Referee #2 (Corwin Wright)

The authors use output from a specified-dynamics run of the Canadian Middle Atmosphere Model (CMAM) to investigate how its orographic gravity wave parameterisation captures the variability at short time and space scales (‘intermittency’) of gravity wave forcing. They focus on three northern-hemisphere wave ‘hotspots’, namely the Himalayas, Rocky Mountains and the generalised high topography of north-east Asia. They compare their results to intermittency in observational data from limb-sounding satellites and to previous studies. They conclude that (i) the assumption of vertically-propagating waves in parameterisations is realistic in the lower stratosphere; (ii) that peak events can have opposing effects on the upper stratosphere and above depending on the specific region chosen, and (iii) that in the zonal mean positive OGWD anomalies contribute to mesospheric cooling.

The paper is clearly laid-out and well-written, and I have no major criticisms of the content or the presentation. In particular, the work is well-contextualised within the literature, the figures well-designed (with good use of colour to avoid significant issues for red/green colourblind readers), and the implications of their results are made clear. There are a few minor spelling and grammar errors, but these should be easily resolvable during the copy-editing stage. I support publication with at most minor revisions.

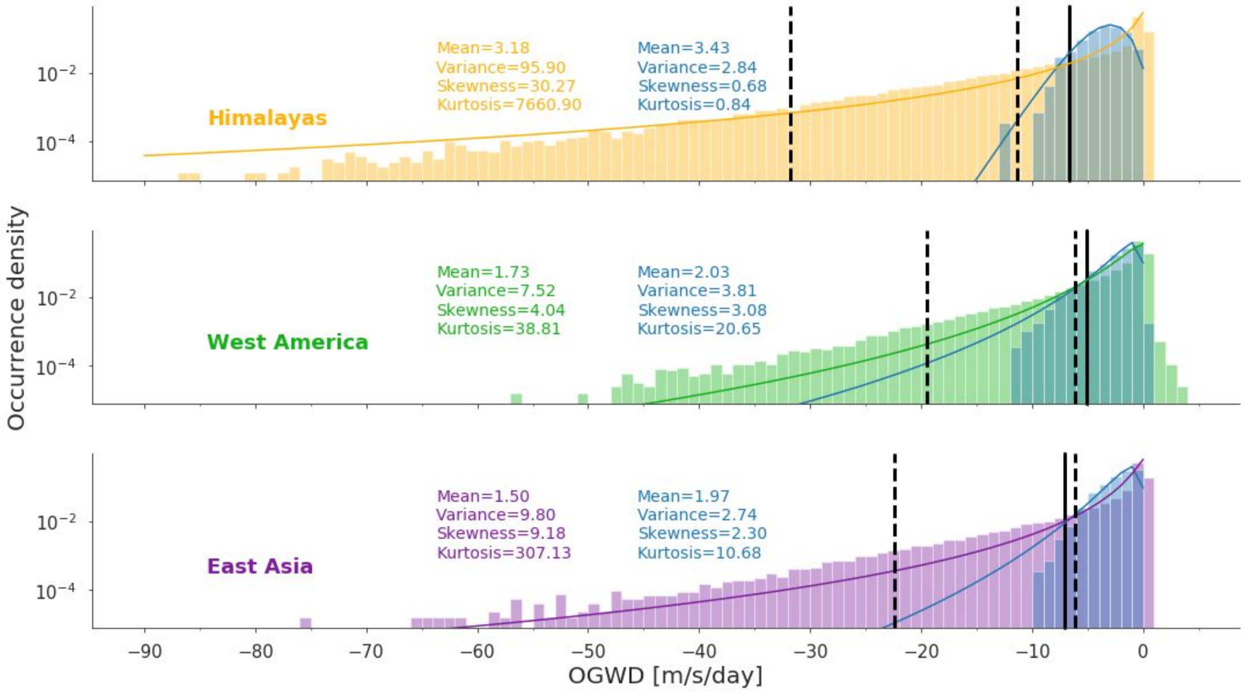
We thank Corwin Wright for his thoughtful comments and suggestions and the overall enthusiastic review. Our response including changes made to the manuscript is listed below.

