Answer to anonymous Reviewer #2

We thank the reviewer for her/his comments and suggestions. We answer point by point in the following with the reviewer's comments added in *red/italics*. Text added to the revised version of the manuscript is included here in *blue/italics*.

5

General comments

The results of this study are in line with experiments carried by other groups, now and in years prior. As a reviewer, and a modeler myself, I appreciate these types of studies that are difficult to publish. There is really no science in them, they are an engineering exercise, but one that is necessary to document. With this preamble in mind, I find the manuscript too long and

10 *delving into details that are not interesting to the general reader. I do think the study needs to be published, but it also needs some thinning, which I am suggesting in the detailed comments below.*

We thank the reviewer for the encouraging remark. We have shortened the revised manuscript by moving some of the Figures to a supplement.

15

From a general point of view, one difficulty in tuning the gravity wave drag parameterization is the coupling with interactive chemistry. When comprehensive models like WACCM were developed, it was often found that the while a set of tuning parameters reproduced a desirable climatology, the timing of the SH warming (i.e., reversal of the winds) was unrealistic and ozone chemistry was adversely affected. In this sense, what I find missing is a discussion of the timing of the springtime warming

20 *in the lower stratosphere, like Figure 10 in doi: 10.1175/2009JAS3112.1. Moreover, the Richter et al. paper is sorely missing from the present discussion; consider also discussing Sassi et al. (doi:10.1029/2003JD004434).*

As documented in the manuscript, the ozone is prescribed from a monthly, zonally averaged climatology derived from a HAMMONIA simulation, using MOZART chemistry. The effect of a self-reinforcing temperature bias due to increased ozone

25 depletion in the presence of an initial negative temperature anomaly does not occur in our simulations. However, this is an important issue and has to be considered when using UA-ICON with interactive chemistry, which is planned in the near future. We have added and briefly discussed Richter et al. (2010) and have added Sassi et al. (2004) when mentioning the influence of ENSO on the NH winter variability in the middle atmosphere in the revised manuscript.
The transition of the monthly mean zonal mean zonal wind from the wintertime westward-directed (easterly) flow to the sum-

30 mertime eastward-directed (westerly) flow in the SH is shown in a supplementary Fig. S4. All data of the UA-ICON simulations is interpolated to constant pressure levels for this analysis.

We have added the following sentences to the discussion section:

35 *Whereas our tuning attempt for the NGWD focused on changing the saturation conditions for NGWs and the source strength, by varying parameters without latitudinal variation, Richter et al. (2010) tuned the NGWD in WACCM3 by introducing a source-oriented GW parametrization, accounting for variability in convective activity and frontal systems. One detail of their tuning success concerns the SH springtime transition of the zonal wind, which shifts to an earlier date. The transition to an easterly flow in $60\,°S$ at 10-hPa occurs in early November for ERA-5 reanalysis data, which the UA-ICON(ECHAM) simulation*

40 *matches perfectly. The UA-ICON(NWP) simulations show the transition in late November, which is only slightly shifted to a later date by the GWD tuning (Fig. 4, supplement), in comparison with the WACCM experiments shown in (Fig.10 of Richter et al., 2010) which appear at end of December.*

Detailed comments

Abstract, line 13: This concept is repeated many time throughout the paper and I might have missed some important details. My reservation about this statement is that the forcing (SST, etc.) is climatological: how do you expect to represent realistically the occurrence of phenomena that are so intimately associated with tropospheric weather, such as wave-2 forcing?

Despite the fact of constant climatological forcing in the troposphere, UA-ICON shows considerable variability in all SSW statistics, including the wave-2 preconditioning. However, it is rare, but not completely absent. We find it still worth mentioning, that the model with a constant forcing has a dynamically self-generated bias in some aspects.

Figure 1. Are these the results of one model simulation or of averaging (climatology) of several instances? If the latter, then you should be able to say something about the range of uncertainty; if the former, I would caution overinterpreting these results. In the following, when difference from observed climatologies are reported, the reader needs to know how significant they are.

