

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 REF-C1-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 $3.5 \times 10^9 \text{ kg s}^{-1}$ and $3.3 \times 10^9 \text{ kg s}^{-1}$ in the REF-C1-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 REF-C1-SD models nudged to the same RA dataset, it is likely that the differences between REF-C1-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?

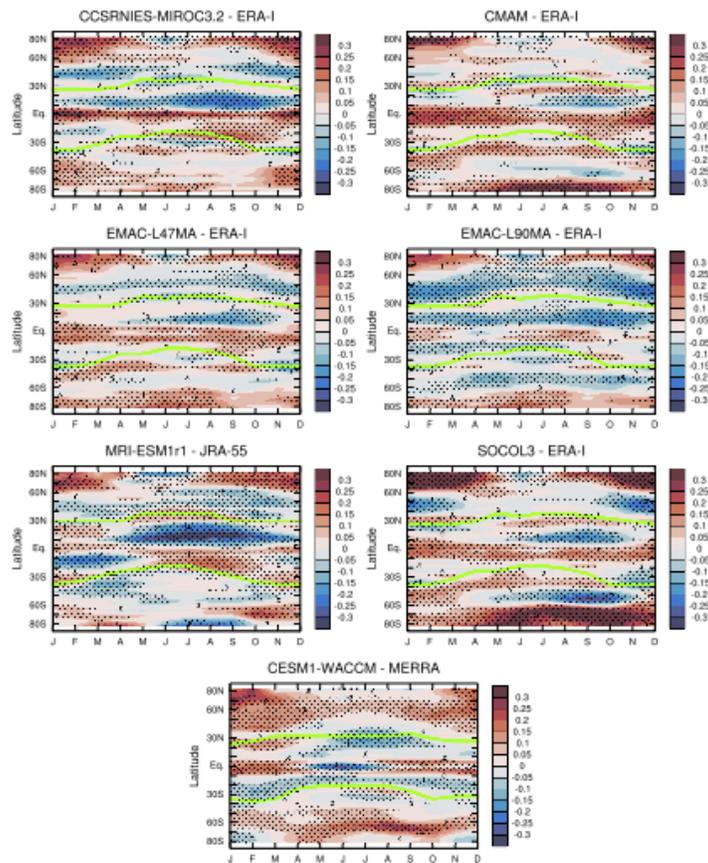
Thank you for highlighting that this sentence was unclear. The higher vertical resolution version of EMAC (L90) simulates weaker directly estimated TUMF than its L47 counterpart in both the C1 and SD experiments. We have edited the text to clarify this :

“...seen in both REF-C1 and REF-C1SD experiments...”

L.427-436: The differences in the annual cycle of \bar{w} 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.

REFC1SD vs RA 1980-2009



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 \bar{w} 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 $\sim 3^\circ$ 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 $\sim 3^\circ$ 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-CISD runs. The negative linear trend in ERA-I is in strong contrast to the positive trends found in the other RAs and the REF-CISD 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-CISD 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-CISD 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-CISD 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-CISD 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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., Maréchal, 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1029/2000RG000101), 2002.

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 #2

In this paper, the authors compare the free running and specified dynamics simulations of the CCM1 models. They find that although the nudging involved in the SD does make the models short term variability line up, the residual circulation in the nudged runs matches neither the original models nor the reanalyses to which the models are nudged.

This paper is important. It is interesting and provides a cautionary tale against using these SD runs to examine transport of long-lived trace gases, which one would expect would be significantly impacted by these inconsistencies. It is well reasoned and well written. I recommend publication following some minor revisions. An expanded introduction would be extremely valuable to the reader, and I have included recommendations for that below. The primary scientific points that would add significantly to the paper are:

1. Discuss the choice of metric for the BDC in more detail. (Line 191)
2. Reconsider the results of the multiple linear regression in light of the importance of the relative phasing of the annual cycle and the QBO.
3. Consider the differences between the models and the reanalyses they are being nudged towards, especially vertical profiles, to try to better understand why the results are what they are.

