First, we would like to thank both the referees and the editor Prof. Dr. Jürg Schmidli for facilitating an open minded and constructive review process. The thoughtful comments and well-minded suggestions of both reviewers helped us to improve the readability and clarity of the text and graphics in the revised version.

In particular we would like to thank the Reviewer#2 for his comments, suggestions and thoughtful questions and points raised. We carefully considered all the comments and suggestions. Below, you will find the original reviewer's comments (in blue) and our answers (in black).

Major comments:

1. However, it also appears to me that the role of the mean flow in shaping the drag and its seasonal cycle could be worked out more nicely. This mostly concerns the seasonality in the southern hemisphere over the Andes, where the southernmost region is not exhibiting most intensive fluxes in winter, but in summer. If I understand them correctly, the authors argue that this is because the mean zonal wind in this region does not reverse its vertical derivative in the lower stratosphere. They claim that this is in line with linear wave dynamics. This might well be true, but I think it should be demonstrated. This could be done by single-column calculations of the drag given the diagnosed winds, with a simple wave emitted in the troposphere. Analytical arguments would be very welcome as well.

   We would like to thank the referee for pointing out that the reasoning behind the importance of the zonal wind curvature above the UTLS jet maximum and of the valve (zero wind) layer has not been clear enough. For illustrating the connection of GW breaking with the zonal background winds we repeat here the saturation flux argument developed originally by Lindzen (1981), which estimates a maximal momentum flux that can be propagated vertically without breaking for a given zonal wind profile:

   One can derive (see also e.g. Fritts, 1984) that the following equality holds for convectively unstable regions, $u' + \overline{u} = c$, where u' is the wind perturbation, $\overline{u}$ is the mean wind and c is the phase velocity. Taking the limiting case for the saturated monochromatic waves $|u'_s| = |\overline{u} - c|$, considering the polarisation relations and the equality for the vertical wavenumber $m = N/(\overline{u} - c)$ in the WKB settings, with the Brunt-Vaisala frequency N, we can write

$$w_s' = -\frac{k}{m}u_s' = \frac{k}{N}(\overline{u} - c)^2.$$

   Therefore, the maximal vertical flux component can be written as

$$\overline{\rho}\,\overline{u_s'w_s'} = \frac{1}{2}\overline{\rho}\,u_s'w_s' = -\frac{1}{2}\overline{\rho}\,\frac{k}{N}(\overline{u} - c)^3.$$

   In some form, this criterion is central to a majority of GW parameterization schemes for identifying regions of GW dissipation in the free atmosphere. It makes explicit that in the situation of the increasing wind speed in the vertical, the wind field counteracts the exponential decrease of density with height and the wave can propagate vertically without reaching the convective instability threshold.

Above the center of the UTLS jet, the wind is no longer increasing with height and the saturation flux decreases sharply indicating a potential GW dissipation region. Moreover, for orographic GWs, the region above the UTLS jet is important also because of the occurrence of a region of near zero winds (termed also the valve layer or a neck region), which serve as a critical level for OGW propagation.

It has to be noted that the curvature of the wind above the UTLS jet may be too strong for the WKB approximation to hold and the existence of breaking cannot be assessed rigorously by a simple nonlocal argument. From this perspective, we refer to our results showing a clear GWD maximum above the UTLS jet as reassuring, documenting the general validity of the simple saturation criterion.

Changes in text:

- We added reference to the saturation hypothesis after mentioning it in the context of the zonal mean profiles (L146) and we expanded the explanation:

  “It is a direct consequence of the saturation criterion employed in the parameterizations (Lindzen, 1981; Nappo, 2012) that requires momentum flux convergence in the region of negative wind shear above the jet center, where the increase of the mean wind does not balance the decrease of the density anymore.”

- To the explanation of the missing winter minimum for the southernmost Southern Andes subdomain (L202), we added the text: “Therefore, during the Southern Hemisphere winter, the vertical profile of the zonal wind above the subdomain does not suggest any regions of potential instability or critical level filtering according to the saturation hypothesis and the resolved wave field behavior confirms this.”
- To increase clarity, corresponding section in the discussion (L311) reformulated to:

  “The zonal mean drag climatology in the stratosphere exhibits a vertical distribution consistent with the saturation hypothesis, as predicted in simplified models, featuring pronounced GWD maxima above the center of the UTLS jet across all seasons in both hemispheres.”

