We would like to thank Referee #2 for appreciating our efforts and for the very helpful comments and suggestions that will definitively improve the manuscript!

Please find below our point-by-point reply to the reviewer concerns. Comments by Reviewer #2 are given in red, our reply is given in black, and changes in the manuscript are indicated in blue.

Reply to the Main Concerns by Reviewer # 2:

(Main Concern 1a:) l.270-272 It is unclear whether a purely zonal approach of assuming k>20 is adequate to extract model-resolved GWs. This would only be the case if zonal propagation of GWs is dominant in the tropics. Otherwise, a more sophisticated approach as proposed by Watanabe et al., 2008, or Becker et al., 2018 would be needed. The contribution of resolved GWs should be shown for more information.

Of course, some care has to be taken whether the major part of the model-inherent gravity wave (GW) drag is from model-resolved GWs, and whether methods as suggested by Reviewer 2 should be applied.

As suggested, we will include the following two references in the revised manuscript as a guidance for readers who want to analyze model-resolved GWs in detail:

Watanabe, S., Kawatani, Y., Tomikawa, Y., Miyazaki, K., Takahashi, M., and Sato, K.: General aspects of a T213L256 middle atmosphere general circulation model, J. Geophys. Res., 113, D12110, doi:10.1029/2008JD010026, 2008.

Becker, E., and Vadas, S. L.: Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model, J. Geophys. Res.: Atmospheres, 123, 2605-2627, https://doi.org/10.1002/2017JD027460, 2018.

However, there are three reasons why we think that a more sophisticated approach is not required in our paper:

- The GW drag of model-resolved waves with k>21 is only a minor part of the GW drag proxy derived in our paper. This is the case because the spatial resolution of the reanalyses is relatively coarse. As a consequence, they resolve only a limited part of the GW spectrum. This can be seen below from Figs. 1 and 2 which show our estimates of the total and the resolved GW drag. Uncertainties in the resolved GW drag will therefore not much affect our estimates of total GW drag. Introducing a limit of k>21 is somewhat arbitrary, anyhow.
- In the tropics, zonal propagation of GWs should be dominant because zonal winds are usually stronger than meridional winds, and GWs propagating opposite to the background wind can attain larger amplitudes because the intrinsic phase speed of these waves, and thus their saturation amplitudes, are increased. Therefore, the purely zonal analysis should be justified even for the part of the wave spectrum that is resolved in the models (in the tropics, many other studies also use this kind of zonal approach — for example, even the KANTO model group did so in the follow-up paper Kawatani et al., 2010, with S. Watanabe as one of the coauthors).
- The validity of a zonal-only approach is also supported by the fact that at most altitudes all our GW drag proxies (reanalyses and SABER) show strong correlation with either the zonal wind, or its vertical gradient.

These three points will be mentioned in the revised manuscript in the new Sect. 5.2.1, In addition, we will include in the paper — among other new figures — the figure showing the resolved GW drag for the four reanalyses.

A brief discussion of Figs. 1 and 2:

As explained in the main paper, the total zonal GW drag $\overline{X}_{GW}$ in models consists of three different contributions and can be written as follows:

$$\overline{X}_{GW} = \overline{X}_{res}(k > 20) + \overline{X}_{param} + \overline{X}_{imbalance} \tag{1}$$

with $\overline{X}_{res}(k > 20)$ the GW drag due to model-resolved waves with zonal wavenumbers $k > 20$, $\overline{X}_{param}$ the parameterized zonal GW drag, and $\overline{X}_{imbalance}$ the "residual", or remaining imbalance that is caused, for example, by data assimilation.