Please note that we do not conclude from our results that the assumption of vertically-propagating waves in parameterisations is realistic in the lower stratosphere. We only refer to the results Kalisch et al. (2014), who found good agreement between GWD with oblique and vertical-only propagation in the lower stratosphere to argue that our results may bear some general similarities with how these hotspots are formed and affect the atmosphere in reality.

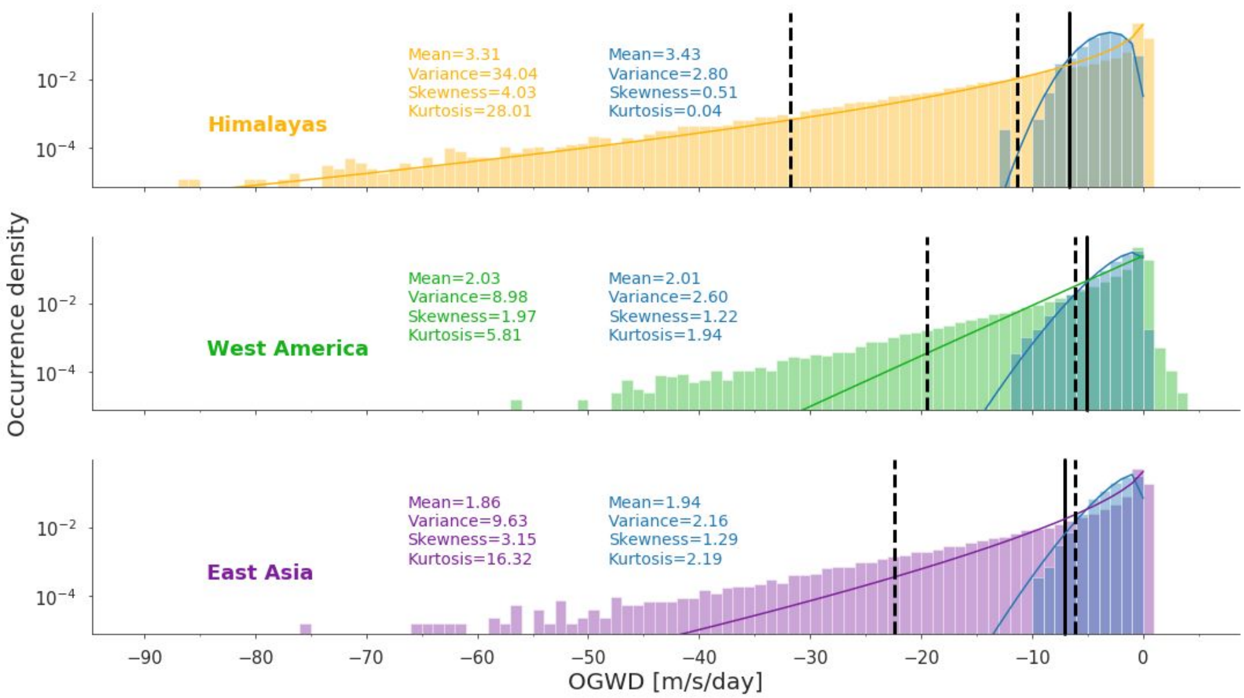
Major comment

I have only one significant comment, which should not impede publication. This is that in Section 3.3 (Composite Analysis) they justify their choice to analyse the data at the regional rather than local scale with the statement that "Figs. 6 and 7 showed that spatial averaging inside the hotspots maintains the intermittent feature of the OGWD and only the information of the long tails of the distribution with extreme (and apparently very localized) drag values of −20m/s/day is lost". I disagree with this statement, especially as relates to Figure 6 - to my eye, the form of the regional-scale distribution (blue bars) is quite distinct to that of the grid-scale distribution (coloured bars). I agree that the appropriate step to take in Section 3.3 is to analyse at the regional scale, and in particular that this is necessary to make the results useful for model developments - I just disagree with this justification, and think it would be better to admit that the distributions are a bit different and explain why their results are still meaningful despite this.
We revised Fig. 6 according to a comment of Ref#1 and added the corresponding fits of log-normal distributions and their corresponding moments (see below). Note that the negative

values of OGWD were used as positive values for fitting (and the originally positive part of the OGWD spectrum was not fitted).