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

Specific comments:

1) Recent work (Abalos et al. 2015, Linz et al. 2019) has raised the issue that the differences due to different calculations of the residual circulation are as significant as (or more than) the differences between the reanalyses themselves (c.f. Seviour et al. 2012, also). Therefore, I would appreciate more discussion of the choice made here to use the TUMF and downward control (which is used to distinguish the different wave forcing components). How do these choices impact the results, if at all? A thermodynamic calculation might be very different, since the temperature is nudged, which will introduce spurious diabatic terms. How is TUMF different from w^* averaged between turnaround latitudes?

Thank you for this suggestion. We agree that some further discussion of the chosen measures of residual circulation is important to place the results into a broader context. As indicated by the reviewer, the choice of diagnostics is likely to be a particular issue for interpreting the nudged

models and reanalyses due to the additional tendencies imposed in the model equations. Part of our choice is pragmatic, based on the availability of output from the CCM1 model dataset, which provides direct estimates of residual circulation diagnostics following Andrews et al. (1987) and resolved and parameterized wave forcing components. As TUMF is an integrated measure of w^* between TA latitudes, the results regarding interannual variability and long-term trends are similar to if an average w^* measure is used. Arguably, the TUMF is a first-order measure in order to evaluate whether the mass circulation is affected by nudging and it has been widely used in previous multi-model comparison and reanalysis studies and so we include it here (e.g. Butchart et al., 2006, Butchart et al., 2010, Butchart et al., 2014 and references therein; Rosenlof, 1995; Seviour et al., 2012; SPARC CCMVal, 2010). The important point is that we use consistent diagnostics in the free running and nudged experiments and the reanalyses, so that we are always comparing the same measures across the different simulations. Unfortunately, as heating rates are not provided by all models in the CCM1 output, it would not be possible to compute a consistent (i.e. with the appropriate model radiation scheme) estimate of the residual circulation using the thermodynamic equation. It is possible that for the nudged simulations calculation of the residual circulation based on the thermodynamic equation might result in different behaviour to the direct TEM formulation used here. We now suggest this as an interesting topic for future research in the conclusions. The downward control principle (DCP; Haynes et al., 1991) calculations provide a framework to assess the extent to which the wave forcing and directly estimated residual circulation become decoupled as a consequence of the additional tendencies imposed by the nudging.

Based on the above we have expanded/edited the relevant parts of the text to the following :

End of section 2.2.1, (just before the introduction of DCP) :

“The TUMF has been used widely as a measure of the strength of the BDC (Rosenlof, 1995; Butchart et al., 2006, Butchart et al., 2010, Butchart et al., 2014 and references therein; Seviour et al., 2012), so its use here enables a direct comparison with earlier studies. Arguably, the strength of the TUMF is a first-order metric to evaluate changes to the mass circulation as a consequence of nudging. As mentioned above, by calculating the annual means of TUMF accounting for the seasonal cycle of the turnaround latitudes we capture the correct evolution of the intraseasonal (not shown) and interannual variability in the TUMF.”

End of section 2.2.2, (after introduction of DCP) :

“Both the direct and DCP methods rely on the applicability of the quasi-geostrophic approximation to interpret the results. We note that in addition to the direct and DCP approaches used here, the residual circulation can also be estimated using the thermodynamic

equation. Unfortunately, heating rates were not available from all CCM1 model simulations to perform this calculation. Studies have shown that the estimates from the different methods for evaluating the residual circulation can differ (Abalos et al., 2015; Linz et al., 2019; Seviour et al., 2012), particularly in reanalyses where the normal conservation laws are not required to be met. Similar issues are likely to beset the nudged model simulations owing to the additional tendencies included in the model equations. The differences between the calculation methods for the residual circulation can be as large as, or larger than, the differences between reanalysis datasets for the same diagnostic (Abalos et al., 2015; Linz et al., 2019), and may further depend on choices around averaging between fixed latitudes or the turnaround latitudes (Linz et al., 2019), so it is important to bear this in mind in interpretation of the results presented here. Nevertheless, we compute the diagnostics for the residual circulation in a self-consistent manner in the models and reanalyses to enable comparison with earlier multi-model studies (e.g. Butchart et al., 2006, Butchart et al., 2010, SPARC, 2010).”

