Author response to referee comments on Chrysanthou et al. "The effect of atmospheric nudging on the stratospheric residual circulation in chemistry-climate models" submitted to ACPD

Reply to Anonymous Referee #1

Recommendation: Publication after minor revision

The paper is very well organised and written. The topic discussed here − the effect of nudging CCMs on the stratospheric residual circulation − is of very high relevance, because nudged or specified dynamics CCM simulations are a common tool for analysing and interpreting observed changes of the stratospheric trace gas composition (e.g. Froidevaux et al. 2019). This work will most likely trigger many more studies related to the problems induced by driving CCMs (or more generally GCMs) with specified dynamics derived from meteorological reanalysis However, some further steps to disentangle the main source(s) of the problem could already be done in this paper or the author could at least explain, why this could not be done here (see comments below). The paper should be submitted after addressing the comments below.

Thank you for your positive comments and suggestions for improving the paper. We reply to the specific points raised below in red.

General comments:

The topic is important, and it is time to carefully analyse the effect of nudging CCMs towards meteorological reanalysis data (RA) on the simulated stratospheric residual circulation and consequently also on the transport of chemical species. The latter is not the focus of this paper here, but the results in this paper will hopefully stimulate further studies on the impact of nudging on stratospheric tracer transport and composition, e.g. on CCM-SD simulations of the recent lower stratospheric ozone trend reported by Ball et al. (2018).

I highly appreciate the topic and the great effort made by the authors and I agree in general to the conclusions derived from the results presented in this very valuable multi-model analysis. However, the manuscript should and could be focused more on the main topic and the new aspect that is clearly announced in the title: "The effect of nudging on the stratospheric …". The differences between specified dynamic (SD) and free-running (FR) simulations is to my point of view important, but only one (minor) aspect of this paper and should be contrasted with the differences between the SD simulations and the reanalysis datasets (RA) for the individual measures of the stratospheric residual circulation. This would describe the quantitative effect of nudging CCMs.

As outlined above, the paper is to my opinion focused too much on the difference between FR and SD simulations – this becomes obvious in most Figures where only the differences of REF-C1SD-REF-C1 (SD-FR) are shown and the differences REF-C1SD-Reanalysis (SD-RA) is missing. The systematic analysis of SD-RA and SD-FR for all the metrics of stratospheric residual circulation would help to get a more quantitative view and help to better understand the effect of nudging on the stratospheric residual circulation. It is clear that the underlying mechanisms that trigger the observed discrepancies in the stratospheric residual circulation between nudged CCMs and reanalysis is not in the scope of this paper, but the differences SD-RA should be shown in the same way as the differences SD-FR, so that the effect of nudging – reflected by the differences to the reanalysis – could be contrasted and discussed more quantitatively compared to the differences between specified dynamics and free-running simulations.

Frankly spoken, it would be better to streamline this paper and to concentrate only on the CCMs that have carried out SD and FR simulations in a consistent way. To my opinion, the other non-nudged models could not add any information to the main topic of the paper – the effect of nudging CCMs. The consequence would be to skip 4 out of 9 models or 4 out of 10 simulations which do not provide SD simulations. Including 4 non-nudged CCMs to the multi- model-mean (MMM) of the FR simulations is not really helpful to analyse the effect of CCM nudging. In the best case, the results are not blurred by these models, but to my opinion the analysis of SD-FR vs. SD-RA for the 6 simulations providing all relevant information would help to get a much more stringent and systematic analysis of the role and impact of nudging CCMs on their stratospheric residual circulation. This paper would gain, if the also very interesting inter-model stratospheric residual circulation comparison of all participating CCMI models would be done in a separate paper.

In response to this comment we have made several changes to the manuscript. As suggested, we have removed the models that did not perform the SD simulation from the main body of the paper (GEOSCCM, NIWA-UKCA and ULAQ-CCM). For completeness a subset of diagnostics is shown for those models in the supplementary material, but the main text now concentrates on the 7 models that performed both free running and nudged simulations (nb: since the initial submission we have now obtained SD data for SOCOL3).

