oscillation initiative (QBOi) are directly compared to Strateole-2 balloon observations made in the lower tropical stratosphere from November 2019 to February 2020 (phase 1) and from October 2021 and January 2022 (phase 2). The parameterizations used span the 3 standard techniques used in GCMs to represent subgrid scale non-orographic GWs, the two globally spectral techniques developed by Warner and McIntyre (1999) and Hines (1997) respectively and 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 instantaneous meteorological conditions found underneath the balloons. In general, the amplitudes are in fair agreement between measurements of the momentum fluxes due to waves with periods less than 1 hr and the parameterizations. The correlation of the daily values between the observations and the results of the parameterization can be around 0.4, which is statistically elevated considering that we analyse around 1200 days of data and sometime good considering that the parameterizations have not been tuned: the schemes used are just the standard ones that help producing a Quasi-Biennial Oscillation (QBO) in the corresponding model. These correlations nevertheless vary considerably between schemes and depend little on their formulation (globally spectral versus multiwaves for instance). We therefore attribute it to dynamical filtering all schemes taking good care of it, whereas only few relate the gravity waves to their sources. Except for two parameterizations, 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.

Plain Language Summary

In most large-scale atmospheric models, gravity wave parameterizations are based on well understood but simplified theories and parameters which are keyed to reduce systematic errors on the planetary scale winds. In the equatorial regions, the most challenging errors concern the Quasi Biennial Oscillation. Although it has never been verified directly, it is expected that the parameterizations tuned this way should transport a realistic amount of momentum flux in both the eastward and westward directions when compared to direct observations. Here we show that it is the case, to a certain extent, using constant-level balloon observations at 20 km altitude. The method consists in comparing directly, each day and at the location of the balloon the measured momentum fluxes and the estimations from the gravity wave parameterizations used in the global models that participate to the Quasi-Biennal Oscillation intiative and when using observed values of the large-scale meteorological conditions of wind, temperature, precipitation, and diabatic heating.

1 Introduction

It is well known that the large scale circulation in the middle atmosphere is in good part driven by gravity waves (GWs) that propagate in the stratosphere (Andrews et al., 1987). These waves carry horizontal momentum vertically and interact with the large scale flow when they break. The horizontal scale of these waves can be quite short, much shorter than the horizontal scale of conventional atmospheric General Circulation Models (GCMs) so they need to be parameterized (Alexander & Dunkerton, 1999). In the tropics, the convective GWs are believed to dominate largely (Fovell et al., 1992; Alexander et al., 2000; Lane & Moncrieff, 2008), they 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, the parameterization of convective GWs is necessary for most GCMs to explicitly realize the QBO.

Although gravity wave parameterizations are now used in many models with success including in the tropics (Scinocca, 2003; Song & Chun, 2005; Beres et al., 2005; Orr et al., 2010; Lott & 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. There exist observations of GWs using global satellite observations (Geller et al., 2013) but the GWs identified this way still have quite large horizontal scales, and some important quantities like the Momentum Fluxes (MFs) are often deduced indirectly, for instance from temperature measurements using polarization relations (Alexander et al., 2010; Ern et al., 2014). For these two reasons, in situ observations are essential, and the most precise ones 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 in 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 than when the GW field is more uniform. This property, 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, for instance on the timing of the final warming in the Southern Hemisphere polar stratosphere (de la Cámara et al., 2016), or on 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 predicted offline using reanalysis data (Kang et al., 2017; Alexander et al., 2021).

However, the evaluation of parameterizations using balloon observations done in the past were often quite indirect , and concern more their statistical behaviours (Jewtoukoff et al., 2015; Kang et al., 2017; Alexander et al., 2021) than there capacity to directly predict instantaneous values of momentum fluxes. Maybe a good reason to believe so is that parameterizations are based on simplified quasi-linear wave theory, they assume spectral distributions that are loosely constrained, and they ignore lateral propagation almost entirely (some attempt to include it can be found in Amemiya and Sato (2016)). Nevertheless, some factors could mitigate these weaknesses. One is that in most parameterizations the wave amplitude is systematically limited by a breaking criterion that encapsulates nonlinear effects. An other is that some parameterizations explicitly relate launched waves to sources, and there is constant effort to improve the realism of the convective ones (Liu et al., 2022). Also, observations systematically suggest that dynamical filtering by the large scale wind is extremely strong for upward propagating GWs (Plougonven et al., 2017), and this central property is represented in most GW parameterizations. For all these reasons, it may well be that GW parameterizations keyed to the large scale conditions found at a given place and time gives MFs that can be directly compared to the MFs measured by a balloon at the same place.

Based on the relative success of the offline calculations done in the past 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 result of interest. The first is that a state of the art convective gravity wave drag scheme due to Lott and Guez (2013) predicts momentum fluxes in the low equatorial stratosphere which amplitude can be directly compared with those measured during phase 1 of the Strateole-2 balloon campaign. It gives a direct in-situ observational confirmation that the theories and modelling of the QBOs developed over the last 50 years were in good part correct about the significance of the GWs to the QBO forcing. Also interesting, the comparison showed a good level of correlation between the day to day variability in momentum fluxes between measured and observed fluxes, a correlation that is much better for waves carrying momentum fluxes in the eastward direction than in the westward direction. It was suggested that such a good correlation was due to the fact that the Lott and Guez (2013)'s scheme analysed relate the gravity waves to their convective sources (not all schemes do) and that the GWs experience significant dynamical filtering in the middle troposphere and lower stratosphere. Lott et al. (2023) nevertheless revealed that a scheme that relates gravity waves to convection exclusively somehow failed in predicting the right statistical behaviour of the momentum fluxes, the probability density function of the momentum fluxes amplitude showing long tails for low values of the MFs, suggesting missing processes like lateral propagation or the presence of a background of waves which origin remains a challenge to predict.

The purpose of this paper is to continue such a direct comparison including more recent Strateole 2 observations and near all the gravity wave parameterization schemes used by the modelling groups participating to the Quasi-Biennial Oscillation initiative (QBOi, Butchart et al., 2018). We will follow for that 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. For each of the flights and each time, we have identified the grid point in the ERA5 reanalysis (Hersbach et al., 2020) that is the nearest and used the vertical profiles of wind and temperature as well as the surface value of precipitations to emulate the parameterization of GWs used in the global models that participated to QBOi. We also extract from the analysis and short range forecast, diabatic heatings and the cloud base and top altitudes needed in some schemes to predict gravity waves.

2 Data and method

2.1 Parameterizations of non orographic gravity wave schemes

The parameterization schemes used to predict non-orographic gravity waves belongs to two well separated families, dating back from the 1980's when it becomes evident that a good simulation of the middle atmosphere by global atmospheric models could not be done without taking them into account. The first family roots in the formulation by Lindzen (1981), where the gravity wave field is represented by gravity waves that are monochromatic in the horizontal and time. It was 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 it roots in linear theories where sources like convection and/or fronts can be introduced using closed form theories (Beres et al., 2005; Song & Chun, 2005; Richter et al., 2010a; Lott & Guez, 2013; de la Cámara & Lott, 2015). In the following we will refer to such schemes as "multiwave", they are expensive because they request a large amount of harmonics to represent well a realistic wave field, but this limit can easily be circum-

	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	1000hPa	$\sqrt{\text{Precip}}$	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 & 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.

vented by using stochastic approaches (Eckermann, 2011; Lott et al., 2012). As an alternative, but also to better represent breaking, globally spectral schemes have been developed and tested 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 wavenumber, 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, McFarlane, & McLandress, 1997), and will be referred to as "HDS" for "Hines Doppler Spread" in the following. As an alternative, the theory in Warner and McIntyre (1996) imposes gravity wave saturation according to an empirical spectra but treat vertical changes in the spectra following GWs propagation invariant character. The theory has been simplified and/or optimized to permit implementation, for instance in the UKMO model (Warner & McIntyre, 1999; Scaife et al., 2002) and in the CMAM model (Scinocca, 2003) respectively, and will be refered to has "WMI" for "Warner and McIntyre" in the following. 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 precipitations in Bushell et al. (2015).

In the present paper, we are going to compare the GWs schemes used in 12 of the models that participate to QBOi, all belonging to one of the three type of schemes described above (WMI, HDS, and Multiwave). As all the multiwave schemes used relate GWs to their convective sources and as only one of the spectral scheme is doing so, the UMGA7gws WMI scheme in Bushell et al. (2015), the former will be discussed with the source-related multiwave schemes.

Among the 12 models, three use the Scinocca (2003)'s version of WMI, CMAM, IFS and ECEarth, their version for QBOi are further detailed in Anstey et al. (2016),Orr

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

et al. (2010), and Davini et al. (2017) respectively. They 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).