2) Although the method used here to do the multiple linear regression is somewhat standard, it does not actually remove all of the variability that is related to these signals. In particular, the relative phasing of the QBO and the annual cycle will cause an enhanced wintertime upwelling in some years, so the result that the leftover variability is highly correlated could be related to this. Randel et al. 1999 have a nice treatment. I would consider using the method in that paper (described on p. 458, see also Randel and Wu 1996) to determine if the relative timing of the annual cycle and QBO is the cause of the covariation of the nudged time series. This phenomenon is also mentioned in the cited Baldwin et al. 2001 QBO review.

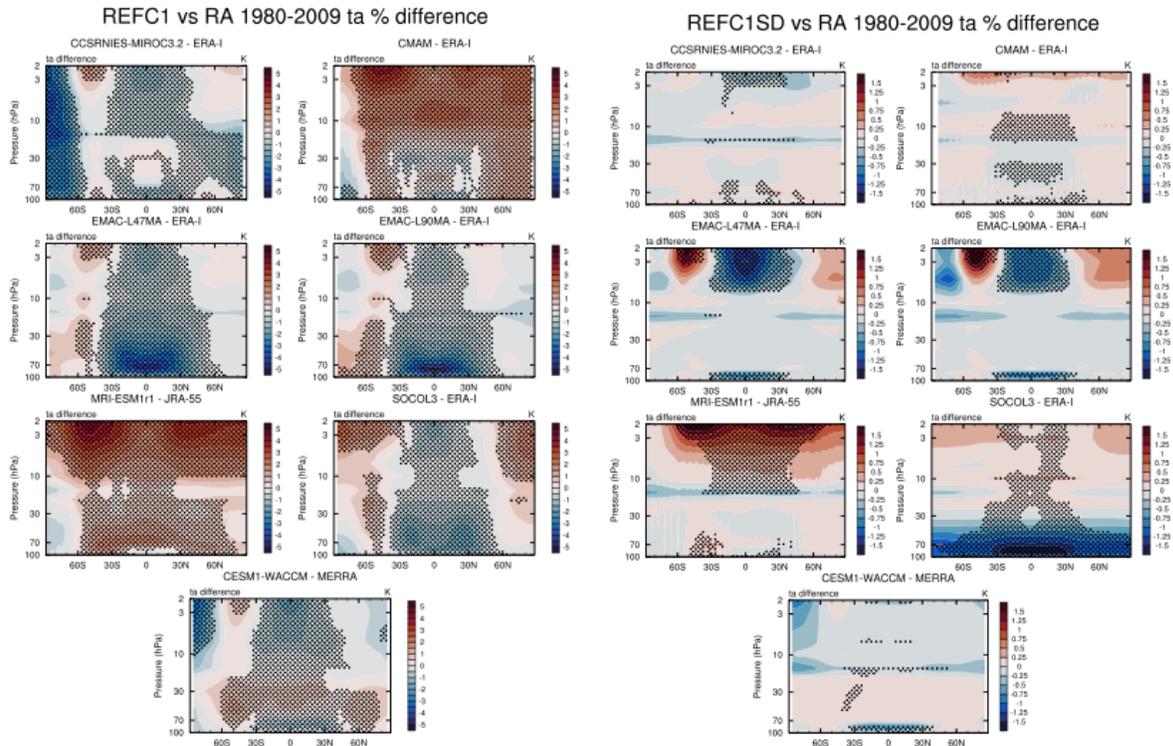
Thank you for raising this interesting point. The MLR model is applied to the annual mean TUMF timeseries and hence the QBO terms are resolved at an annual timescale. Therefore, the method presented by Randel et al. (1999) to account for interannual variations in the relative phasing of the QBO with the annual cycle is not readily applicable in our analysis. We further note that all of the models except CMAM and MRI-ESM1 also nudge the phase of the QBO in the REF-C1 simulations (Morgenstern et al., 2017). Hence, if this is a main factor for explaining the MLR residuals, we expect that such an effect would also be seen in the REF-C1 simulations. However, the temporal correlation of the residuals is much weaker in the REF-C1 experiments for those models that nudge the QBO, suggesting this is not a major factor for explaining the correlated residuals in the REFC1-SD experiments.