Additionally, in line with the reviewer's recommendation, we now include additional figures showing differences between the SD experiments and the reanalysis (RA) datasets. Specifically, we have moved Figure S2 from the supplement into the main text and have added reanalysis data (now Figure 2), and have added additional panels to Figures 3 and 4 showing the SD - RA differences for comparison with the SD - REF-C1 results. Additionally, the supplement now includes figures showing differences in the climatological wbar_star between each SD model and the respective reanalysis it was nudged towards (Figure S1). We also show the 70 hPa wbar_star differences in the climatological annual cycle between each SD model and the respective reanalysis it was nudged towards (Figure S3). The MLR analysis presented in section 3.5 and the trend analysis in section 3.6 was also performed for the RA datasets and the results are discussed in the main text with the relevant figures also presented in the Supplement (Figures S4 and S7). We hope that the reviewer finds that these changes to the manuscript address the comment about focusing the paper on the SD simulations and comparing them more closely with the RA data.

Specific comments:

L.79: "… from which age-of-air (AoA) can be estimated (Waugh and Hall, 2002)." For clarity, I would suggest to add: …, however AoA represents the combined effects of residual circulation and mixing. Waugh and Hall (2002) is a review paper and not the original reference for AoA from tracer measurements and the credit should be given to the original work. To my knowledge, the first stratospheric AoA estimates from tracer measurements has been reported by Schmidt and Khedim (1991) and the concept of AoA was first applied to the stratosphere by Kida (1983).

Thank you for highlighting the original work that defined the AoA concept. We have added the references to Kida (1983) and Schmidt and Khedim (1991). Additionally, we have added a word of caution regarding what AoA represents :

"tracer measurements from which age-of-air (AoA) can be estimated (Kida, 1983; Schmidt and Khedim (1991); Waugh and Hall, 2002). Note that AoA represents the combined effects of both the residual circulation and mixing processes."

L.79-80: Engel et al. (2009) found no decrease in AoA, but the observed increase of AoA was statistically non-significant.

Thank you for pointing this out. We have now corrected this sentence:

"Engel et al. (2009) and Stiller et al. (2012) using observations from tracers found a statistically non-significant increase in AoA in the middle stratosphere at northern midlatitudes"

L.191: The reference for the citation "(Rosenlof, 1995)" is missing.

Thank you for spotting this. We have now added the reference.

L. 276-284: To my opinion, the Supplemental Figure 2 should be moved to the manuscript, because the topic is discussed here and the TA-latitudes are a simple measure and very indicative for the different structure of the residual circulations derived from SD and FR simulations. This Figure could be extended with the weighted mean TA-latitudes of the different reanalysis datasets (depending how often they are used for SD runs: 4x ERA-I, 1x MERRA and 1x JRA55). This would give a measure of the differences between the SD simulations and the reanalysis induced by nudging.

We have moved the Supplemental Figure 2 to the main text and have added the reanalysis datasets to the figure. We decided not to add a "multi-reanalysis mean" as we think it is important to highlight the differences between the reanalysis themselves.

The relevant part of the manuscript that was referring to this figure has also been expanded:

"To show the differences in vertical structure of the upwelling/downwelling regions between the REF-C1SD and REF-C1 experiments and between the REF-C1SD and RA, Figure 2 shows vertical profiles of the climatological TA latitudes for MMM-C1, MMM-SD and the three RA datasets used for nudging. It should be noted that since five of the REFC1-SD models were nudged towards ERA-I the MMM-SD is quite heavily weighted towards ERA-I. The REF-C1SD experiments consistently simulate a wider NH TA latitude (Figure 2b) throughout most of the stratosphere as compared to both the REF-C1 runs and the RA. In contrast, in the SH up to 30 hPa the TA latitude for MMM-SD is consistently wider than both MMM-C1 and RA (Figure 2a), while in the middle and upper stratosphere this reverses and MMM-SD shows a narrower Southern hemisphere upwelling region than both MMM-C1 and RA. Interestingly, above 10 hPa in the SH (Figure 2a) the MMM-SD does not show a widening of the upwelling region as seen in the RA. This is reflected in the structural differences in w in the SH upper stratosphere found in some models (Supplemental Figure S1). It should be noted though that the differences in TA latitudes between the REF-C1 and equivalent REF-C1SD simulations are comparable to the differences found between the three RA datasets."*

