Reviewer (Comments):

Review of "The effect of atmospheric nudging on the stratospheric residual circulation in chemistry-climate models" by Andreas Chrysanthou et al.

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.

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 analyse 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.

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).

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

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

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.

L327-338: Here, the author discuss some 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.

L350-364: Here, the author 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.

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?

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.

L392-397: It sounds not reasonable 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.

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 (non-steady 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.

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 holds for both EMAC SD simulations? Does the vertical resolution also matters here?

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.

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.

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.

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?

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.

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.

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)

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?

References:

Froidevaux, L., Kinnison, D. E., Wang, R., Anderson, J., and Fuller, R. A.: Evaluation of CESM1 (WACCM) free-running 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.