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.

Fritts, D. C. (1984). Gravity wave saturation in the middle atmosphere: A review of theory and observations. *Reviews of Geophysics*, *22*(3), 275-308.

2. The authors decide for a rhomboidal instead of a triangular horizontal spectral filter in order to extract the gravity-wave signal from the horizontal wind. It is not quite clear to me whether this is the more appropriate approach. After all it should be the total wave number that decides, not the meridional wave number. Zonally symmetric gravity waves are not excluded by definition. Whatever, the authors state that the triangular filter gives the same results. It would be good to demonstrate this with a single figure.

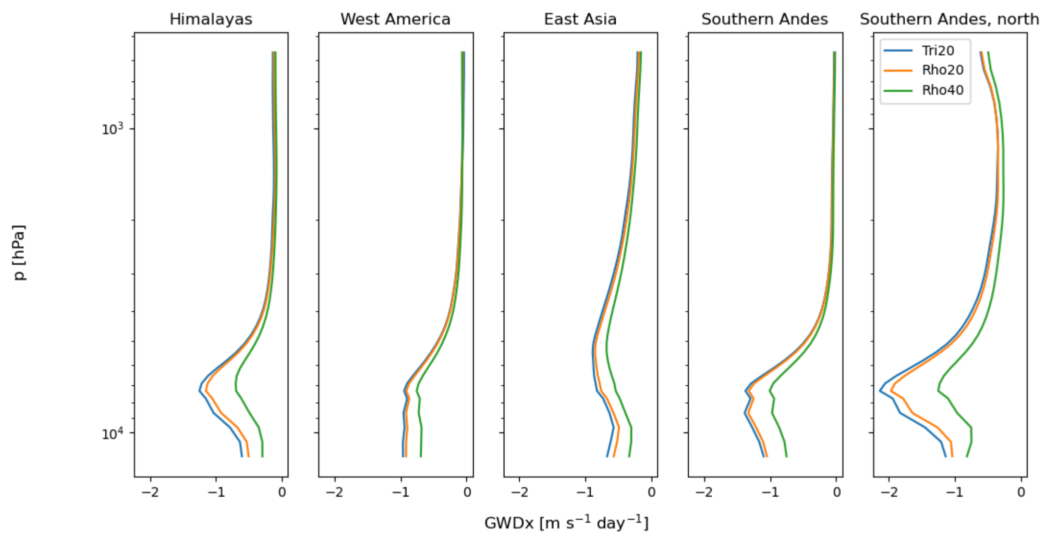
Thank you for commenting on this methodological aspect. The advantage of the rhomboidal truncation is the possibility to potentially adjust the filter according to one's needs and not-allowing modes that are longer in both zonal and meridional directions (see Fig. 2 in the manuscript). In our understanding, the triangular truncation is not something clearly related to GW. In this light, we see the rhomboidal truncation as superficial to the triangular truncation for GW separation. But, as we already stated in the manuscript, in the average sense the two methods give almost identical results and we document this by adding a figure with the comparison of the GW drag by the methods in the appendix as Fig. B2, which is also shown below.

Also, we added a description of the figure:

"Vertical profiles of GWD for their peak season evaluated using different filtering methods are displayed in Fig. B2. The triangular method provides very similar results to the rhomboidal method for equivalent cut off, with the rhomboidal method giving slightly lower amplitudes, which aligns with the comparison of the filtered fields in Fig. 3."

and a reference to the figure to Section 3.2:

"The vertical profiles do not deviate significantly from profiles that would be obtained by the filtering with triangular truncation (see Fig. B2)."



Minor comments:

l. 70 and Fig. 3: Would it not be more consistent to also filter the vertical wind? True, linear theory tells us that to leading order all vertical wind is gravity waves, but there is large-scale balanced dynamics for vertical flow (omega equation yielding the balanced ageostrophic flow) so that some large-scale vertical wind cannot really be attributed to gravity waves.