Furthermore, we enclosed the following figure into the supplement where fits of the Weibull Minimum Extreme Value distributions and their corresponding moments are added to document the fact that it fits more closely the spatially averaged values.



The sentence mentioned above was revised in the following way:
"Figs. 6 and 7 showed that spatial averaging inside the hotspots changes the distribution (compare corresponding moments displayed in Fig. 6) as the information about the long tails of the distributions with strong (and apparently very localized) drag values of −15 m/s/day is lost and also the approximation by the log-normal distribution does no longer characterize the spatially averaged OGWD distribution as good as the Weibull Minimum Extreme Value distribution (see Fig. S15). On the other hand, the mean value of both distributions is approximately equal justifying the utilization of composites from spatially averaged OGWD for dynamical analysis. Magnitude of higher order moments (especially kurtosis as a measure of the "tailedness" of the distributions) is another measure of the intermittency (Zel'Dovich et al., 1987; Mahrt, 1988). The values given in Fig. 6 show that the intermittency is highest for HI and lowest for WA. After spatial averaging, the intermittency is reduced at most for HI which is, however, caused by a poor accuracy of the log-normal fit for EA and WA.".

Minor specific comments

L019: the paper comes in at quite a high level without much preamble, and in particular the mention of the BDC without any contextualisation might be hard for someone entering the field to follow. Would suggesting adding a brief description of what this to soften it for e.g. a starting research student.

This part of the introduction has been rephrased: "The meridional overturning circulation in the middle atmosphere is known as the Brewer-Dobson circulation (BDC), whereby air rises in the tropics and then moves upward and poleward before descending in the middle and high latitudes (Butchart, 2014). While the BDC is believed to be driven mainly by Rossby waves, its upper part consisting from the mesospheric inter-hemispheric circulation from the summer to winter pole is dominated by GWs (Plumb, 2002; Alexander, 2013). "

L031: you refer to the GWs as being spatially asymmetric, but don't say in what direction. From later in the paper I assume this is E/W rather than (e.g.) N/S, but you should mention this here.

We mean spatial asymmetry in general. We revised according to a comment of Ref#1.

L051: This sentence expands out to "The study is based on a simulation performed with the [Canadian Middle Atmosphere Model] simulation". Suggest rephrasing to be a bit less repetitious.

Revised according to a technical comment of Ref#1.

L053: how variable a vertical resolution? Is it high enough for your purposes in the regions you study? Suggest briefly clarifying this.

According to Fig. 1 in Scinocca et al (2008), there are 71 vertical levels with a resolution varying from 1 km around the tropopause to about 2.5 km in the mesosphere. As we cannot compare the results of simulations with different resolutions, it is currently not so easy to answer the question regarding sufficient resolution. However, the resolution of this CMAM simulation ranges among the higher resolved models in current multi-model assessments (e.g. CCMI or CMIP) and these are the models we want to understand. To make this point, we revised the sentence in the following way: “CMAM is a chemistry-climate model with 71 levels in the vertical spanning from the surface up to 7 · 10−4 hPa (about 100 km) with a vertical resolution varying from 1 km around the tropopause to about 2.5 km in the mesosphere (see Fig. 1 in Scinocca et al., 2008)”.

L055: you refer to "spatial scales of <T21". I assume by this you mean scales which are physically larger than T21, but this is not what the sentence says - it instead refers to (relatively) small-physical-scale features. Suggest rephrasing for clarity.
We rephrase it in the following way: “This means that Newtonian relaxation ("nudging") on spatial scales of wavenumbers <21 to the 6-hourly horizontal winds...”.