L327-338: Here, the authors discuss some of the differences between SD and reanalysis for wbar_star on the 70 hPa level as shown in Fig. 2b). As noted in the general comments, it would be helpful to add here the corresponding differences SD-RA.

As explained in the reply to comment 1, we have now added the differences between each SD model and the respective reanalysis it was nudged towards in Figure 3d.

The relevant part of the manuscript was also expanded to discuss this new result :

"Figure 3d shows the absolute differences in w between the REF-C1SD simulations and the respective RA dataset used for nudging. In the upwelling region, the REF-C1SD experiments generally show stronger upwelling around the equator than in the RA. The differences near 10 - 15°N highlight a lack of inter-hemispheric asymmetry in the double-peaked w* structure in the REF-C1SD experiment compared to the RA. Outside of the tropical pipe, the REF-C1SD experiments show weaker downwelling, particularly in the NH mid-latitudes, while at subpolar/polar latitudes (>65° latitude) the REF-C1SD runs simulate consistently stronger downwelling than the respective RA. This difference in w* at high latitudes between the REFC1-SD and RA datasets extends throughout the depth of the stratosphere (see Supplemental Figure S2)*."

L350-364: Here, the authors discuss the discrepancies between SD and RA using Figure 3b. Again, I would suggest to add a fourth panel (Figure 3d) that clearly shows the differences SD-RA (same as panel 3c) for the vertical profile of the climatological TUMF.

Thank you for this, as per your suggestion we have now added the differences between each SD model and the respective reanalysis it was nudged towards in Figure 4d.

The relevant part of the manuscript was also expanded based on this new result :

"In the lower and middle stratosphere, most of the REF-C1SD models show a smaller TUMF compared to the RA (Figure 4d), while in the upper stratosphere the majority of the REF-C1SD models simulate higher TUMF values than the respective RA."

And:

"There is a high degree of similarity in the vertical structure of the differences in TUMF between the REFC1-SD simulations for EMAC-L47/L90 and SOCOL3 as compared to ERA-I, which may be related to similarities in the implementation of nudging in these models; for example, vorticity and divergence were nudged with the same relaxation parameters (see Table 2)."

L.364-366: It is not clear to me, what the total spread here really is. Is it the standard deviation of all models TUMF in the range 100-30 hPa? Could you please clarify?

Thank you for noting that this was not stated in the manuscript. The model spread is quantified using the inter-model standard deviation. The relevant part of the text was updated to:

"The inter-model standard deviation in TUMF in the lower stratosphere (100-30 hPa), where the largest spread is found, is up to 3.5×10^9 kg s^{-1} and 3.3×10^9 kg s^{-1} in the REFC1-SD and REF-C1 experiments, respectively."

L. 377-379: The second model, for which TUMF from wbar_star exceeds TUMF from total wave forcing for the REF-C1 simulations, is the GEOSCCM and not the EMAC-L90MA.

Thank you for pointing this out, we have corrected this. However, as GEOSCCM has been removed from the main part of the study, the text was updated to:

"...the estimated TUMF from the total wave forcing for the majority of the models (apart from CESM1-WACCM), clearly exceeds the TUMF calculated directly from w..."*

L392-397: It sounds unreasonable to me that the SD-RA discrepancies for TUMF should be related to the individual RA dataset, it is much more reasonable that the individual model itself and how the nudging is implemented is causing these differences.