Still among the 12 models, 5 uses the HDS parameterization presented 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)) respectively. Between them change 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 that eventually changes, a minimum value for the the cutoff vertical wavenumber m_{min}, and two parameters that control smoothing in the vertical of the GWs root mean square variance and cut-off vertical wavenumber, the coefficient C_{smo} and the number of time the smoothing is applied N_{smo}. Importantly nevertheless, in ECham5 the choice has been made to chose the variability 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 links GWs to sources (convection or precipitation), 3 are multiwaves and have been developed independently one from the others: LMDz, HadGEM2, and WACCM, their version for QBOi are described in Lott and Guez (2013), Song and Chun (2005), and Richter et al. (2010b) and 1 uses the ultra simple version of the WMI schemes presented in Bushell et al. (2015) rather than the (Scinocca, 2003)'s version. Note nevertheless that for both HadGEM2 and WACCM, we do not use the exact version used in QBOi models but rather offline versions developed by Kang et al. (2017) and Alexander et al. (2021) respectively, and which were adapted to interpret observations. The differences between the 3 multiwave schemes are numerous it is impossible to detail them, the reader is referred to the corresponding papers, but some salient differences are in the source term, the launching levels and the intrinsic phase speed of the launched waves. More specifically, in LMDz is made the choice to relate the launched MF 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). Still in LMDz, the waves are launched from the mid troposphere, whereas they are launched from the surface in the UMGA7gws model. In the HadGEM2's scheme (Song & Chun, 2005; Choi & 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 launching level can vary between $2km$ and 15km typically and the depth of the heating between 1km and $15km$. We will take the same parameters to run the WACCM scheme, using the version in Alexander et al. (2021), and despite that in this paper the WACCM scheme was adapted and partly re-written to use direct satellite observations of convective heating. Finally, an important difference is that LMDz span harmonics which intrinsic phase speeds typically range between $-30\text{m/s} < C_{abs} < 30\text{m/s}$, whereas in both UMGA7gws and WACCM the choice is made to have absolute phase speeds in the range $-100\text{m/s} < C_{abs} < 100\text{m/s}$.

2.2 Offline parameterization runs

To activate the schemes in offline mode we will use ERA-5 hourly data of precipitation and 3-hourly data of winds, surface pressure, temperature, cloud liquid and ice water content at $1^o \times 1^o$ horizontal grid to mimic a large scale climate model resolution. Winds, surface pressure, temperature, and water contents are then linearly interpolated on 1hr time step to be synchronised with precipitation. In the vertical we use data at 67 model levels, taking one every two ERA5 levels again to mimic large scale models vertical resolution but also to speed up calculations. To estimate convective heating rates vertical profiles, we follow Fueglistaler et al. (2009) and evaluate diabatic heating using ERA5 hourly data from short range forecast and as a residual between the parameterized temperature tendency and the radiative heatings (longwave plus shortwave). When needed, we also evaluate the cloud bottom and top altitudes using the cloud water content (liquid+ice) given in ERA5.

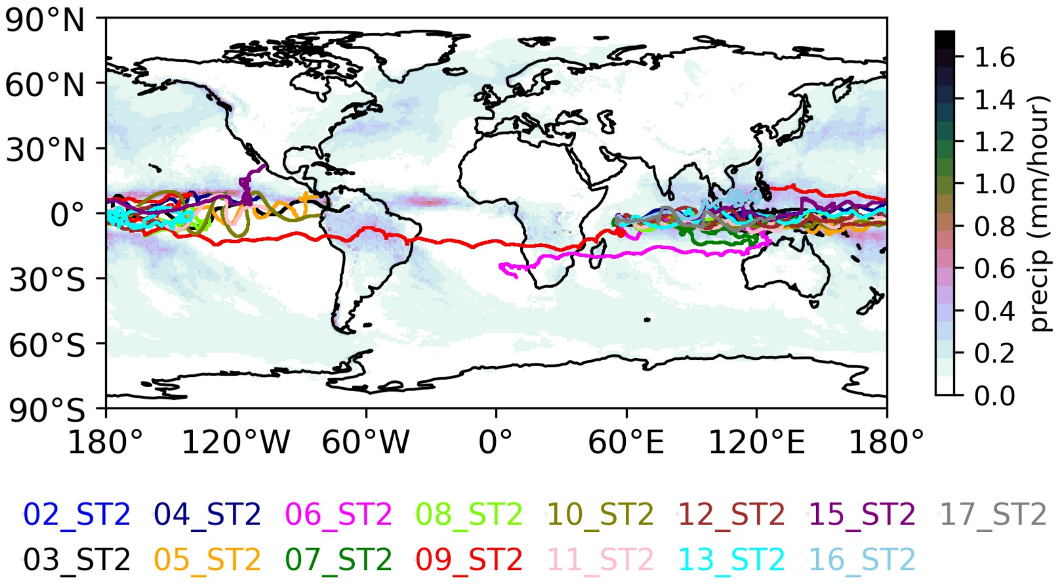


Figure 1. Strateole 2, Phase 2 balloon trajectories taking place between October 2021 and January 2022. Shading presents the precipitation field from ERA5 averaged over the period.

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 lower tropical stratosphere between November 2019 and February 2020 and from the 15 balloons that flew 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 the averaged precipitation (the same Figure but for phase 1 is in Lott et al. (2023)). In the MFs calculated from observations Corcos et al. (2021) distinguish the waves with short periods (1hr-15mn) from the waves with period up to one day (1d-15mn). They also distinguish the

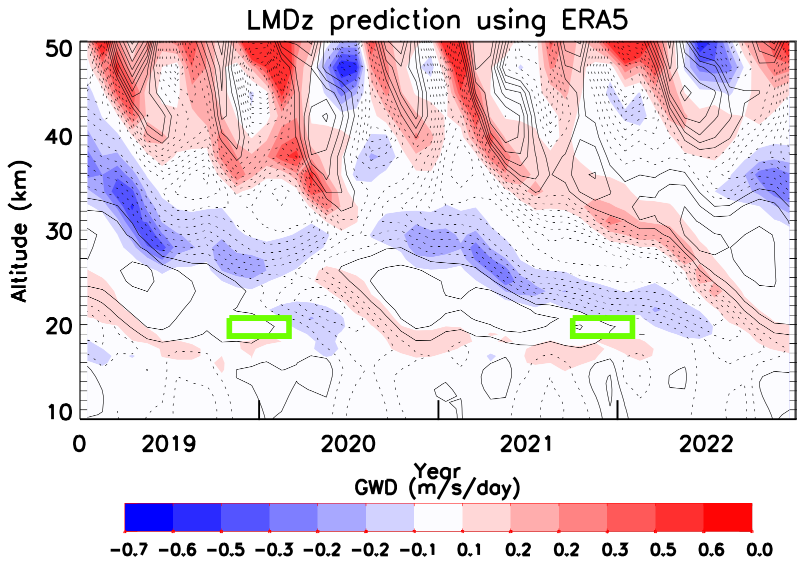


Figure 2. Time vertical sections of the zonal mean zonal wind (CI=10m/s, negative values are dashed and non-orographic gravity wave tendency averaged over the Equatorial band ($-6^oS \quad - \quad +6^oN$). Input data are from ERA5 reanalysis and GWs prediction from the LMDz scheme. The 2 green boxes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

eastward waves giving positive MF in the zonal direction from the westward waves giving negative MF. To characterize the QBO condition during the balloon flights, Fig. ?? shows time altitude sections of the equatorial zonal winds and GWD predicted by the scheme globally and in offline mode usin LMDz scheme between 2018-2023. In it we see that the gravity wave drag is negative (positive) where the zonal mean zonal wind vertical shear is negative (positive) consistent with the fact that it contributes to the descent of the QBO. We also note that the amplitudes vary between ±0.5m/s/day, a range characteristic of the parameterized GW tendency used in GCMs that produce a quasi-biennial oscillation (Butchart et al., 2018). The figure also indicates with a green rectangle the region and period during which the balloons operated, typically during the end of easterly QBO phase for both phase 1 and 2. As we shall see this yield quite comparable results during the two phase, and despite the fact that during phase 1 and above flight altitude the 2nd documented QBO disruption started (Anstey et al., 2021).

In the following we will compare the momentum fluxes derived from the balloon data, emphasize the intrinsic frequencies that the scheme represents (the intrinsic periods below $1hr$) and consider the ERA5 data at the points that are the nearest from the balloon. The prediction is then made every hour and averaged over the day, partly because it is the time scale needed for the some schemes to sample realistically a GW field, and also because it takes around a 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 shows time series of daily values of momentum fluxes predicted 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), and where was also shown the time

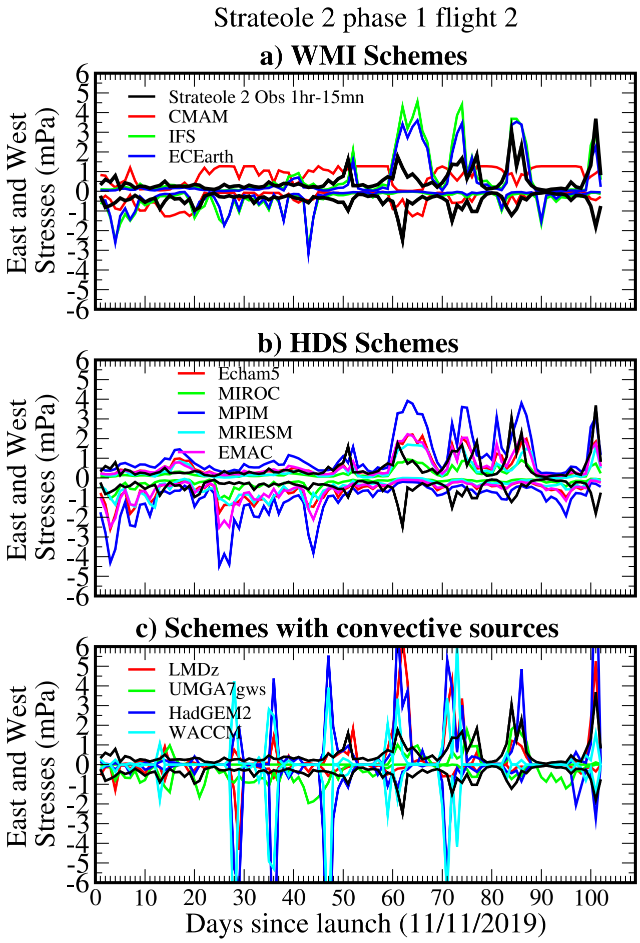


Figure 3. Comparison between daily averaged values of the eastward and westward MFs measured by the balloons during Strateole 2 phase 1 Flight 2 and estimated by the GW schemes at the balloon location and altitude. Colored curves are for the GW schemes predictions using ERA5 and from different models, 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.