L064: you say the parameterisation launches "two vertically propagating zero-phase-speed waves" but not how often. Per timestep? Per day? Please clarify.
They are required to be launched at every time step for each GCM grid cell, when conditions prone to launch freely propagating GWs are met (Scinocca and McFarlane, 2000). It was clarified in the manuscript accordingly: “The OGW scheme launches every time step (when the conditions for sourcing of freely propagating waves are met) two vertically propagating zero-phase-speed waves with orientation and magnitude depending on the near-surface static stability, wind speed and direction relative to the subgrid topography (anisotropic effects; Scinocca and McFarlane, 2000).”.

L071: you say the parameters are tuned to exert a "reasonable" drag. In what sense?
The sentence was revised in the following way: “ These parameters have been tuned for polar-ozone chemistry studies using CMAM (McLandress et al, 2013).”.

L083: you also need to clarify: (i) whether HIRDLS was assimilated in MERRA [I assume not] and (ii) whether HIRDLS or SABER were assimilated by JRA55.
The sentence was revised in the following way: “Note that neither SABER nor HIRDLS are assimilated in the reanalyses (Wright and Hindley, 2018)”.

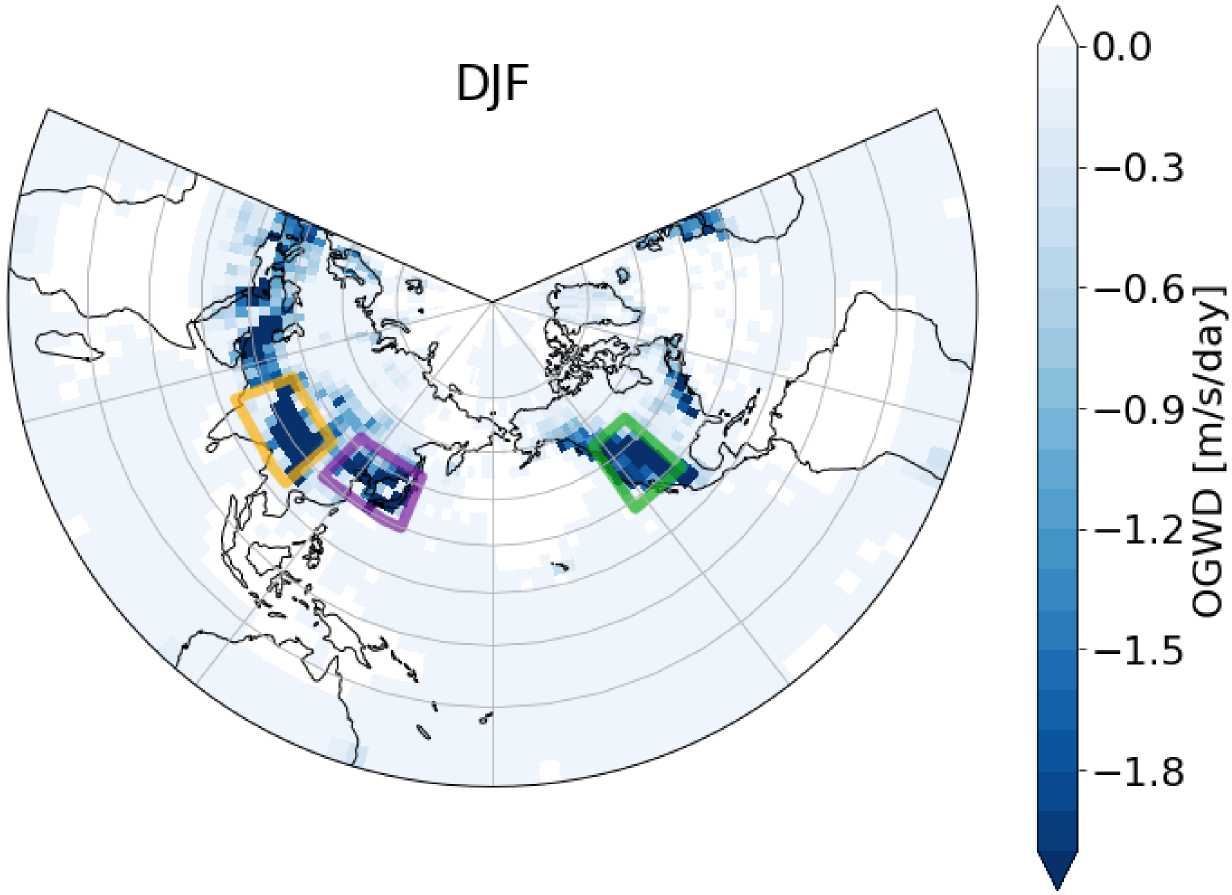
L088: missing close-bracket.
Thank you for reading carefully and for providing a good suggestion. We revised the text accordingly.