The results are based on only one model simulation in each panel. However, in the case of the UA-ICON (NWPD) and UA-ICON(NWP) simulations 60 years are included to characterize the mean state, which we believe is sufficient in this context. The largest uncertainty for the climatology arises from the effects of major sudden stratospheric warmings (SSWs). We show the large variability in the NH stratospheric polar vortex in terms of SSWs per decade (Fig. 12(a), initial submission; Fig. 9(a), revised submission), confirming that periods of high and low SSW frequency are included, leading to a robust climatological average. We have tested the climatological averages for statistical significance by comparing the UA-ICON seasonal averages to ERA-5 (temperature and zonal wind) and SABER (temperature). This reveals that in large areas significant differences between ERA-5/SABER and the UA-ICON simulations exist. The results of the significance test are included in the Figure 1 of the revised submission showing the seasonal mean temperatures. The results for the seasonal differences in zonal mean zonal wind to ERA-5 are not included in the Figure 2 (revised submission), as we compare to the URAP climatology, for which no individual monthly means are available. We have added the following statement about the significance of the differences to observations/reanalyses to Section 3 of the manuscript.

However, overall, the differences in zonal mean temperature between all UA-ICON simulations and SABER are significant in most areas, as indicated by the non-hatched areas in Fig. 2. The same holds for the zonal mean wind compared to ERA-5, where all UA-ICON simulations show significant differences in most areas up to 80 km altitude (not shown).

Line 171. This is odd sequencing. You cannot show a result and say we will explain later the methodology: the reader needs to see first the methodology and then we can make an informed decision on the validity of the results. Otherwise, the study is reduced to a beauty contest. I suggest moving section 4 ahead of section 3 and streamline the discussion. This will also help the flow of the paper, which instead appears is interrupted.

We followed the reviewer's suggestion and shortened Section 4 by moving the vertical profiles and the probability density functions to the supplement. Only the Figure showing the zonal-mean anomalies stays in the manuscript. However, although we can understand his/her objection, we would like to stay with the current sequence of the sections, as it allows us to introduce the motivation for tuning the GWD parameterizations first. After specifying the tuned GWDs, we show their spatial structure (Fig 3 and 4) and refer back to T and U (Fig 1 and 2) for comparison with the former structures. So we keep the zonal means from simulations and observations in one compact figure. The details of tuning experiments are thus removed from the main text to the supplement. Accordingly, the shortened Table 2 moved from Section 4 to Section 2.1 (Model and Setup), as well as the description of the parameters for the tuning of the gravity wave parameterizations. We now hope the reader will easily follow the content of Section 3.

We followed the reviewer's suggestion and moved Figures 5, 7, and 8 to an additional supplement, as it also allows to shorten the manuscript.

Figure 5. Are these hemispheric averages? It would be more informative to exclude the tropics and average only poleward of 30 degrees in each hemisphere. More generally, it is hard to follow each single line in Figure 5. The vertical line plots are useful in the process of tuning a parameterization, but as they are, they become confusing, hard to read and not very useful in a scientific paper.

The latitudinal range of the averages is varying as indicated on each of the sub-panels of Figure 5. The average over the complete hemispheres is only applied for V*. In order to make this section better readable we shifted Fig. 5 with detailed tuning information into the supplement.

Moreover, since the arguments are clearly developed referring to a global circulation, the reader's mind goes to the zonal mean plots. It would be much more useful to describe the un–tuned simulations in the previous section and introduce the tuned GWD schemes in this section.

To follow the reviewer's suggestion would require to split Figures 1 and 2 and has the potential to enlarge the manuscript even more, as the discussion of both types of simulations at once allows for a more compact presentation.