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 GWs fluxes to their sources. In all panels the black curves are for the daily observations. For clarity we present results for the eastward and westward MFs only. Overall ones sees that the schemes predict momentum flux values that somehow compare with the observed one, at least in term of amplitude. There are nevertheless significant differences in behaviour. For instance, the IFS's schemes present substantial peaks in eastward flux during the second half of the flight, which is a period during which the zonal wind at flight altitude becomes westward potentially favoring eastward waves, a process we refer to as dynamical filtering in Lott et al. (2023) (see Figure 3 and Eq. 3 there and the following discussion). Note that in this paper, we showed that the 3 peaks in measured fluxes around days 60, 75, and 83 also correspond to dates when there are precipitations near the balloon location. These correspondences made us believe that the relation with convective sources is essential, we see here that dynamical filtering alone may well be the main cause. Although having smaller amplitudes, the Fig-

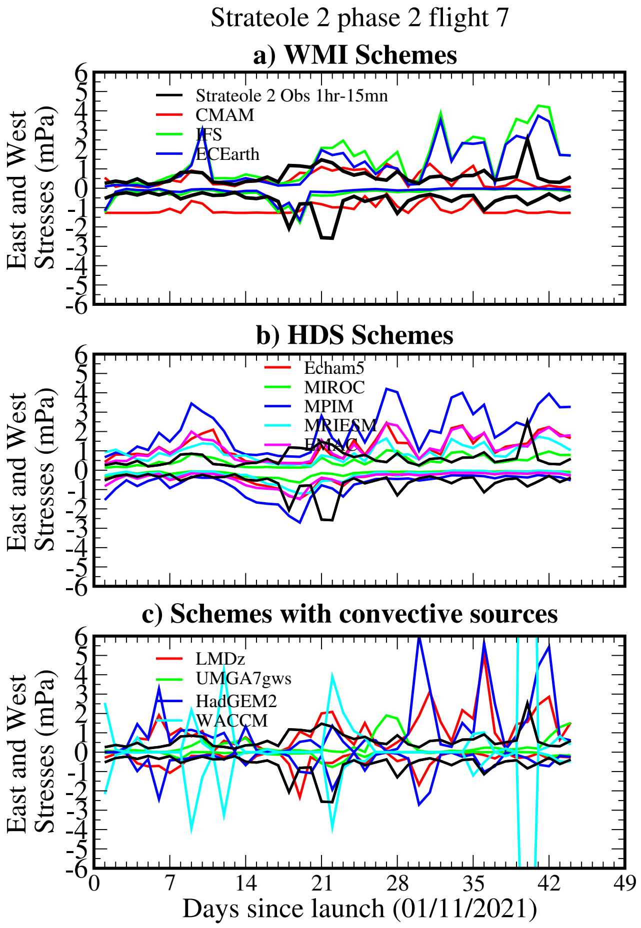


Figure 4. a) Same as Fig 2 but for Strateole 2 Phase 2 Flight 7.

ure show that in EC-Earth, the momentum fluxes behave almost as in IFS. The results from CMAM are quite different nevertheless. In this model it was chosen to place the launching altitude near the tropopause. As a consequence the daily series fluctuate less and present long lasting "plateaus". Clearly in this model, the distance between the launching level (100hPa see Table 1) and the balloon altitude is too small for dynamical filtering to be efficient. The second panel 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 behaviours of the source related schemes (multiwave for LMDz an HadGEM2, WMI for UMGA7gws) in the last panel are more contrasted. As expected, there are long periods during which the schemes predicted small and null momentum fluxes fluxes, interrupted by short lasting peaks with values sometime going beyond ±5mPa, values that were never reached by any of the globally spectral schemes in Panels. 3a) and 3b). In contrast with LMDz and HadGEM2, the UMGA7gws scheme present smaller amplitude and broader peaks, we attribute this to that it relates the launched flux to $\sqrt{P_r}$ rather than P_r^2 in LMDz, or the square of heating in HadGEM2's and WACCM.

An other example of timeseries is provided in Fig. 4, which corresponds to a flight during the second phase of strateole 2. Beyond the fact that the flight is shorter than in Fig. 3, a difference in duration that characterize most of the flights during phase 2 compared to phase 1,the overall behaviours stay about the same, with the spectral schemes presenting fluctuations with broader peaks, except maybe CMAM, again because the launching altitude is quite high and dynamical filtering not yet efficient at balloon flight altitude. The last panel also shows that UMGA7gws present long periods with almost no fluxes, in it, the fact that the launching height is near the surface produces much more critical level situations during the propagation through the troposphere. Finally, in the version of WACCM we use, there is extreme outlayers at day 40, with values exceeding ±10mPa, we only found few of them over the entire campaign, and only in WACCM. They translate that WACCM sometimes and rarely predicts extreme values in MFs, extreme values that significantly contribute to the averaged MFs.

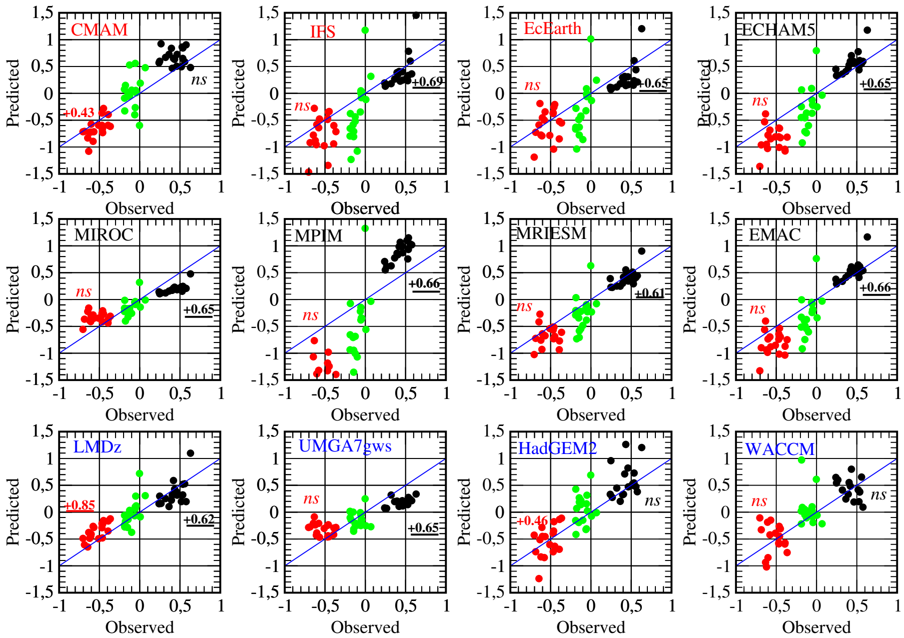


Figure 5. Scatter plot of the momentum fluxes measured by the balloon versus parameterized using different models. Only considered here the 18 balloon flights that last more than a month (East: black; West: red; Cumulated: green). Also shown are the correlations between observations and predictions, 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.

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 it we see that the predicted values align quite well with the observed one, some schemes having tendency to slightly underestimate the fluxes (MIROC, LMDz), other to overestimate them (CMAM, HadGEM2), with the tendency to overestimate being in general more pronounced for the westward fluxes. The numbers in each panels also show the correlation between the 18 values averaged over each flights, showing that the correlations become strong in many models, at least in the eastward direction. Interestingly some models also have significant medium to high correlations in the westward direction (CMAM, LMDz, HadGEM2).

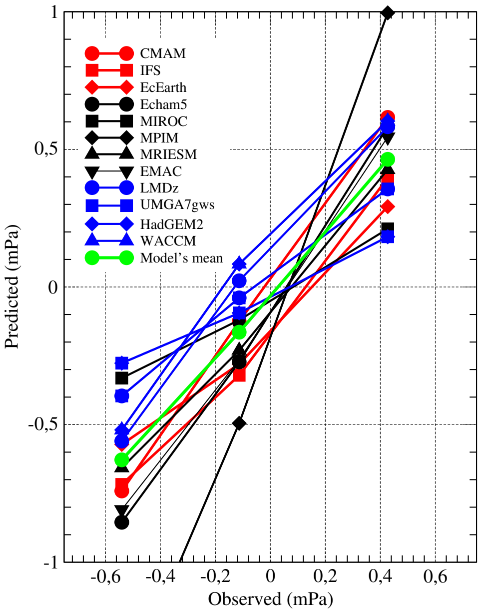


Figure 6. East, West and cumulated zonal momentum fluxes averaged over the Strateole 2 phase 1 and 2 period and according to participating models.