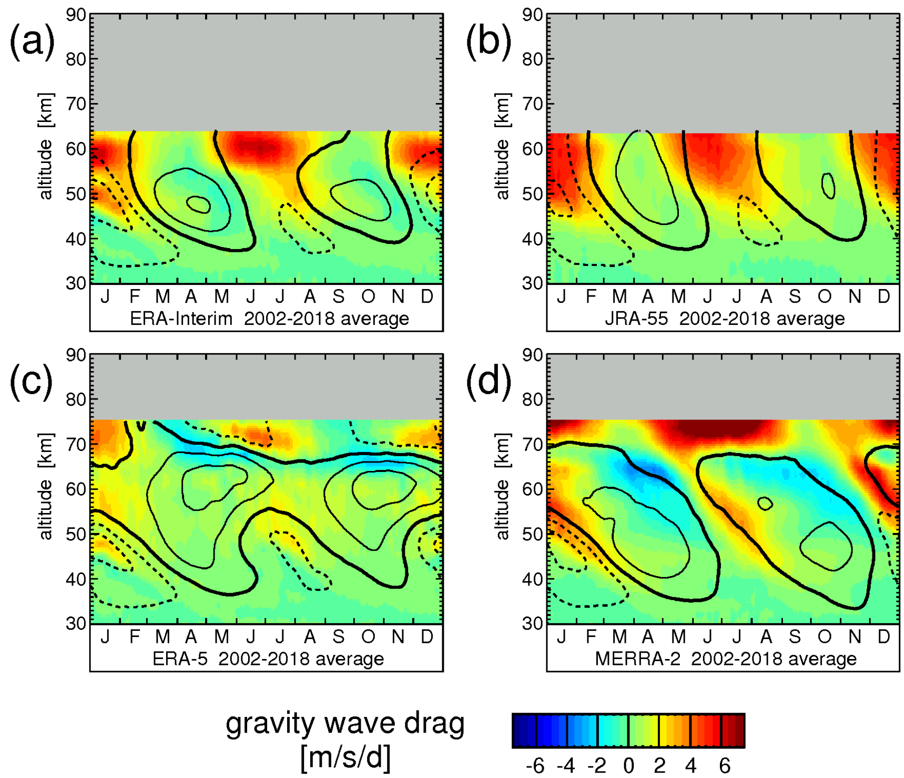

Figure 1: Total zonal GW drag $\overline{X}_{GW}$ for **(a)** ERA-Interim, **(b)** JRA-55, **(c)** ERA-5, and **(d)** MERRA-2. Overplotted are contour lines of the corresponding zonal average zonal winds for the respective reanalysis dataset. Contour line interval is 20 m/s. The zero wind line is highlighted in bold solid, and westward (eastward) winds are indicated by dashed (solid) contour lines.

As can be seen from Figs. 1 and 2 of this reply, the resolved GW drag is negligible in ERA-Interim, JRA-55, and MERRA-2. (Please note that in Fig. 1 the range of the color scale is $\pm$7.5 m/s/d, while it is only $\pm$0.25 m/s/d in Fig. 2a, 2b, and 2d, and only $\pm$1.25 m/s/d in Fig. 2c.) Only for ERA-5 below 55 km $\overline{X}_{res}(k > 20)$ sometimes contributes as much as about 50% to $\overline{X}_{GW}$. In both $\overline{X}_{GW}$ and $\overline{X}_{res}(k > 20)$ eastward GW drag is stronger than westward GW drag in the upper stratosphere and lower mesosphere, which is a consequence of the QBO wave filtering in the stratosphere below.

References:

Kawatani, Y., Sato, K., Dunkerton, T. J., Watanabe, S., Miyahara, S., and Takahashi, M.: The roles of equatorial trapped waves and internal inertia gravity waves in driving the quasi-biennial oscillation. Part I: Zonal mean wave forcing, J. Atmos. Sci., 67, 963-980, doi:10.1175/2009JAS3222.1, 2010.