We agree that the SD-RA differences are not RA specific and that the method for implementing nudging is important. We have edited the text as follows :

"However, given there are substantial differences in TUMF amongst REFC1-SD models nudged to the same RA dataset, it is likely that the differences between REFC1-SD and RA are related to how nudging was implemented in each model; a wide variety of relaxation timescales and vertical nudging ranges were used by the models."

L.401-404: Is the large positive difference between TUMF derived from wbar_star and derived from total wave forcing not what one would expect, if one assumes that nudging CCMs might lead to additional forcings to the residual stratospheric circulation induced by the inconsistencies with the modelled physics and not by wave breaking? In my naive way of thinking, I would expect that TUMF from total wave forcing derived from the downward control principle (DCP) is at least equal or maybe slightly larger than the directly calculated TUMF from the residual vertical velocity due to the possible wave-wave interactions and the slight imperfections of the "exact downward control" in transient (nonsteady state) cases. The internally more consistent free-running simulations without the additional tendencies induced by nudging seem to corroborate this hypothesis. For SD simulations only the EMAC-L90 model behaves different (more realistic?) with a slightly larger TUMF derived from wave forcings. Could this be the consequence of a different nudging procedure? The EMAC-L90 and to a less degree the MRI-ESM1r1 are also the only models for which TUMF at 70 hPa derived from the SD simulation is smaller than that derived from the applied RA datasets, ERA-I and JRA55 respectively. Both models also show the smallest difference between directly calculated TUMF and TUMF derived from total wave forcing. To my opinion, it would be worth to extend the discussion on these topics (RA vs. SD) a bit more to focus the paper more on the main topic stated in the paper's title.

This is an interesting hypothesis and appears to be corroborated by most models from the 2010 SPARC CCMVal report (Figure 4.10(a), p.121, Chapter 4). The free-running simulations are internally consistent, and the imperfect match between the direct and DCP estimates shows the limitation of certain assumptions in the way the DCP is applied here, such as assuming a steady-state response to a steady mechanical forcing. As noted by the reviewer, the nudging procedure adds an additional non-physical tendency in the model equations, but from our perspective it is difficult to infer how this acts to decouple the wave forcing from the modelled residual circulation. For example, the nudging technique in EMAC-L90 is identical to that in EMAC-L47 (Morgenstern et al., 2017), so the different behaviour of those two models in Figure 4b indicate their must be sensitivities to multiple factors such as vertical resolution. In the EMAC-L90 there are far more levels throughout the stratosphere which might affect the results in that vicinity (Jockel et al 2016, ESCiMo_supplement, Figures S4 and S5). As a consequence, a dedicated study of the sensitivities of the residual circulation within one model to factors such as relaxation timescales, nudging parameters, nudging height range, vertical resolution etc would be needed to offer a detailed explanation for these differences. We therefore think it is prudent to adopt a cautious approach in trying to explain the differences between the direct and DCP calculations for the REFC1-SD experiments as these may be the result of different factors for different models.

To accommodate the points raised by the reviewer about the differences between the direct and DCP estimates in the REF-C1 experiments, we have added the following paragraph to the main text:

"On the contrary, as seen from the smaller residuals in Figure 5a, the TUMF estimated from the total wave forcing is similar to or slightly larger than the directly calculated TUMF in most of the REF-C1 simulations. This is comparable to results for the free-running CCMVal-2 models in SPARC (2010) (Figure 4.10(a), p.121). Since these simulations are internally consistent, the imperfect match indicates that the downward control principle as applied here relies on the applicability of certain assumptions such as the system being in a steady-state in response to a steady mechanical forcing (Haynes et al., 1991). The larger differences between the direct and downward control principle TUMF calculations in the REFC1-SD simulations shows that nudging adds an additional non-physical tendency in the model equations which acts to decouple the wave forcing from the residual circulation; the details of how this decoupling is manifested is

likely to vary from one model to another depending on multiple factors such as nudging timescales, nudging parameters, nudging height range, and model resolution."