Thank you for mentioning that. We agree that filtering of the vertical wind would be an option. However, due to the large uncertainty in the filtering methods (e.g., the selection of the cutoff), the filtering would add this uncertainty to the field, and the low-passed vertical wind appears to be generally very small (see e.g., Fig. 2 in Sun at al, 2023). In any case, this should be definitely discussed more in the paper and we modified the text in the methodology (L72 in the revised manuscript). We changed

“(the vertical velocity field is assumed to be dominated by GWs in the stratosphere and hence no filtering is needed).”

-> “We do not apply any filtering of the vertical velocity field as the theory and existing literature (Sun et al., 2023) suggest the dominance of the GW perturbations to the leading order and a filtering procedure might possibly introduce some artifacts to the resulting fields.”

Sun, Y. Q., Hassanzadeh, P., Alexander, M. J., & Kruse, C. G. (2023). Quantifying 3D gravity wave drag in a library of tropical convection-permitting simulations for data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, *15*(5), e2022MS003585.

l. 77: I understand that the factor n was necessary in Prochazkova et al (2023) where WRF data had been analyzed, i.e. from a non-hydrostatic model. However, here IFS data are used, i.e. from a hydrostatic model where by definition n = 1. This should not be presented as an assumption but rather as a consequence of the model formulation. Perhaps one could even set n = 1 directly?

Thank you for the comment. The derivation of the equation is shown with the variable n for the generality. However, as mentioned in the methodology subsection, for most parts of the dataset, we use n=1 as the computation of n for upper levels significantly increases the computational costs. Regarding the 5 lower levels, the values of n actually range between 0.9 and 1.1 with the average at 1. These differences from the mean are present because although IFS is a hydrostatic model, the assimilation schemes in the ERA 5 reanalysis are slightly pushing it away from the hydrostatic equilibrium. Therefore, we prefer to leave the equations in a more general form with a variable n.

l. 167 – 168: Not quite sure whether the weakness of the zonal-mean meridional circulation is a good argument why weaker meridional drag is significant. To leading order, the time-mean residual circulation is determined by the zonal drag!

Thank you for identifying a possibly confusing formulation. The term zonal mean meridional circulation probably gave an impression that we imply some sort of nonlocal dynamical mechanism. Sorry for this. The text was modified so that it does not suggest the interpretation of means from:

“However, for context, the zonal mean meridional circulation is many times slower than the zonal mean zonal winds.”

to

“However, since the meridional wind is generally lower than the zonal wind, weaker meridional drag can have a relatively strong effect on the circulation.”

l. 174: Even without oblique propagation, in classic simple single-column gravity-wave parameterizations horizontal fluxes and their convergence are allowed. They are not considered, for simplicity, but they are there. Oblique propagation will horizontally redistribute vertical and horizontal fluxes.

Thanks for pointing out the misleading formulation. Sentence

“Our methodology allows to compute regional drag estimates, which we do in the following for selected major extratropical hotspots, taking fully into account the effects of the oblique propagation”

changed to

“Our methodology allows us to compute regional drag estimates, which we do in the following for selected major extratropical hotspots, evaluating also the effects of horizontal fluxes.”

l. 179 – 180: A diurnal signal in the gravity-wave drag could also be a signature of coupling with solar tides.

Excellent comment, thanks for the remark. We added a sentence (L187 in the revised manuscript)

“This can be caused either directly by solar heating of the air and the Earth's surface or secondarily by coupling with the solar tides.”

Fig. 5: Please mention in the caption that the spectra are for 70hPa.

Added.

Fig. 7 and 8: Replace ‘winter’ by ‘peak season’?

Thanks, it is now corrected.

l. 231: Should it not be shorter (instead of longer) orographic waves that propagate mostly in the vertical direction?

Thanks for pointing out the possibly misleading formulation. Based of your comment and the comment of the other reviewer, we changed the formulation to:

“A possible reason for this can be that the horizontal scales and geometry of the Himalaya orography together with its orientation with respect to the predominantly zonal background flow favor sourcing of longer orographic GW modes that propagate vertically more efficiently compared to longer GWs for other hotspots.”