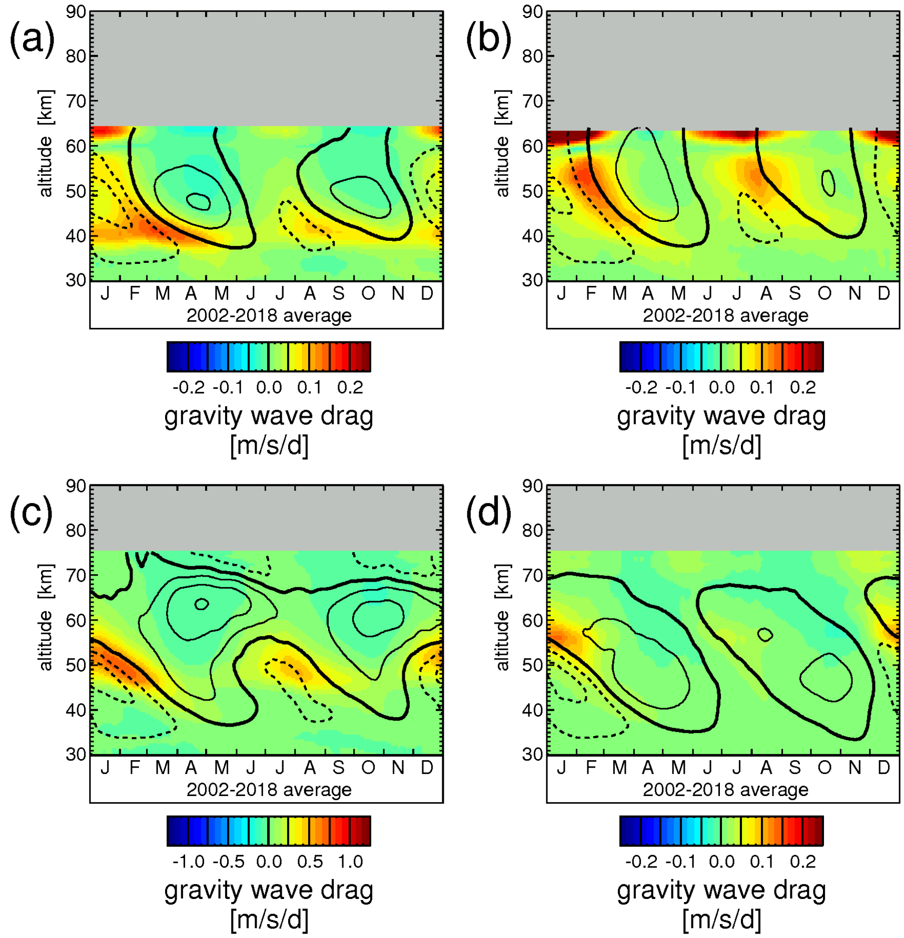

Figure 2: Resolved GW drag $\overline{X}_{res}(k > 20)$ for **(a)** ERA-Interim, **(b)** JRA-55, **(c)** ERA-5, and **(d)** MERRA-2. Overplotted are contour lines of the corresponding zonal average zonal winds for the respective reanalysis dataset. Contour line interval is 20 m/s. The zero wind line is highlighted in bold solid, and westward (eastward) winds are indicated by dashed (solid) contour lines.

(Main Concern 1b) Assuming zonal propagation of GWs may contradict the flux estimate of SABER, in which the along-track wavelength is regarded as the horizontal wavelength. If zonal propagation of GWs is dominant in the tropics, an error of flux estimate could be large. Please mention the error of flux estimate in more detail.

This concern is the same as **Main Concern 3a** by Reviewer 1. Therefore our corresponding reply is repeated here:

As recommended, We will add more discussion about the shortcoming of using SABER along-track GW horizontal wavenumbers, instead of "true" GW horizontal wavenumbers.

Generally, the use of along-track GW horizontal wavenumbers as a proxy for the true GW horizontal wavenumbers will lead to a low-bias of SABER momentum fluxes (the momentum flux is proportional to the horizontal wavenumber) because the along-track GW horizontal wavenumber will always underestimate the true horizontal wavenumber.
The AIRS satellite instrument has a similar orbit geometry. Because AIRS provides 3D temperature observations, it is possible to determine from AIRS observations true GW horizontal wavenumbers, as well as along-track GW horizontal wavenumbers. This opportunity has been taken by Ern et al. (2017) to compare true and along-track GW horizontal wavenumbers: AIRS observations indicate an underestimation of the along-track wavenumber (corresponding to an underestimation of momentum fluxes) by a factor between 1.5 and somewhat above

2.
In addition, for SABER there will be aliasing effects (undersampling of observed GWs) and effects of the instrument sensitivity function of limb sounding satellite instruments (cf. Ern et al., 2018), which should both lead to an even stronger underestimation of GW momentum fluxes. Therefore the error of SABER GW momentum fluxes should be at least a factor of two, and momentum fluxes are likely strongly underestimated.