The Figure 6 group the models averaging the eastward and westward fluxes over all the balloon flights, confirming again that the parameterizations used fall around the observed values. There is variabilities between the models, but there is no systematic tendency among the modellers to overstate or understate the MFs flux amplitude. This is summarized by the green curve which represents the average over models and over balloon flights. The average amplitude of the eastward flux is very near that observed (a 10% overestimation between 0.45mPa in parameterizations against 0.40mPa observed), whereas the westward flux are overestimated by the models by less than 20% (-0.65mPa parameterized against -0.55mPa observed). This 10%-20% errors explain the quite large relative error (50%) in the cumulated flux but for it the large relative error is in good due to the fact that large positive and negative fluxes opposed each other.

The daily series in Figs 3 and 4 also suggest that observations and offline estimations sometimes evolve similarly day after day, a reason could be that both measured and parameterized MFs are sensitive to dynamical filtering, some schemes also taking into account sources. In the two examples given here, it is quite apparent in the first (Figure 3) and for instance for the peaks in the eastward direction as already discussed. Correspondences are less obvious to visualize in the second case (Figure 4) where the evolution of the measured MFs present less variations than the predicted MFs. In Lott et al. (2023) these daily variabilities were analysed flights by flights, in some flights the series correlating well whereas in others they do not. The contrast between flights made that in the end the correlations where significant but "medium" in the eastward direction $C \approx 0.5$ and "low" in the westward direction $C \approx 0.3$. Here and in the following, we referred to "medium" positive correlations with $0.3 < C < 0.5$ and small correlations when $0.1 < C < 0.3$. As such a result was obtain from the LMDz parameterization during Strateole 2 phase 1 the coefficients are given again in the 9th column of Table 4. In it are also given the same coefficients but for Phase 2 and measured over

East	Day	CM	IFS	ECE	Ech	MI	MPI	MRI	EM	LMD	UMG	HadG	WAC
	Dof	AM		ARTH	am5	ROC	M	ESM	AC	z	A7gws	EM2	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**	0.2
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**	0.13
West	Day	CM	IFS	ECE	Ech	MI	MPI	MRI	EM	LMD	UMG	HadG	WAC
	Dof	AM		ARTH	am5	ROC	M	ESM	AC	z	A7gws	EM2	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 ds 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.

the 2 phases. Consistent with the results found for phase 1, we found during phase 2 medium correlation in the Eastward phase ($C = 0.4$) and in the westward phase ($C = 0.40$), the values evaluated over the two phases being medium and small, $C = 0.46$ and $C = 0.34$, repectively. Here and for completeness, note that as in (Lott et al., 2023), and to test the significance, we measure the number of Degrees of Freedom (DoF) present in each dataset, and calculate for that the decorrelation time scale, which we take as the lag in day beyond which the lag-autocorrelation of the series falls below 0.2. As this time-lag varies from one series to the other, we give explicitly in column 5, the number of DoF, which is the duration of the flight divided by the decorrelation time scale. Note that for their decorrelation time, we consider for simplicity that evaluated with 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 schemes used in the other models, the result are contrasted but quite in agreement. A lot a variations between flights (not shown) the overall behaviour being well summarized in the global correlation coefficients shown in Table 4. First, and as for LMDz, the correlations evaluated using Phase 2 data stay robust when 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 launching level is quite. In CMAM it is always 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 them and in case of deep convections waves are launched from quite high levels in the troposphere (not shown) suggesting that in them as well and for waves with strong eastward flux, there is not enough space between launching levels and balloon flight for dynamical filtering to be efficient. The results in the westward direction are more intriguing, the correlations are always small except for 1 scheme (LMDz) and some but "low" correlations are found for two schemes that launch waves quite near the tropopause (CMAM and HadGEM2). We have difficulties in interpreting this last result, it may be tells 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 are somehow in contradiction, in it the launching level is in the mid troposphere, as many other schemes according to tables 3-2-1. Maybe its skill come from the fact that LMDz explicitely launch waves according to their intrinsic frequency, a choice that directly affect dynamical filtering, whereas in the globally spectral schemes the dynamical filtering is more indirect and while 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 LMDz6.

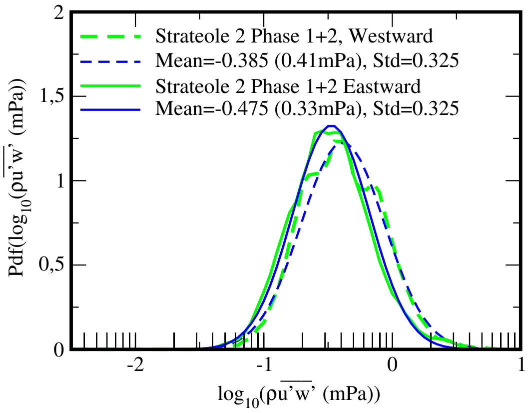


Figure 7. PDFs of daily values of Momentum flux distribution evaluated from Strateole Phases 1 and 2. The PDFs are calculated from histograms of 1291 MFs daily value within intervals of $\Delta\left(\log_{10}\rho\overline{u'w'}(\text{mPa})\right) = 0.05$, thereafter smoothed by a 5 point non-recursive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

As said 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 global statistical behaviour. A reason is that observed gravity waves show a strong level of intermittency such an intermittency impacting the the effect of the waves on the large scale flow and climate in the middle atmosphere. In a recent Paper, Green et al. (2023) showed that this intermittent behaviour is well captured when the GWs MFs have pdfs following a log-normal distribution. These authors even concluded that in all directions of propagation, momentum fluxes characteristics could be summarized in terms of the mean and variance of log normal distributions. As shows Fig. 7, such lognormal distributions also describe well the Strateole-2 data. In it, one sees that the balloons measure fluxes with amplitude between 0.1mPa and 10mPa, the pdf of the westward fluxes being shifted toward higher values compared to that for eastward flues the shapes being little changed. The Figure also shows that the shifts in pdf between eastward and westward fluxes are also well described by shifts in mean and variance of log-normal distributions.

Next, and to analyse the QBOi schemes in this framework the Figure 8 presents PDFs of the distributions of the predicted daily values of the momentum fluxes. In it we notice that in the WMI schemes (model names in red) te pdfs are quite broader than the observed pdfs, and often far from log-normal. CMAM and EC-earth for instance present peaks in PDFs not located in the middle of the distribution. Quite remarkably, the HDS

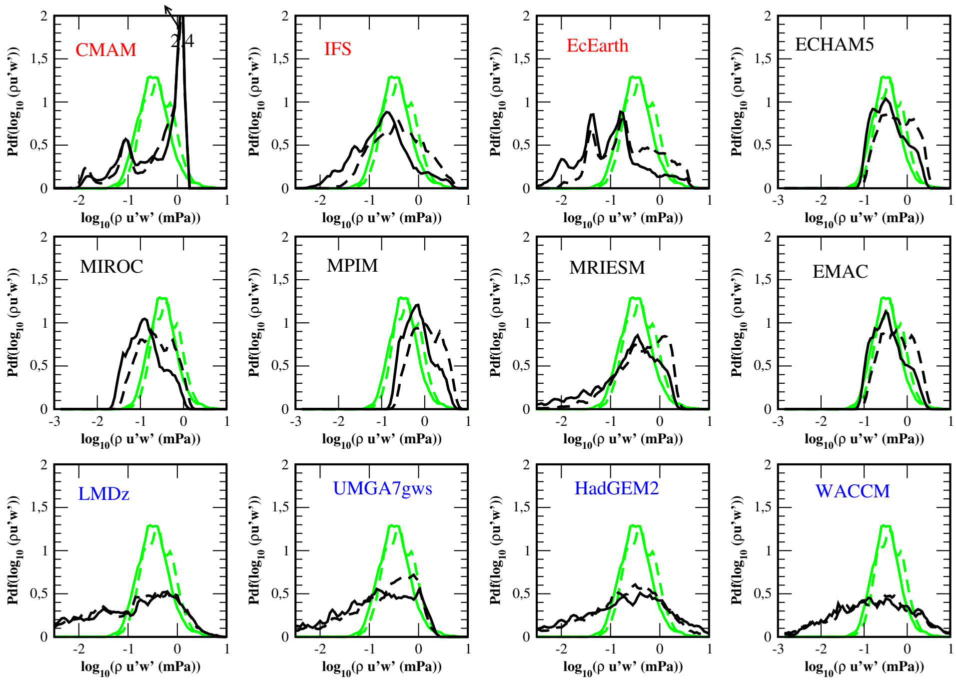


Figure 8. PDFs of daily values of Momentum flux distribution, same method as in Fig. 7. Measured values are in green, estimations using ERA5 data and the parameterizations are in black. Solid lines are for Eastward, dashed lines are for Westward.

schemes (model names in black) seem more realistic: in them the pdfs are narrower and somehow distributed quite along log normal distributions. Importantly, and in all the globally spectral schemes without convective sources (WMI and HDS) the shift of the westward pdf toward higher values compared to the eastward pdf is represented. Finally, the schemes that relate GWs to convection (names in blue) systematically have much broader pdfs, they all present a tail toward small values of the MFs, a tail that is not realsitic and that suggest that in them miss a background of wave activity existing even in the absence of convection nearby. In them also, the shift of the westward pdf toward higher values than the eastward pdfs is not much apparent, larger westward fluxes are eventually captured through changes in pdf shape than through translations (see for instance UMGA7gws and HadGEM2). If we now return to the conclusions in Green et al. (2023) that difference 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.

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 stress 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. Some scheme for instance present "medium" correlations in the eastward direction, telling that such correlation levels can be reached.