Figure 7. I suggest marking the ERA-5 and SABER PDF with thick bold lines. This is a very busy figure; it will be hard to read for many people: I cannot tell the light green from the magenta. I recommend reducing the number of cases reported in these panels (same for Figure 8) to make reading easier. An additional figure with all the cases can be added as a supplemental image.

The Figures 7 and 8 are now in the supplement, however, it is technically not possible to emphasize the ERA-5 and SABER PDFs with matplotlib.hist(histtype='step').

Figure 8. As noted above, these figures are too busy. Also, I recommend discussing the lower stratospheric temperature, important for ozone depletion.

As the Figure 8 is now included only in the supplement, therefore we have not added an extra discussion of the lower stratospheric temperature.

Figure 10. Very difficult to tell colors apart; use different type of line style.

We changed the line style for the UA-ICON(NWP) simulation to dashed, to improve comparability.

Line 320. Aren't these biases the result of excessive dynamical heating?

Dynamical heating may contribute to these biases in the stratosphere. For example, the larger temperature anomaly of the tuned UA-ICON(NWP) simulation in early winter may be related to the stronger orographic GWD. However, we see a larger influence of dynamical heating on the temperature near the stratopause by increasing non-orographic GWD.

We have added the following to the manuscript:

The warm bias of the tuned UA-ICON(NWP) simulation is slightly larger in early winter. This larger dynamic heating is probably due to a stronger orographic GWD.

Fig. 11e,f. Note the persistent divergence at higher altitudes. Why?

The divergence of the Eliassen-Palm flux (EPF) is diagnosing the drag of the resolved waves in the model. Before and after the SSWs a persistent EPF divergence near 110 km is present. This is related to eastward propagating resolved waves in the middle atmosphere up to the lower thermosphere. During the SSWs, these eastward propagating resolved waves already dissipate in the mesosphere. We have added the following to the manuscript:

Westward directed (negative) planetary wave drag up to 100km has been found in several model simulations (Zülicke and Becker, 2013; Limpasuvan et al., 2016; Okui et al., 2021) as a result of baroclinic/barotropic instability (Sato and Nomoto, 2015). The persistent positive $\nabla \cdot \boldsymbol{F}$ (divergence) in the lower thermosphere near 110 km before and after the SSWs is related to eastward propagating resolved waves. The reason for the appearance in a focussed layer around 110 km in UA-ICON is most likely in the particular vertical structure of incorporated WM96 NGWD parameterization.

Fig. 11, the V* and W* panels. Are these necessary? Again, this looks like more a dump from a PhD thesis; these panels do not add anything to the paper, except space and length.

The intention of showing the residual circulation is to control the consistency of response, in particular the link between dynamic warming and temperature patterns. Especially the W* is interesting, as the vertical velocity is a prognostic variable in the ICON model. Most GCMs/CCMs including the MLT are hydrostatic and therefore diagnose the vertical velocity. Figure 11 is included to demonstrate the ability of UA-ICON to plausibly model the physical mechanism of SSWs, which also includes a consistent alteration of the mean meridional circulation. We therefore want to stay with Figure 11 as it is, including V* and W*.

Figure 12. What are the empty bars? I know it is described in the text, but you don't want the reader to go hunting for that information. Repeat in the caption.

The missing information is now included in the caption of Figure 12.

Figure 12: (a) Time series of the 10-year moving averaged major sudden stratospheric warming (SSW) frequency in events per decade within the NH winter season (November–March) for UA-ICON simulations and reanalyses; the shading indicates the 95% confidence interval. (b) Bar chart of SSWs per decade with error bars indicating the 95% confidence interval; (c) the monthly distribution of SSWs per decade. The statistics are based on a period of 20 years for UA-ICON(ECHAM), and 60 years for UA-ICON(NWPD), UA-ICON(NWP), NCEP1, ERA-5 (1963-2022), and 23 years for MERRA2 (1980-2022). The open bars in panels (b) and (c) indicate the averaged accumulated easterlies ($|\Delta U|$) per event.