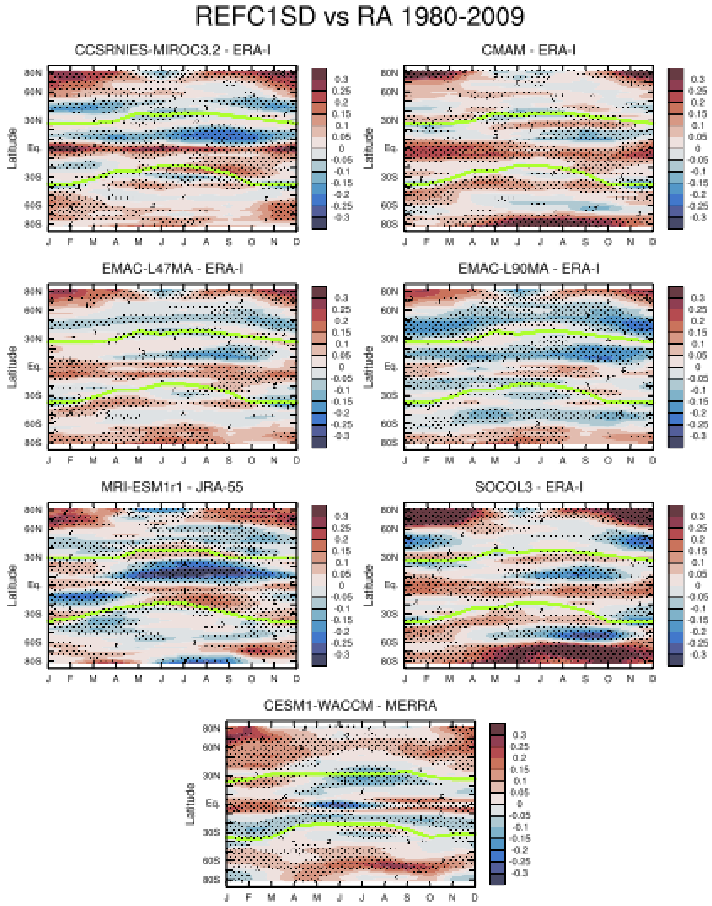
L.404-406: The author states that the lower directly calculated TUMF values for the higher vertically resolved EMAC-L90 compared to -L47 is in line with the finding for the SOCOL3 model reported by Revell et al. (2015b), i.e. 90 vs. 39 layers. This is true for the FR simulations which has been used in the sensitivity experiments by Revell et al. (2015b), but does this conclusion also hold for both EMAC SD simulations? Does the vertical resolution also matters here?

Thank you for highlighting that this sentence was unclear. The higher vertical resolution version of EMAC (L90) simulates weaker directly estimated TUMF than its L47 counterpart in both the C1 and SD experiments. We have edited the text to clarify this :

"...seen in both REF-C1 and REF-C1SD experiments…"

L.427-436: The differences in the annual cycle of wbar_star at 70 hPa between SD and RA should be added to the Figure 5 and discussed here.

Thank you for your suggestion. Similar to your comment about calculating a "multi-reanalysis mean" for the previously Supplemental Figure 2, we think it is important to highlight the differences between each SD model and the respective reanalysis instead (now added as Supplemental Figure S3). The following plot illustrates that the SD models simulate a rather different annual cycle compared to the RA they were nudged towards.



We have also expanded the relevant part of the manuscript based on this result :

"However, it should be noted that the annual cycle in lower stratospheric w in the REF-C1SD runs generally differs from that in the respective RA they were nudged towards (Supplemental Figure S3). In fact, the REF-C1SD – RA comparison highlights a wide variety in both the magnitude and the spatial patterns of their absolute differences, exhibiting on average larger differences than the REF-C1SD – REF-C1. Within the tropical upwelling region similarly to the differences between REF-C1 and REF-C1SD (the narrow band between the equator and 10°N) the local minimum associated with the weaker upwelling simulated by REF-C1SD is extended in most cases throughout the year when compared with the RA (Supplemental Figure S3)."*

L.431-432: The author states that NH midlatitude downwelling during summer is stronger for the SD simulations. To my view, there are only blueish colours north of the TA-latitudes during JJA in Figure 5c indicating weaker and not stronger downwelling.