In the westward direction, the day to day correlations are ”low”, to the best and in 1 model, we can only say that such a level can be reached in the tropical regions.

Due to the low to medium level of 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 5) and frequently medium in the westward direction, which is probably enough in the context of the QBO forcing, the QBO evolving over time scales much larger than a month.

An other substantial difference concerns the pdfs of the parameterized momentum fluxes against those of the measured fluxes. Many spectral schemes have log-normal pdfs consistent with observations, providing that the launch level is not to close from the balloon location, whereas the schemes that relate the GWs to their convective sources all present tails toward small values which seem unrealistic. As intermittency is a key factor controlling the altitude at which GWs break, a factor that can have climatic impacts ((de la Cámara et al., 2016)), this should be considered seriously, at least by introducing a background in wave launching amplitude in the schemes that only consider convective sources. This issue may well also be partly sorted out by introducing lateral propagation (Amemiya & 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 to cover much wider regions. We should also test if improving the schemes parameters to improve the fit with observations improve or do 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 the QBOi models (Holt et al., 2022). Also, we could also hope that a better fit with observed values would help reduced persistent systematic errors in the QBO simulations, one of them being that models underestimate the QBO amplitude in the low stratosphere. Unfortunatly, our results so far are not much positive: a common beleive is that such an error could well be reduced by launching waves from near the tropopause, the parameterizations which do so here are not much realistic when it comes to predict MFs variabilities (over day or moths).

5 Open Research

Balloon data presented in Haase et al. (2018) can be extracted from the STRATEOLE 2 dedicated web site: https://webstr2.ipsl.polytechnique.fr

ERA5 reanalysis data are described in Hersbach et al. (2020) and can be extracted from the COPERNICUS access hub: https://scihub.copernicus.eu/

The LMDz-6A GCM used for CMIP6 project is described in Hourdin et al. (2020), it can be directly installed from the dedicated webpage: https://lmdz.lmd.jussieu.fr/utilisateurs/installation-lmdz

Acknowledgments

This work was supported by the VESRI Schmidt Future project "DataWave".

Appendix: Running the offline code

To run the models parameterizations in offline mode and compare with daily values of momentum fluxes measured during strateole 2, download the file offline_v9_Strateole_QBOi_Open.tar, on the web page:

wget `https://web.lmd.jussieu.fr/~flott/DATA/offline_v9_Strateole_QBOi_Open.tar.gz`

Then gunzip and do tar -xvf offline_v9_Strateole_QBOi_Open.tar

In the directory, offline_v9_Strateole_QBOi_Open:

Run directory It basically contains a script that compile the programs, link to the input dataset and produce various outputs. The Makefile certainly needs to be adapted to the computer.
To launch predictions for Strateole-2 phase 1, launch: ./laun_ph1ball_gwd_era5.sh
For phase 2, ph1→ph2.

Fortran Codes: all the fortran routines are located in prog.
laun_gwd_era5.f90: Main program loading input data in netcdf format and calculating drag and momentum fluxes at the balloon place.
preci_gwd_LMDz_QBOi.f90: LMDz Multiwaves routines predicting gwdrag from precipitation
gwsat_Modnam.f90: the globally spectral scheme using the Warner and McIntyre (1996)'s scheme version by J. Scinocca.
hinesgw6g_plus_subs.f HDS scheme
gw_ussp_core.f90: The WMI scheme with amplitude keyed to precipitation used in some UMGA7gws runs.
cgwcalc.f90: Multiwave scheme developped at HadGEM2's university

Input Data: All the input data are located in the directories hourly_ph1 and hourly_ph2 for phase 1 and 2 respectively. For instance, 1hr average of the strateole2 momentum fluxes are in
ALL_STRATEOLE2_Balloon_ph1_1day15min.nc
and
All_STRATEOLE2_Balloon_ph1_1hrs15min.nc
for the waves with periods between 1day and 15mn and between 1Hr and 15 mn respectively.
Still in this directory, the ERA5 reanalysis products, which include winds temperature, cloud liquid water, and surface log pressure, over a 5°x5° domain centered at the balloons drifting locations are in Input_ERA5_data_all_variables_balloons_ph1.nc.
Precipitation every hours are also included. The diabatic heatings are from fore-

cast. All datas that are only provided every 3hr are linearly interpolated in time to give hourly values.

Output data (Part 1)

All the ouputs are in the output_ph1 and output_ph2 directories:

Netcdf: contains the output of the schemes in netcdf format on the vertical column and over the 5°x5° domain over which the ER5 data are provided. There is one netcdf dataset by balloons flight each contains output from all the schemes.

Balloon_alt After post processing by the python scripts launch_script_obs.py, are extracted the MFs at balloon flight altitude.

Python Scripts

A serie of Python scripts, located into python_script are proposed to compare the outputs of the scheme to the balloon data.

launch_script_obs.py: Reads the balloon flight data of MFs and averaged over 1day and writte them in text format (ending with '.dat') and stored in **output/Balloon_alt/obs_output**

launch_prediction_eachB_ysei.py: extract from the prediction the values of the MFs at the balloons place and altitude. Results stored in text format (".dat" in **Balloon_alt/Pred_output_Balloon_altitude/**.

The next python scripts are cosmetic in the sense that they use the above two datasets to make plots of timeseries balloon averaged values, evaluate correlations, and histograms.

timeseries_obs_pred_plot_all.py Produces a lot of time series for each model and flights.

Output data (Part2) As a result, you can visualize timeseries of each flight here:

output_*ph*1/Balloon_alt/figure_timeseries

Histograms here: **output_ph1/histo**

Scatter plots and correlations here **output_ph1/correlation**

For phase 2, change ph1 in ph2.

xmgrace Alternative to calculate these diagnostics using fortran programs and xmgrace, the programs permit to combine statistics over the 2 phases of Strateole2. Just go in the directory and launch or read the README.sh file to produce the figures of the paper once the daily timeseries associated with phase 1 and 2 are produced.

Overleaf Texmaker file including all the references, figures, and texfiles to compile this version of the ms.

References

Alexander, M. J., Beres, J. H., & Pfister, L. (2000). Tropical stratospheric gravity wave activity and relationships to clouds. *Journal of Geophysical Research: Atmospheres*, *105*(D17), 22299-22309. doi: https://doi.org/10.1029/2000JD900326

Alexander, M. J., & Dunkerton, T. J. (1999). A Spectral Parameterization of Mean-Flow Forcing due to Breaking Gravity Waves. *J. Atmos. Sci.*, *56*(24), 4167–4182. doi: 10.1175/1520-0469(1999)056⟨4167:ASPOMF⟩2.0.CO;2

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., ... 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. doi: https://doi.org/10.1002/qj.637

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*(7), e2020JD033954. (e2020JD033954 2020JD033954) doi: https://doi.org/10.1029/2020JD033954

Amemiya, A., & Sato, K. (2016). A new gravity wave parameterization including three-dimensional propagation. *Journal of the Meteorological Society of Japan. Ser. II*, *94*(3), 237-256. doi: 10.2151/jmsj.2016-013

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

Anstey, J. A., Banyard, T. P., Butchart, N., Coy, L., Newman, P. A., Osprey, S., & Wright, C. J. (2021). Prospect of increased disruption to the QBO in a changing climate. *Geophysical Research Letters*, *48*(15), e2021GL093058. doi: https://doi.org/10.1029/2021GL093058

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*(4), 1649 - 1665. doi: https://doi.org/10.1175/JAS-D-15-0099.1

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., ... Takahashi, M. (2001). The quasi-biennial oscillation. *Rev. Geophys.*, *39*(2), 179-229. doi: 10.1029/1999RG00007

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*(D10). doi: https://doi.org/10.1029/2004JD005504

Bushell, A. C., Butchart, N., Derbyshire, S. H., Jackson, D. R., Shutts, G. J., Vosper, S. B., & 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*(11), 4349–4371. doi: https://doi.org/10.1175/JAS-D-15-0022.1

Butchart, N., Anstey, J. A., Hamilton, K., Osprey, S., McLandress, C., Bushell, A. C., ... 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*(3), 1009–1032. doi: 10.5194/gmd-11-1009-2018

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*(5), 923 - 941. doi: 10.1175/1520-0469(2002)059⟨0923:GWFFPA⟩2.0.CO;2

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*(4), 739 - 759. doi: https://doi.org/10.1175/2010JAS3552.1

Christiansen, B., Yang, S., & Madsen, M. S. (2016). Do strong warm enso events control the phase of the stratospheric qbo? *Geophysical Research Letters*, *43*(19), 10,489-10,495. doi: https://doi.org/10.1002/2016GL070751

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*(15), e2021JD035165. doi: https://doi.org/10.1029/2021JD035165

Davini, P., von Hardenberg, J., Corti, S., Christensen, H. M., Juricke, S., Subramanian, A., … 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*(3), 1383–1402. Retrieved from `https://gmd.copernicus.org/articles/10/1383/2017/` doi: 10.5194/gmd-10-1383-2017

de la Cámara, A., & Lott, F. (2015). A parameterization of gravity waves emitted by fronts and jets. *Geophys. Res. Lett.*, *42*(6), 2071-2078. doi: 10.1002/2015GL063298

de la Cámara, A., Lott, F., & Hertzog, A. (2014). Intermittency in a stochastic parameterization of nonorographic gravity waves. *J. Geophys. Res.: Atmospheres*, *119*(21), 11905-11919. doi: 10.1002/2014JD022002

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. *J. Atmos. Sci.*, *73*(8), 3213-3226. doi: https://doi.org/10.1175/JAS-D-15-0377.1

Eckermann, S. D. (2011). Explicitly Stochastic Parameterization of Nonorographic Gravity Wave Drag. *J. Atmos. Sci.*, *68*, 1749–1765. doi: 10.1175/2011JAS3684.1

Ern, M., Ploeger, F., Preusse, P., Gille, J., Gray, L. J., Kalisch, S., … Riese, M. (2014). Interaction of gravity waves with the QBO: A satellite perspective. *Journal of Geophysical Research: Atmospheres*, *119*, 2329 - 2355. doi: https://doi.org/10.1002/2013JD020731

Fovell, R., Durran, D., & Holton, J. R. (1992). Numerical simulations of convectively generated stratospheric gravity waves. *Journal of Atmospheric Sciences*, *49*(16), 1427 - 1442. doi: 10.1175/1520-0469(1992)049⟨1427:NSOCGS⟩2.0.CO;2

Fueglistaler, S., Legras, B., Beljaars, A., Morcrette, J.-J., Simmons, A., Tompkins, A. M., & 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*(638), 21-37. doi: https://doi.org/10.1002/qj.361

Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., … Zhou, T. (2013). A comparison between gravity wave momentum fluxes in observations and climate models. *J. Atmos. Sci.*, *26*(17).