L100: "GWD constitutes about half of the net forcing in the LS" - at the global, zonally-local, or regional scale?
The sentence was revised in the following way: “In reanalyses, zonally averaged GWD constitutes about half of the net forcing in the LS (Albers and Birner, 2014; Abalos et al., 2015; Sato and Hirano, 2019).”

Figure 2: I would strongly suggest extending this figure further south - the boxes you

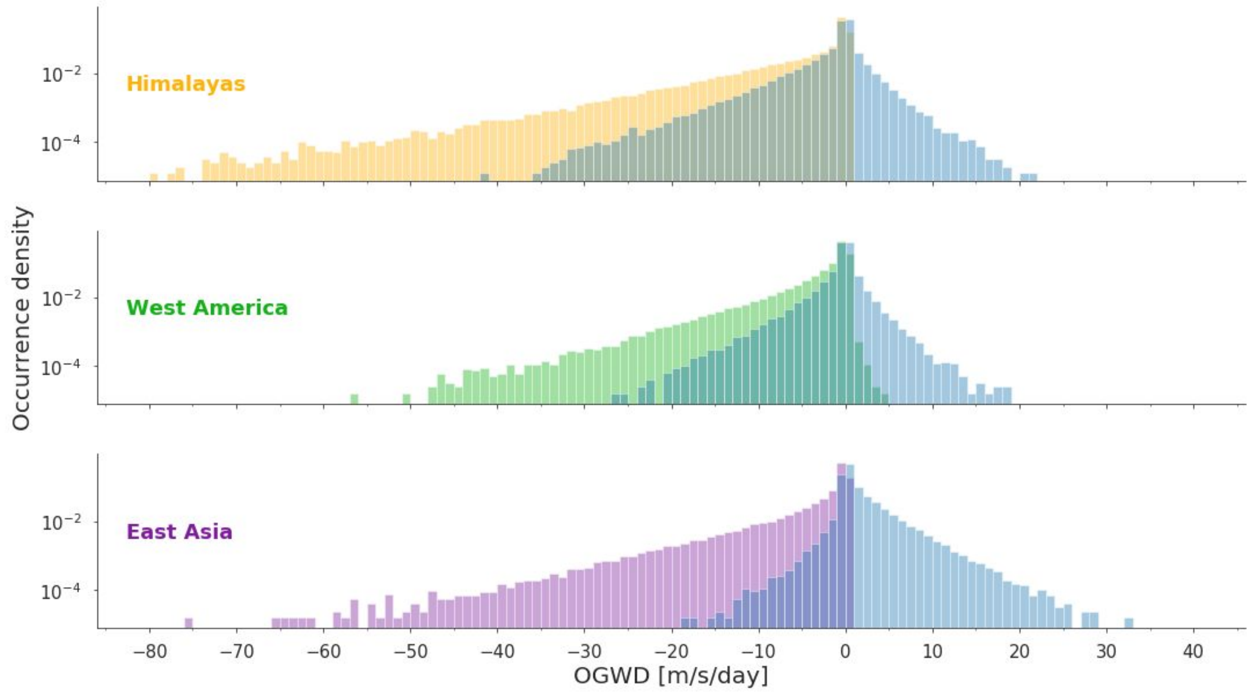
highlight are so close to the "horizon" that it is very difficult visually to work out the areas they cover. This is particularly the case for the Himalayas box - I have real trouble working out which parts of the mountain range it covers.