This detailed discussion will be included in the revised manuscript in the newly introduced Sect. 6.1.1.

References:

Ern, M., Hoffmann, L., and Preusse, P.: Directional gravity wave momentum fluxes in the stratosphere derived from high-resolution AIRS temperature data, Geophys. Res. Lett., 44, 475-485, doi:10.1002/2016GL072007, 2017.

Ern, M., Trinh, Q. T., Preusse, P., Gille, J. C., Mlynczak, M. G., Russell III, J. M., and Riese, M.: GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings, Earth Syst. Sci. Data, 10, 857-892, doi:10.5194/essd-10-857-2018, 2018.

(Main Concern 2) Figures for respective years are shown throughout this paper. However, there is little discussion of interannual variation. It is better to reduce unnecessary figures and make them larger and easier to see.

This comment is similar to **Main Concern 1a** by Reviewer 1.

We will move the single-year figures into the supplemental material and only keep the multi-year averages in the main paper. Further, we will merge figures in the main paper.
In addition, in former Figs. 17 and 18 we will omit the column showing the zonal wind, because it is redundant. Further, we will move three rows of former Fig. 17 into the Supplement.

(Main Concern 3) It is mentioned that the effect of tide is large for satellite data above 80 km and contaminate the GW contribution to the SAO driving throughout this paper. Although I understand that it is the signature of the interaction between the tide and GWs and important, it looks far from the primary purpose of this paper. I recommend moving it to another paper.

Because the local time of SABER observations is continuously changing, a brief discussion of local-time effects, including tides, cannot be avoided. Further, the correlation between SABER vertical gradients of absolute momentum flux and the background winds has very different characteristics in different altitude ranges.
For a full understanding, these differences need to be explained, even if these differences are caused by the QBO or tides, and not directly by the SAO. Otherwise, readers will start to question our methods.

Therefore we still mention the effect of tides, but significantly shorten the discussion — particularly in Sects. 7.1.3 and 7.2.2.

Reply to the Minor Comments by Reviewer # 2:

(Minor Comment 1) The SPARC climatology is regarded "true" and the difference from it is expressed as "bias" throughout this paper. However, it may be better to change the expression because it could be caused by the difference in time resolution and interannual variation as shown in L. 225-228. At least, it is better to compare monthly averages between SPARC climatology and reanalysis/satellite data, which clarify the effect of time resolution.

As recommended, the word "bias" has been removed in connection with the SPARC climatology.

Just to clarify: we do not regard the SPARC climatology as "true"!
We show the SPARC climatology just for comparison because our general knowledge of the SAO is very poor. In spite of its shortcomings, this climatology gives some information on the basic structure of the SAO, which is helpful to guide the discussion throughout the paper. This might have caused the impression we would consider the SPARC climatology to be the "truth".

In order to clearly emphasize that the SPARC climatology is not the "truth", we feel that more discussion on the weaknesses of this climatology is needed. A number of weaknesses of the URAP/SPARC climatology is listed below, and a brief version of this listing will be included as additional discussion in the revised paper in Sect. 4.1.

Some potential limitations of the URAP/SPARC climatology:

- The SPARC climatology is based on the URAP zonal wind climatology, which uses direct wind observations by the HRDI instrument. HRDI observes the Doppler shift of spectral lines from satellite. An inherent problem of this method is that the zero-wind is not clearly defined (e.g., Hays et al., 1993; Baron et al., ACP, 2013), which involves assumptions and may introduce biases.

- HRDI uses only daytime data and covers only 50 to 75% of local times in the mesosphere (cf. Fig. 2 in Swinbank et al. 2003) which can introduce biases, although a correction of tidal effects was attempted.