Table 5. I feel these statistics are an overkill. If the goal is to evaluate the climatological behavior, one doesn't need much beyond Figure 12. If the goal is instead to accurately represent individual SSW, then the whole premise is wrong, since, as I stated above, my understanding is that the boundary forcing is climatological.

We are convinced that it is worth including the statistics in Table 5. The reviewer is right to mention that the statistics derived from the UA-ICON simulations are not directly comparable with statistics derived from reanalyses, as only climatological con-

ditions are prescribed and an important phenomenon as the QBO is not included. As already mentioned previously, despite this constant forcing considerable variability in all SSW statistics is present and these statistics can be used for comparing the UA-ICON simulations with different physical packages and related tuning for the gravity wave parametrizations. In addition, we benchmark here the mesospheric cooling response to stratospheric warming which is an important part for further MLT studies.

Line 495. I understand the methodology is borrowed from the YS24 paper but a little more information is necessary here. Is the "migrating" component representative of a 60-day composite, or longer?

The reviewer is right that the retrieval of tides from SABER data by YS24 involved a 60-day moving window. This information has been included in the revised manuscript.

It is noted that the tidal analysis of SABER data involves a 60-day window, which might lead to an underestimation of tidal amplitudes.

Line 505. Need to cite Pedatella et al. 2014 (doi:10.1002/2013JA019421)

The paragraph starting at l.505 discusses climatologies of tides and their seasonal dependence. The study by Pedatella et al. (2014), on the other hand, focuses on the response of tides to SSWs, and thus it does not fit well in this paragraph. Since the Pedatella et al. (2014) reference is already cited in the paragraph starting at l.512, which concentrates on tidal changes during SSWs, we believe no changes to the text are necessary in response to this comment.

References

Limpasuvan, V., Orsolini, Y. J., Chandran, A., Garcia, R. R., and Smith, A. K.: On the composite response of the MLT to major sudden stratospheric warming events with elevated stratopause, J. Geophys. Res.: Atmos., 121, 4518–4537, https://doi.org/10.1002/2015JD024401, 2016.

Okui, H., Sato, K., Koshin, D., and Watanabe, S.: Formation of a Mesospheric Inversion Layer and the Subsequent Elevated Stratopause Associated With the Major Stratospheric Sudden Warming in 2018/19, J. Geophys. Res.: Atmos., 126, 1–22, https://doi.org/10.1029/2021JD034681, 2021.

Pedatella, N. M., Fuller-Rowell, T., Wang, H., Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Liu, H.-L., Sassi, F., Schmidt, H., Matthias, V., and Goncharenko, L.: The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models, Journal of Geophysical Research: Space Physics, 119, 1306–1324, https://doi.org/https://doi.org/10.1002/2013JA019421, 2014.

Richter, J. H., Sassi, F., and Garcia, R. R.: Toward a physically based gravity wave source parameterization in a general circulation model, J. Atmos. Sci., 67, 136–156, https://doi.org/10.1175/2009JAS3112.1, 2010.

Sassi, F., Kinnison, D., Boville, B. A., Garcia, R. R., and Roble, R.: Effect of El Niño–Southern Oscillation on the dynamical, thermal, and chemical structure of the middle atmosphere, J. Geophys. Res.: Atmos., 109, https://doi.org/10.1029/2003JD004434, 2004.

Sato, K. and Nomoto, M.: Gravity Wave–Induced Anomalous Potential Vorticity Gradient Generating Planetary Waves in the Winter Mesosphere, J. Atmos. Sci., 72, 3609–3624, https://doi.org/10.1175/JAS-D-15-0046.1, 2015.

Zülicke, C. and Becker, E.: The structure of the mesosphere during sudden stratospheric warmings in a global circulation model, J. Geophys. Res.: Atmos., 118, 2255–2271, https://doi.org/10.1002/jgrd.50219, 2013.