We changed the map projection from the Orthographic to the Lambert Conformal conic to cover all boxes in a clearer way (see below).



L115: You say that the meridional accelerations were not analysed at all. Are you certain that they have cannot make a significant contribution? For example, over the edge of Antarctica (a long way from your regions!), mountain-generated GWMFs often have significant meridional components up to 50% of the zonal component - is that definitely untrue for these Northern hemisphere regions? In particular, since the main ridge of the Himalayas is aligned at a slight angle to the E-W direction, I would perhaps expect some degree of meridional forcing from there, even if not the Rockies since they're so cleanly N-S.

The dynamical role of meridional acceleration is to a large extent unknown as mentioned for example in Šácha et al. (2016). We agree that this shall, however, not be a reason for not analyzing it. Comparing meridional (colored bars) and zonal (blue bars) components of OGWD at 70 hPa at our selected regions in the figure below (mimicking Fig. 6) confirms your statement that the meridional component of mountain-generated GWD can reach up to 50% of the zonal component. It is valid especially for Himalayas and West America. East Asia produces larger fraction od positive (poleward) meridional OGWD in comparison with the other two regions. We inserted the figure in the supplement with the following sentence in the manuscript: “As shown in Fig. S1 comparing meridional and zonal components of OGWD at 70 hPa in our selected regions, the meridional component of OGWD can significantly contribute to the total drag from parameterized OGWs in CMAM-sd. However, the dynamical role of the meridional acceleration is widely unknown (e.g. Šácha et al., 2016). Therefore, we do not consider it in the analysis of the current paper, but encourage its further investigation in future studies.”.



L145: the large difference between HIRDLS/SABER and the model is in fact a lot less than it looks from the raw numbers, and definitely less in reality than the factor of two or more inferred here. This is due to two problems affecting limb-sounder GW measurements. Firstly, and most importantly, these satellites are making a 2d cut at a high angle through the 3d wave field, travelling primarily meridionally. As a result, if you are trying to measure zonally-oriented waves (which is implicit in your comparison), then these instruments will massively underestimate wave momentum fluxes, since they measure the PROJECTION of the horizontal wavelength in the ALONG-TRACK direction. This is discussed well by e.g. Ern et al (JGR, 2004) and discussed in slightly more detail in Wright et al (ACP, 2015), and this effect will massively low-bias the measured momentum fluxes. Secondly, and less importantly, spectral fitting methods such as those used to measure GWs in these data will inherently underestimate peak amplitudes, again skewing the data low. Indeed, I would expect the values in GRACILE to be closer to a lower bound than a true estimate of the actual MF. I’m not sure how to integrate this into your text, but a brief mention that these limb-sounder estimates are

known to be significantly low-biased at a systematic level would be helpful.
We added the following mention in the manuscript: “Note that limb-sounder momentum flux estimates from HIRDLS and SABER are known to be systematically low-biased (e.g. Wright et al., 2015; Ern et al., 2004) which may be one of reasons for such a significant overestimation as shown in Fig. 4.”.

L173: word choice is a little sticky here: a "negative [gravity-wave drag]" is, literally, a gravity-wave acceleration. Would suggest a slight rephrase to make absolutely clear what you mean.
We revised the sentence in the following way: “...amplitude and meridional distribution of the negative values of GWD”.

Figure 4: it’s quite tricky to follow the seasonal cycle in the GRACILE data due to the scale. Maybe try a log-scale ordinate?
We changed Fig. 4 in the following way:

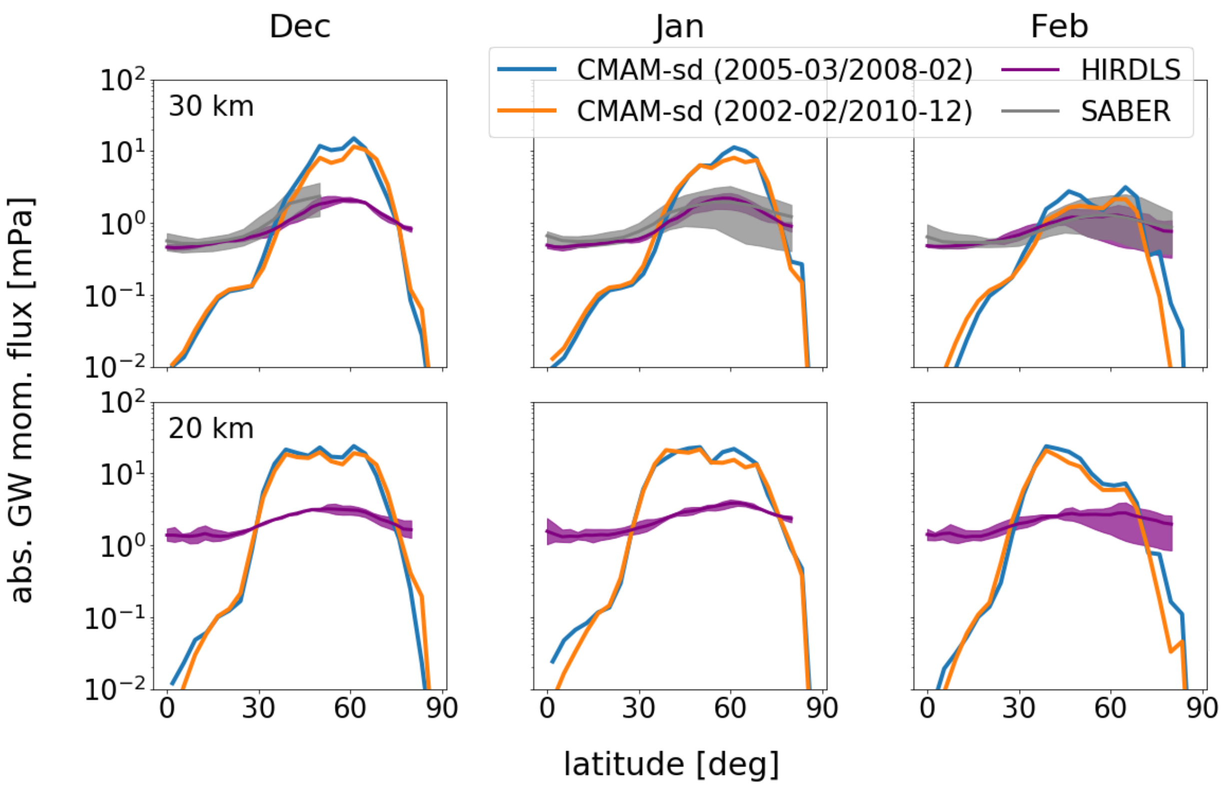


Figure 5: same comment for the NOWD line.
In our opinion, the log axis is not applicable here and we are convinced that all important details can actually be seen well enough in the original depiction. Nevertheless, a reader interested in more details can use the code and data availability.

Figure 6 caption and Line 194: I think you have the colours of EA and WA described the wrong way around.
Thank you for reading carefully and for providing a good suggestion. We revised the text accordingly.

L207: "Surprisingly, the second most frequent OGWD value is a small positive drag for all hotspots". Is this that surprising? For example, +1 is a lot closer to -1 than -50 is, so it seems intuitive to me that a small positive value would occur more frequently than a large negative value given the form of the distribution. Would suggest removing the Line.
Our motivation for the word "surprisingly" comes from the large-scale circulation point of view, where OGWs are considered to produce negative accelerations only. As for the meridional accelerations, the dynamical effect of positive accelerations is also largely unknown. We revised the sentence in the following way: "Note that for WA we find in addition some less frequent positive drags up to 5 m/s/day".

Table 1: Why are the months numbered rather than named, given you discuss them by name? And why are SON and MAM included here but not JJA, when they haven't been in the rest of the paper.
The months in Table 1 were rather named and reduced to DJF only.

Code/data availability statement: excellent, but the first two sentences are ungrammatical.
Corrected.