Green, B., Sheshadri, A., Alexander, M., Bramberger, M., & 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., … Venel, S. (2018). Around the world in 84 days [Dataset]. *Eos*, *99*. doi: https://doi.org/10.1029/2018EO091907

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis [Dataset]. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hertzog, A. (2007). The stratéole-vorcore long-duration balloon experiment: A personal perspective. *Space Research Today*, *169*, 43-48. Retrieved from https://www.sciencedirect.com/science/article/pii/S1752929807800478 doi: https://doi.org/10.1016/S1752-9298(07)80047-8

Hertzog, A., Alexander, M. J., & Plougonven, R. (2012). On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere. *Journal of the Atmospheric Sciences*(11), 3433–3448. doi: 10.1175/JAS-D-12-09.1

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*(11), 1361 - 1379. doi: https://doi.org/10.1175/1520-0469(1991)048⟨1361:TSOGWI⟩2.0.CO;2

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. *J. Atmos. Solar Terr. Phys.*, *59*(4), 387-400. doi: 10.1016/S1364-6826(96)00080-6

Holt, L. A., Lott, F., Garcia, R. R., Kiladis, G. N., Cheng, Y.-M., Anstey, J. A., ... 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*(744), 1541-1567. doi: https://doi.org/10.1002/qj.3827

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., ... 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*(7), e2019MS001892. doi: https://doi.org/10.1029/2019MS001892

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. *J. Atmos. Sci.*, *72*(9). doi: DOI:10.1175/JAS-D-14-0324.1

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*(17), 9690-9707. doi: https://doi.org/10.1002/jgrd.50781

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., ... Kern, B. (2010). Development cycle 2 of the modular earth submodel system (messy2). *Geoscientific Model Development*, *3*(2), 717–752. doi: 10.5194/gmd-3-717-2010

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*(10), 3167 - 3189. doi: 10.1175/JAS-D-17-0053.1

Lane, T. P., & Moncrieff, M. W. (2008). Stratospheric gravity waves generated by multiscale tropical convection. *J. Atmos. Sci.*, *65*, 2598–2614. doi: DOI:10.1175/2007JAS2601.1

Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.*, *86*(C10), 9707-9714. doi: 10.1029/JC086iC10p09707

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*(1), e2021JD035785. (e2021JD035785 2021JD035785) doi: https://doi.org/10.1029/2021JD035785

Lott, F., & Guez, L. (2013). A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *J. Geophys. Res.*, *118*(16), 8897-8909. doi: 10.1002/jgrd.50705

Lott, F., Guez, L., & Maury, P. (2012). A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere. *Geophys. Res. Lett.*, *39*(6), L06807. doi: 10.1029/2012GL051001

Lott, F., Rani, R., Podglajen, A., Codron, F., Guez, L., Hertzog, A., & Plougonven, R. (2023). Direct comparison between a non-orographic gravity wave drag scheme and constant level balloons. *Journal of Geophysical Research: Atmospheres*, *128*(4), e2022JD037585. doi: https://doi.org/10.1029/2022JD037585

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*(D22), 25751-25762. doi: 10.1029/97JD01096

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*(723), 2755-2771. doi: https://doi.org/10.1002/qj.3591

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*(22), 5905 - 5926. doi: https://doi.org/10.1175/2010JCLI3490.1

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*(D14). doi: https://doi.org/10.1029/2004JD004656

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. *J. Atmos. Sci.*, *74*(4), 1075-1093. doi: 10.1175/JAS-D-16-0096.1

Pohlmann, H., Müller, W. A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F. S. E., … Marotzke, J. (2013). Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophysical Research Letters*, *40*(21), 5798-5802. doi: https://doi.org/10.1002/2013GL058051

Rabier, F., Bouchard, A., Brun, E., Doerenbecher, A., Guedj, S., Guidard, V., … Steinle, P. (2010, January). The Concordiasi Project in Antarctica. *Bulletin of the American Meteorological Society*, *91*(1), 69-86. Retrieved from `https://hal-insu.archives-ouvertes.fr/insu-00562459` doi: 10.1175/2009BAMS2764.1

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*(1), 136 - 156. doi: 10.1175/2009JAS3112.1

Richter, J. H., Sassi, F., & Garcia, R. R. (2010b). Toward a Physically Based Gravity Wave Source Parameterization in a General Circulation Model. *J. Atmos. Sci.*, *67*(1), 136–156. doi: 10.1175/2009JAS3112.1

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., … Schulzweida, U. (2006). Sensitivity of simulated climate to horizontal and vertical resolution in the echam5 atmosphere model. *Journal of Climate*, *19*(16), 3771 - 3791. doi: https://doi.org/10.1175/JCLI3824.1

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*(9), 1473 - 1489. doi: https://doi.org/10.1175/1520-0469(2002)059⟨1473:IOASGW⟩2.0.CO;2

Scinocca, J. F. (2003). An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *Journal of the Atmospheric Sciences*, *60*(4), 667 - 682. doi: https://doi.org/10.1175/1520-0469(2003)060⟨0667:AASNGW⟩2.0.CO;2

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*(9), 2147-2162. doi: https://doi.org/10.1029/2018MS001297

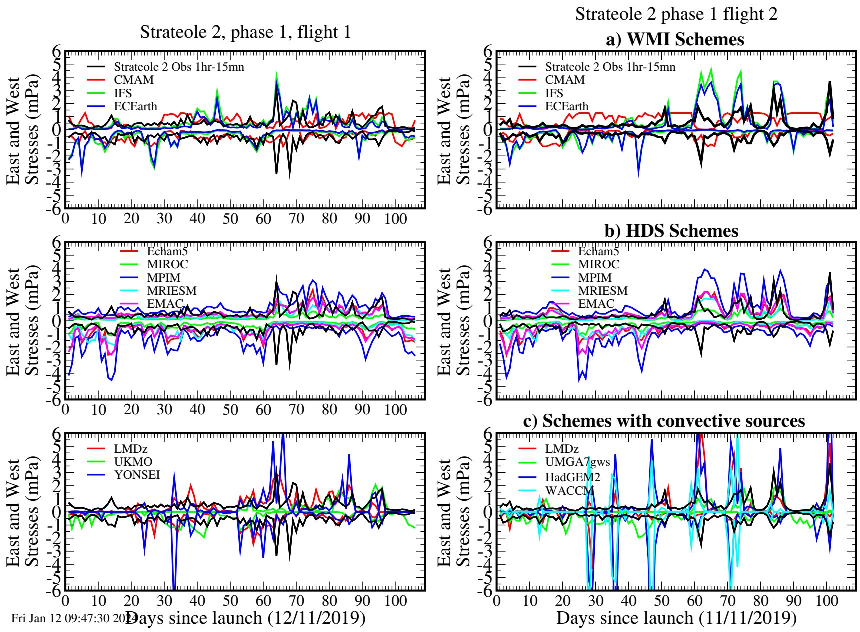
Song, I.-S., & Chun, H.-Y. (2005). Momentum flux spectrum of convectively forced internal gravity waves and its application to gravity wave drag parameterization. part i: Theory. *J. Atmos. Sci.*, *62*(1), 107-124. doi: https://doi.org/10.1175/JAS-3363.1