Thank you for pointing this out, we have now corrected this.

L.451-452: "The REF-C1SD shows…" I assume that this sentence concerns the comparison between SD and RA, however it might be better to clarify this.

Thank you for pointing out this unclear sentence. It was referring to the differences between C1 and SD. As some minor details of the results have changed due to the fact that we are now comparing the models that performed both sets of experiments we have altered this whole paragraph to better communicate the key points.

L.454-456: Why is the annual cycle of wbar_star in the tropical lower stratosphere more consistent for the SD compared to the FR simulations? Above you conclude that the annual cycle and the phasing of SD simulations are weakly constrained and the intermodal spread is 20% larger than for FR simulations. How does this fit together?

We thank you for highlighting this statement which was poorly formulated and did not adequately communicate the interhemispheric and seasonal aspects of this comparison. Based on Figures 7a/b we conclude that the nudging constrains the tropical mean w* in boreal summer (JJA), which exhibits ~20% less spread than the REF-C1 experiments, but it does not constrain the tropical mean upwelling in boreal winter (DJF), which shows a factor of two larger spread than the free-running models. With respect to the turnaround latitudes, the REFC1-SD experiments exhibit less spread (by up to ~3° or 25%) than the REF-C1 experiments in both hemispheres during boreal winter (DJF) and in the SH during austral winter (JJA). In contrast, in the NH during austral winter the turnaround latitude in the nudged simulations shows an increased intermodel spread by 5.5° (25%) compared to the REF-C1 runs. Hence, the picture emerging is rather more complex than was written in the initial manuscript. We have edited/expanded the paragraphs for figures 7a/b and 7c/d in the manuscript and the associated changes with respect to your comment are (nb: black italic below denote unchanged parts of the manuscript) :

"Based on Figures 7a/b we conclude that the nudging constrains the tropical mean w in boreal summer (JJA), which exhibits ~20% less spread than the REF-C1 experiments, but it does not constrain the tropical mean upwelling in boreal winter (DJF), which shows a factor of two larger spread than the free running models. The differences between the REF-C1SD runs and the respective RA they are nudged towards are generally larger in boreal winter than in boreal summer for the majority of the REF-C1SD models. Figures 7c and 7d show the climatological annual cycle in the turnaround latitudes at 70 hPa for the REF-C1 and REF-C1SD runs, respectively. For both REF-C1 and REF-C1SD runs the NH TA latitude varies significantly more than the SH TA latitude during boreal summer (JJA). The NH TA latitude is displaced more poleward for the REF-C1SD runs during boreal summer (JJA) for most models, corroborating the results of section 3.1 associated with an asymmetrical and wider upwelling region (Figure 2)*

in the lower stratosphere. The REF-C1SD runs exhibit less spread (by up to ~3° or 25%) than the REF-C1 experiments in both hemispheres during boreal winter (DJF) and in the SH during austral winter (JJA). In contrast, in the NH during austral winter the turnaround latitude in the nudged simulations shows an increased intermodel spread by 5.5° (25%) compared to the REF-C1 runs. Nevertheless, the amplitude of the annual cycle for the RA is in fact quite amplified when compared with the REF-C1SD (just for the SH) while the inter-reanalysis spread is significantly higher (more than 2.5° in the SH) than both REF-C1 and REF-C1SD runs. To summarize, there is substantial inter-model spread in the turnaround latitudes and the amplitude of the annual cycle highlighting significant interhemispheric differences in the upwelling region between both sets of simulations as well between the nudged experiment and the RA. Hence, the picture emerging is rather more complex one. Arguably, the nudged runs are closer to the free-running simulations rather than the reanalysis they were nudged towards, hence not constraining the annual cycle of the residual vertical velocity in the lower stratosphere."