- URAP uses HRDI observations from the 7 years 1992-1998, but HRDI temporal coverage is reduced to much less than ~50% after mid 1996. Strictly speaking, this means that the SPARC climatology contains only a short period of 4.5 years of quasi-continuous HRDI observations. Only this period should be more reliable because directly guided by HRDI observations. Consequently, interannual variability will still have strong effect on the monthly averages of the SPARC climatology.

- Spatial and temporal gaps in the HRDI wind observations are filled by a climatology, or by interpolation based on model data (UKMO analyses) and balanced winds (cf. Randel et al., 2002). This combination of different data sets, as well as the data processing, may introduce certain biases. Multi-year averaging over a longer dataset with good temporal resolution (like in our paper) was not possible for the SPARC climatology due to the shortness and further shortcomings of the datasets.

- There is a HRDI data gap (no observations) centered around 0.3 hPa (55km, cf. Swinbank 2003, their p.2, and their Figs. 3a and 6) that needs to be filled by climatology, model data, or interpolation, and could introduce biases. This makes the continuously eastward directed winds at 60km in SPARC questionable. Particularly, because there is strong interannual variability, including also periods of **westward** directed winds, at that altitude in all our datasets. Therefore, already in the ACPD paper version we dedicated some discussion to the question whether this feature in the URAP/SPARC climatology could be reliable.

Given these facts, it should be clear that the SPARC climatology can contain biases and should not considered to be the "truth".

Since the SPARC climatology is not considered to be a reference, or the "truth", it does not make much sense to, for example, reduce the time resolution of our better resolved datasets to that of the SPARC climatology. Further, we hope that the better time resolution of our dataset may be considered helpful by others.

References:

Baron, P., Murtagh, D. P., Urban, J., Sagawa, H., Ochiai, S., Kasai, Y., Kikuchi, K., Khosrawi, K., Körnich, H., Mizobuchi, S., Sagi, K., and Yasui, M.: Observation of horizontal winds in the middle-atmosphere between 30°S and 55°N during the northern winter 2009–2010, Atmos. Chem. Phys., 13, 6049–6064, doi:10.5194/acp-13-6049-2013, 2013.

Hays, P. B., Arbreu, V. J., Dobbs, M. E., Gell, D. A., Grassl, H. J., and Skinner, W. R.: The high-resolution Doppler imager on the Upper Atmosphere Research Satellite, J. Geophys. Res., 98, 10713–10723, 1993.

Randel, W., Chanin, M.–L., Michaut, C., and the SPARC Reference Climatology Group: SPARC intercomparison of middle atmosphere climatologies, WCRP 116, WMO/TD No. 1142, SPARC report No. 3, 2002.

Swinbank, R., and Ortland, D. A.: Compilation of wind data for the UARS Reference Atmosphere Project, J. Geophys. Res., 108, 4615, doi:10.1029/2002JD003135, 2003.

(Minor Comment 2) It seems unnecessary to discuss the correlation in QBO and sponge layers in this paper.

The Rayleigh drag exerted in the sponge layers is one of the contributions of the GW drag proxy. For JRA-55 and ERA-Interim, this is even one of the main contributions. In the discussion in Sect. 8 we point out that this contribution is not considered to be very realistic. We feel that this discussion is needed because Reviewer 1 in his **Major Concern 2** was worried that the residual drag from reanalyses would not always be representative of GW drag. For this reason, we keep this discussion in order not to appear to be too uncritical about our results. In addition, the altitude where the negative correlation between the residual drag and the zonal wind strengthens gives valuable information above which altitude the residual drag becomes increasingly unrealistic.

Therefore we decided to keep the discussion about Rayleigh drag in the sponge layers.

The biennial variations seen in our correlation analysis are a striking feature that needs to be explained. Without an explanation, readers would think this correlation would be an artifact, and doubt our methods. Overall, the discussion of the QBO-correlations does not take much space, anyhow.

Therefore, for completeness, we decided to keep the discussion related to the QBO.