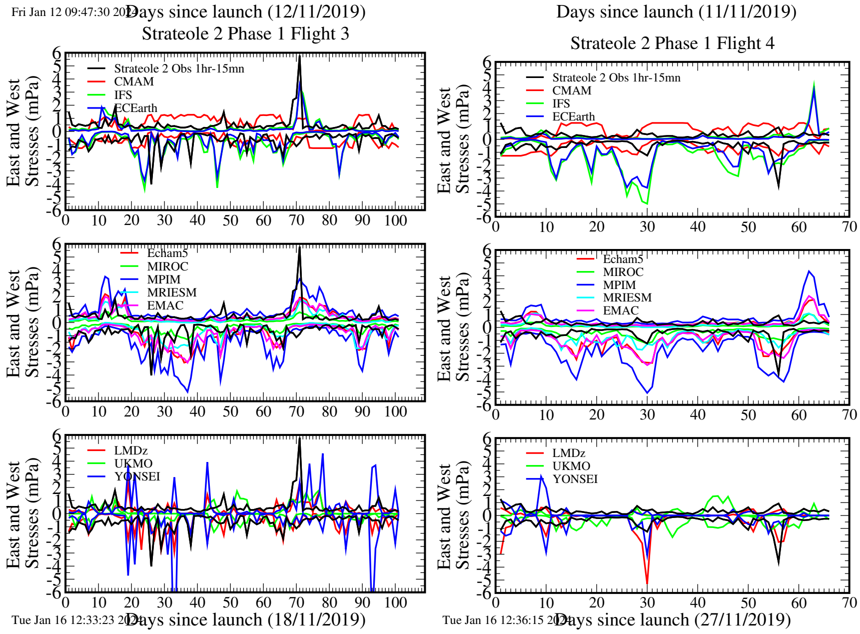
Warner, C. D., & McIntyre, M. E. (1996). On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *J. Atmos. Sci.*, *53*(22), 3213-3235. doi: 10.1175/1520-0469(1996)053⟨3213:OTPADO⟩2.0.CO;2

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. doi: 10.1186/BF03353209

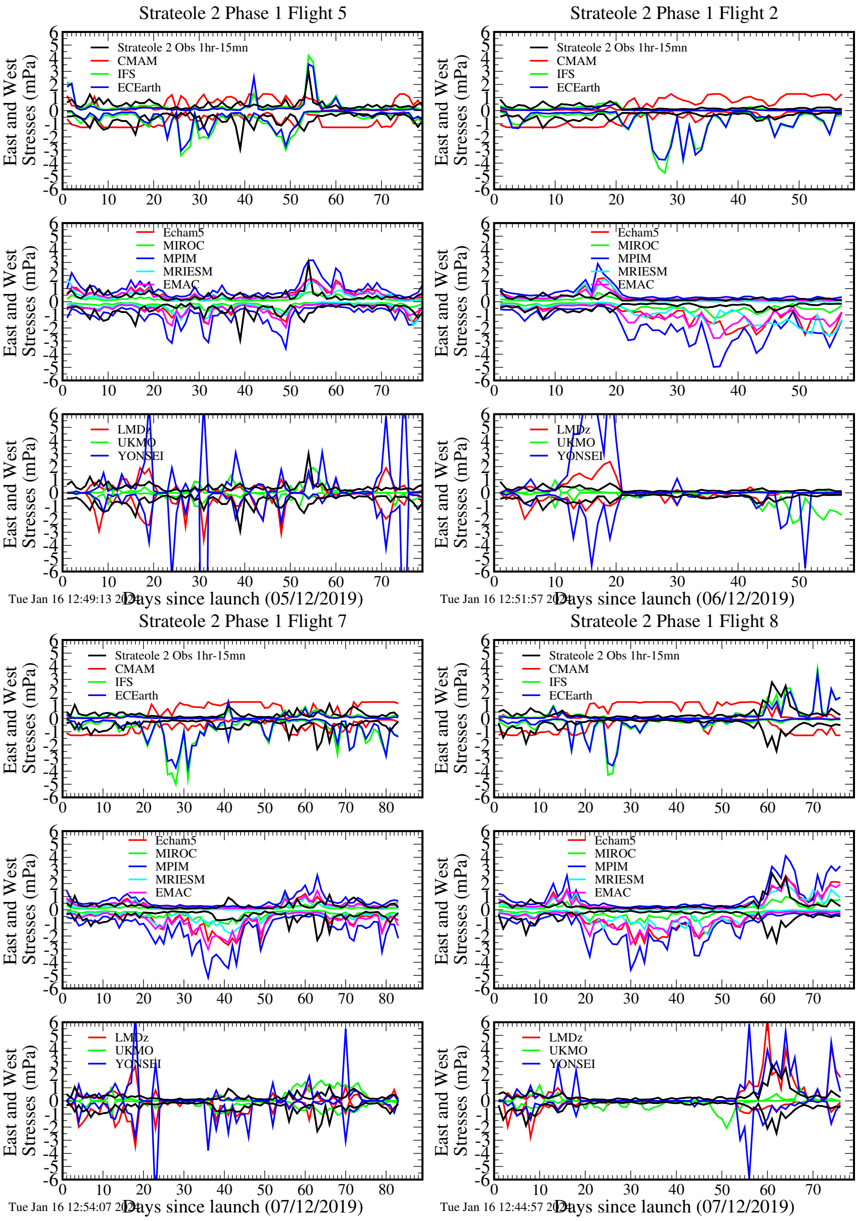
Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., ... Kawamiya, M. (2011). Miroc-esm 2010: model description and basic results of cmip5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. doi: 10.5194/gmd-4-845-2011

822 # Supplementary: Phase 1 flights and models

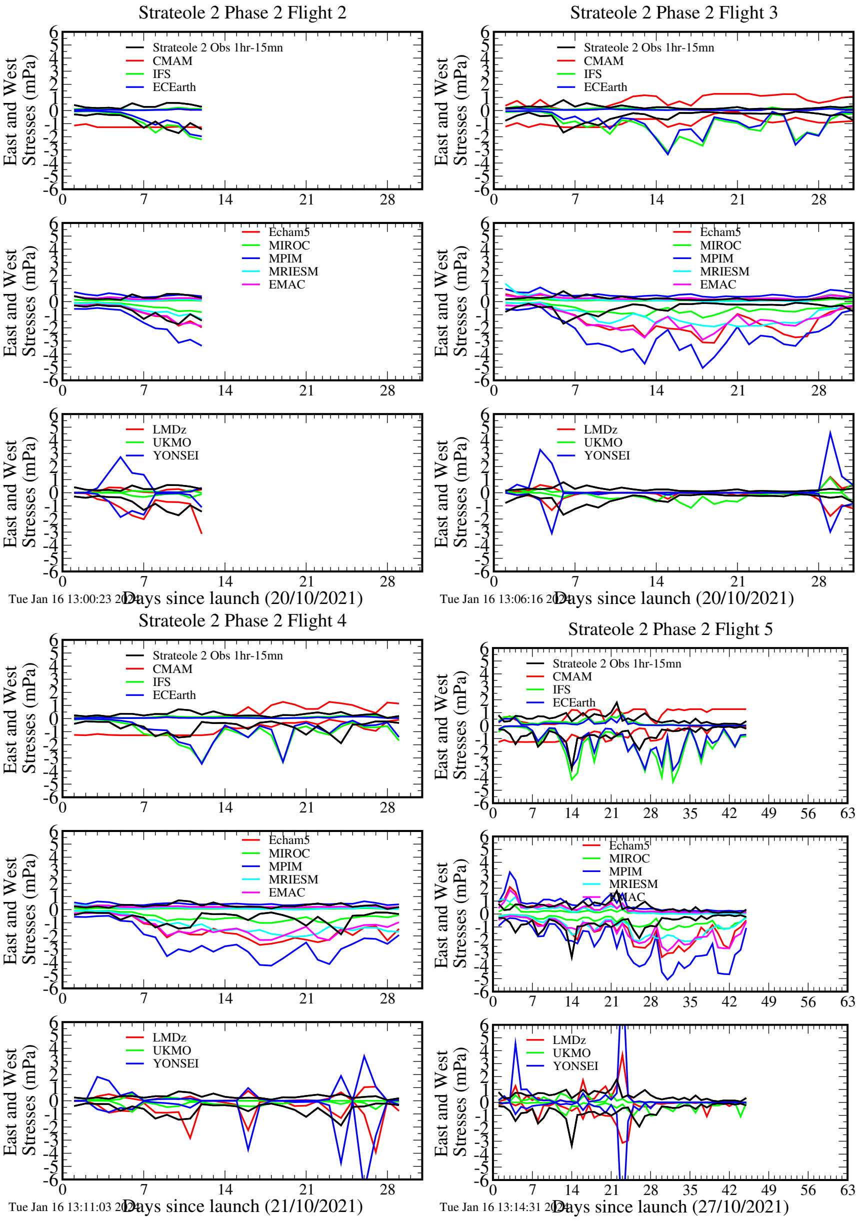




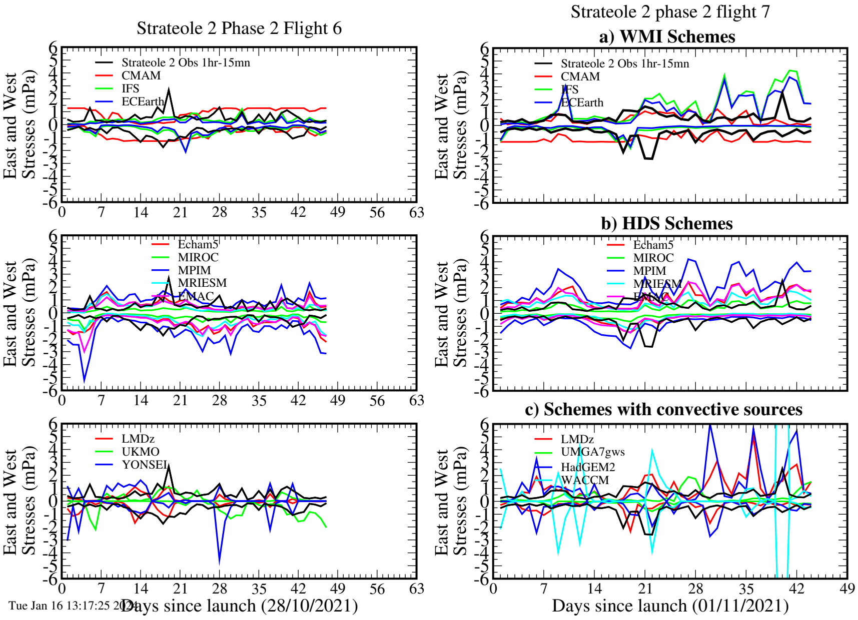
Supplementary: Phase 1 flights and models (continued)