L.456-459: Again, I am missing in the summary here the SD vs. RA comparison relevant for analysing the nudging effect on stratospheric residual circulation in CCMs.

Thank you for highlighting this. We believe we have sufficiently covered this point in the replies in the two previous comments.

L.503-569 (Section 3.5 and 3.6): Would it be possible to add here also the MLR of TUMF at 70 hPa and the related trend analysis for the RA datasets? The latter might give a very interesting insight into the question, if the different linear trends of FR and SD simulations are mainly driven by the different stratospheric residual circulation of the RA datasets or if they are significantly influenced by the nudging of the CCMs.

Thank you for this suggestion. We have applied the MLR analysis and trend sensitivity analysis to the RA datasets and have included the results in the supplement with some discussion in the main text (Figures S4, S7, S8).

The analysis reveals some interesting behaviours. In terms of the MLR analysis, ERA-I shows the highest R^2 value (0.66) while in MERRA the variance explained by the MLR is substantially lower (0.3) than the other RA datasets and the SD models (Figure S4). In terms of long-term trends, none of the SD models nudged towards ERA-I simulate a negative trend in 70 hPa TUMF over 1980-2009 despite the ERA-I dataset showing a negative trend (Figure S8). The trend sensitivity analysis for the RA datasets highlights a common feature of a negative trend in TUMF from the mid-1990s onwards. This feature is not found in any of the SD models (Figure 13), but is captured by some free-running simulations (Figure 12).

The relevant parts of the manuscript have been expanded to discuss the MLR and trend results for the RA datasets:

"The MLR analysis was also applied to the RA TUMF at 70 hPa (Supplemental Figures S4 and S7). This highlights significant discrepancies in attributing the variance in TUMF in the different RA datasets to the various proxies used in the MLR model. Both the volcanic activity and ENSO contributions to the variance in the TUMF is rather suppressed when compared to the REF-C1SD runs. The negative linear trend in ERA-I is in strong contrast to the positive trends found in the other RAs and the REF-C1SD models. The negative trend in ERA-I found in the TUMF in the lower stratosphere over 1980-2009 corroborates the findings of Abalos et al. (2015) who showed trends based on the residual circulation w estimate in ERA-I were also negative over 1979-2012 throughout the depth of the stratosphere. Despite this difference in the representation of long-term changes, ERA-I shows the highest percentage of TUMF variance explained (66%), with MERRA showing a substantially lower R^2 (0.3) compared to the other reanalyses and the REFC1-SD models. The residuals are generally less correlated between the reanalyses on interannual timescales than was found in the REFC1-SD simulations, but are similar on inter-decadal timescales; however, the residuals in the RAs show a different temporal behaviour from those found in the REF-C1SD simulations (Figure 11) (note that the y-axis scale for the residuals in Supplemental Figure S4 is double that for the CCMI models in Figures 10 and 11). In summary, although nudging constrains the interannual variability in the TUMF at 70 hPa, the REF-C1SD runs generally do not resemble the RA they were nudged towards."*

And

"Interestingly, the RA trend sensitivity analysis highlights that nudging does not constrain the underlying trends of the REF-C1SD models in the TUMF at 70 hPa, as the RA datasets exhibit a wide range of trends (Supplemental Figure S8) when compared to the REF-C1SD runs. None of the REFC1-SD models simulate the statistically non-significant decrease in 70 hPa TUMF starting around the mid-1990s seen in the RA datasets."

L585-588: What is the explanation of the differences among the directly simulated TUMF for SD simulations that can be derived from the diagnosed wave forcing using the downward control principle? (See also my comments above: L.401-404)

Please see our reply to the previous comment above. We have added the following sentence in this section of the manuscript:

"The large intermodel spread in the differences between the direct and downward control estimates of lower stratospheric TUMF in the REFC1-SD simulations shows that multiple factors

are likely to affect the decoupling of the wave forcing and residual circulation; this includes nudging timescales, nudging parameters, nudging height range, and model resolution. Further nudging sensitivity studies to examine the effects of these factors within one model would help to offer a more detailed explanation of these differences."