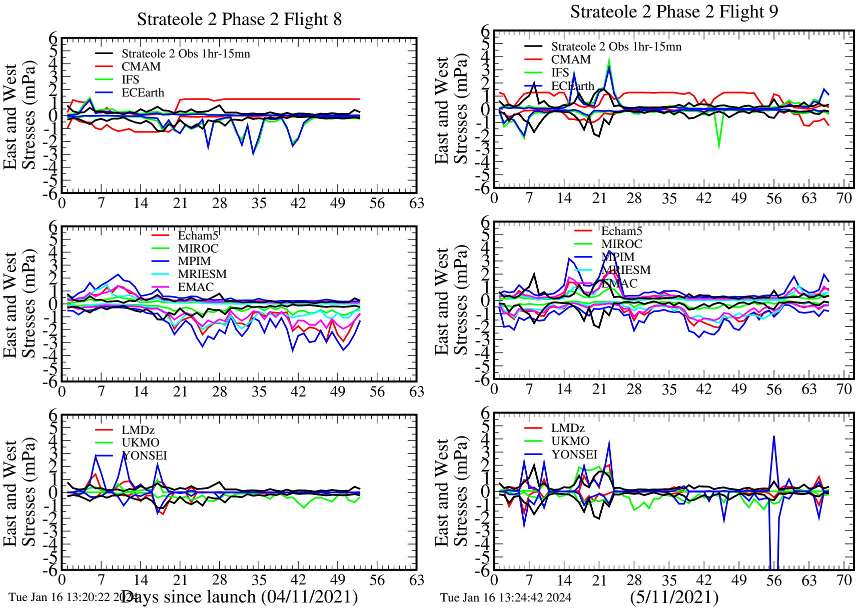


Supplementary: Phase 2 flights and models

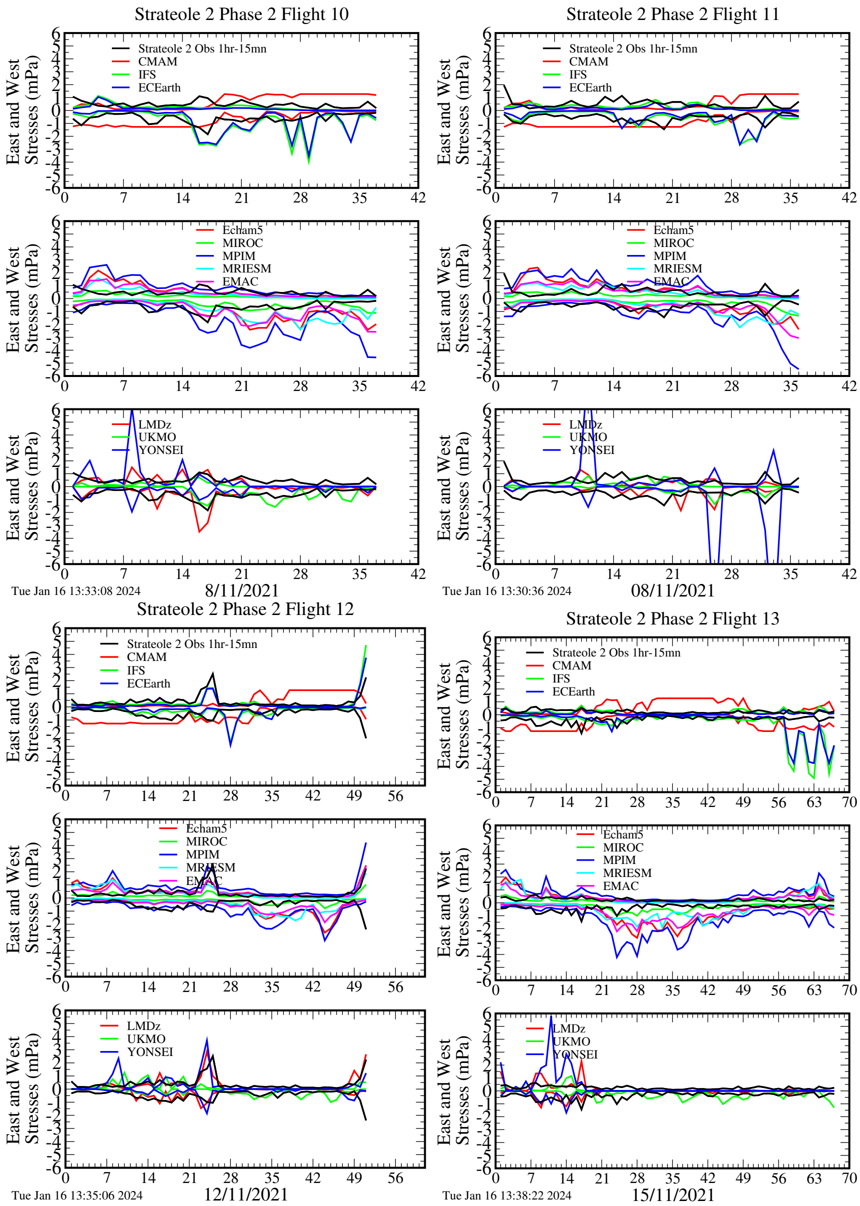


Supplementary: Phase 2 flights and models (continued)

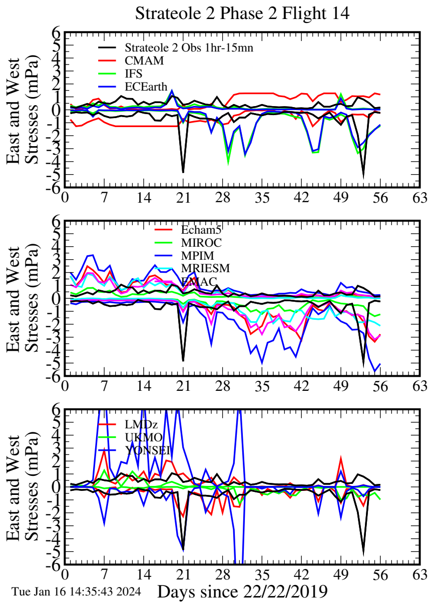


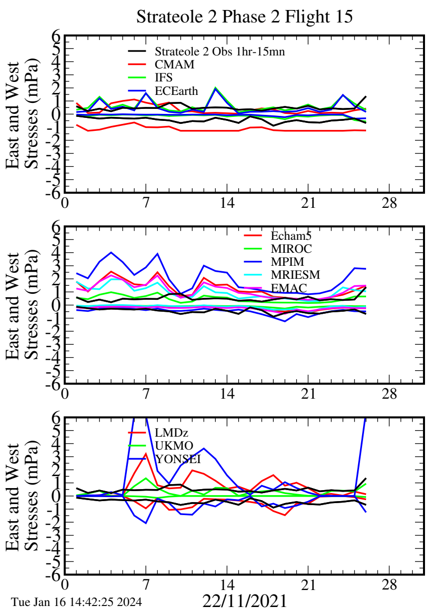


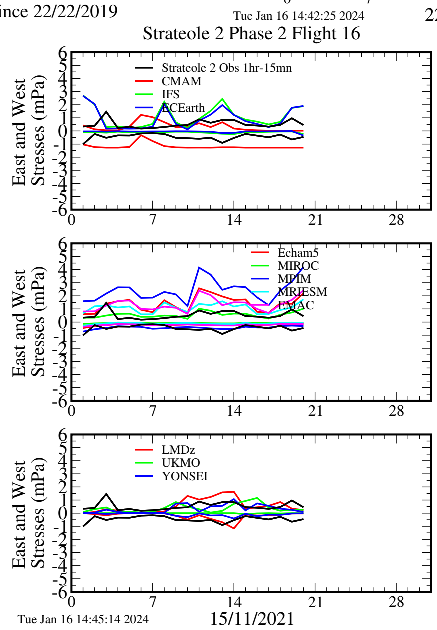
Supplementary: Phase 2 flights and models (continued)



Supplementary: Phase 2 flights and models (continued)


Strateole 2 Phase 2 Flight 14
East and West Stresses (mPa)
Strateole 2 Obs 1hr-15mn
CMAM
IFS
ECEarth
Echam5
MIROC
MPIM
MRIESM
EMAC
LMDz
UKMO
YONSEI
Tue Jan 16 14:35:43 2024
Days since 22/22/2019


Strateole 2 Phase 2 Flight 15
East and West Stresses (mPa)
Strateole 2 Obs 1hr-15mn
CMAM
IFS
ECEarth
Echam5
MIROC
MPIM
MRIESM
EMAC
LMDz
UKMO
YONSEI
Tue Jan 16 14:42:25 2024
22/11/2021


Strateole 2 Phase 2 Flight 16
East and West Stresses (mPa)
Strateole 2 Obs 1hr-15mn
CMAM
IFS
ECEarth
Echam5
MIROC
MPIM
MRIESM
EMAC
LMDz
UKMO
YONSEI
Tue Jan 16 14:45:14 2024
15/11/2021