L.601-602: What is the explanation that ERA-I shows no positive trend in tropical upwelling but a significant positive linear trend in TUMF at 70 hPa?

This sentence was not phrased clearly in the original manuscript. We intended to communicate that ERA-Interim does not show a positive trend in both TUMF and tropical upwelling (see Figure S7 in the revised manuscript for TUMF and also Figure 11 of Abalos et al., (2015); for trends in tropical average upwelling). But the models that were nudged towards ERA-Interim do show a positive trend in TUMF, so there is an inconsistency between what the RA dataset and the nudged models show in terms of long-term trends.

We have rephrased this bullet to:

"6. Most nudged simulations show a statistically significant positive trend in tropical upward mass flux (TUMF) in the lower stratosphere over 1980-2009, which is on average larger than the trends simulated in the free-running models. This is despite the fact that five out of the seven models analysed were nudged towards ERA-Interim, which shows a negative trend in TUMF (also Abalos et al., 2015), while JRA-55 and MERRA show a positive trend."

References:

Froidevaux, L., Kinnison, D. E., Wang, R., Anderson, J., and Fuller, R. A.: Evaluation of CESM1 (WACCM) freerunning and specified dynamics atmospheric composition simulations using global multispecies satellite data records, Atmos. Chem. Phys., 19, 4783-4821, 10.5194/acp-19-4783-2019, 2019.

Kida, H.: General Circulation of Air Parcels and Transport Characteristics Derived from a Hemispheric GCM Part 2. Very Long-Term Motions of Air Parcels in the Troposphere and Stratosphere, Journal of the Meteorological Society of Japan. Ser. II, 61, 510-523, 10.2151/jmsj1965.61.4_510, 1983.

Rosenlof, K. H.: Seasonal Cycle of the Residual Mean Meridional Circulation in the Stratosphere, J. Geophys. Res., 100, 5173-5191, 1995.

Schmidt, U., and Khedim, A.: In situ measurements of carbon dioxide in the winter Arctic vortex and at midlatitudes: An indicator of the ‘age’ of stratospheric air, Geophys. Res. Lett., 18, 763-766, 10.1029/91gl00022, 1991.

Abalos, M., Legras, B., Ploeger, F. and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979-2012, J. Geophys. Res. Atmos., 120(15), 7534–7554, doi:10.1002/2015JD023182, 2015.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A. and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, Nat. Geosci., 2(1), 28–31, doi:10.1038/ngeo388, 2009.

Haynes, P. H., McIntyre, M. E., Shepherd, T. G., Marks, C. J. and Shine, K. P.: On the “Downward Control” of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces, J. Atmos. Sci., 48(4), 651–678, doi:10.1175/1520-0469(1991)048<0651:OTCOED>2.0.CO;2, 1991.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O’Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M. E., Marchand, M., Marécal, V., Michou, M., Oman, L. D., Pitari, G., Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K. and Zeng, G.: Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI), Geosci. Model Dev., 10(2), 639–671, doi:10.5194/gmd-10-639-2017, 2017.

Revell, L. E., Tummon, F., Stenke, A., Sukhodolov, T., Coulon, A., Rozanov, E., Garny, H., Grewe, V. and Peter, T.: Drivers of the tropospheric ozone budget throughout the 21st century under the medium-high climate scenario RCP 6.0, Atmos. Chem. Phys., 15(10), 5887–5902, doi:10.5194/acp-15-5887-2015, 2015.

Stiller, G. P., Von Clarmann, T., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S. and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmos. Chem. Phys., 12(7), 3311–3331, doi:10.5194/acp-12-3311-2012, 2012.

Waugh, D. W. and Hall, T. M.: Age of stratospheric air: Theory, observations, and models, Rev. Geophys., 40(4), 1010, doi:10.1029/2000RG000101, 2002.