3) The authors find that the reanalyses differ from the nudged runs which in turn differ from the free running models. I would like to see some presentation of the difference between the mean states of the reanalyses and the mean states of the models without nudging. The result that the models tend to have stronger upwelling in their SD runs could be caused by a systematic low bias in the model temperature in the tropics compared to the reanalysis temperature. The repeated

nudging with a primarily positive value would then cause an apparent increase in upwelling, without changing the underlying model physics that are the cause of any systematic mean state bias. To understand why the nudging does what it does, I think the difference between the free running models and the reanalyses needs to be explored in much more detail—perhaps for just one model. A continuous spurious forcing because of a mismatch in the mean state (and likely also in the seasonal cycle) should impact the mean while leaving the interannual variability alone (line 612).

Thank you for this suggestion. We present below the percent differences in zonal mean temperature in the stratosphere between the REF-C1 simulations and the respective reanalysis they are nudged towards (stippling denotes statistical significance at 95% based on a two-tailed Student's t-test); the equivalent differences for the REFC1-SD simulations are also shown in the right panels (note the reduced colour scale).

Most models show a cold bias relative to the reanalyses in the tropical tropopause layer, except for MRI-ESMr1 which shows a warm bias. The two versions of EMAC show the largest relative cold bias in the TTL. Much of this bias is alleviated in the nudged simulations, as expected, but EMAC (I47/90), CESM1-WACCM and SOCOL3 still show a (smaller) cold bias in the TTL. Larger differences remain above the level where nudging ceased to be imposed (e.g. 10 hPa in EMAC and 40 hPa in MRI-ESMr1). The opposite sign of the TTL temperature bias in the free-running MRI-ESMr1 model provides a useful test bed for the point raised by the reviewer. The REFC1-SD MRI-ESMr1 simulation still shows stronger upwelling near the equator compared to the free-running model (Figure 3c) despite having a warm bias; this result is similar to the effect of nudging in the other models that show a cold TTL bias. Hence from this cursory investigation, we do not identify a systematic relationship between the intrinsic model bias and the effect of the nudging on the residual circulation. To get a better handle on this, a detailed analysis of the sensitivities within one model to relaxation timescales and nudging parameters such as the altitude range etc. would be needed in order to robustly establish how nudging affects the mean residual circulation.



Minor comments

I find the second paragraph of the introduction to be lacking. Although Abalos et al. (2015) did conclude that there was likely a 2-5%/decade trend in the circulation, they found substantial differences with different calculations for the circulation strength. Ploeger et al. (2019) have a new paper out on the inconsistency of the BDC trends for the different reanalyses, looking at the age spectra. Age of Air is mentioned somewhat abruptly and with no explanation. The results are also misrepresented. Engel et al. 2009 did not find a statistically significant increase in the mean age of air in the midlatitudes. They found no statistically significant change. There are limitations both on the availability of measurements and on the interpretation of age of air as a direct proxy for the residual circulation, and much more recent papers discuss this, rather than Waugh and Hall 2002. There are new satellite based measurements—ACE-FTS (e.g. Ray et al. 2016), MIPAS (Haenel et al. 2015), and MLS (Linz et al. 2017)—and new measurements from Air Core (Engel et al. 2017). The interpretation of the data is also difficult because of the separation of the residual circulation from the mixing (e.g. Garny et al. 2014, Ploeger et al 2015, as mentioned in the first paragraph—perhaps just move this comment to this paragraph?). The only attempt to calculate the residual circulation from tracer data did not calculate the residual circulation directly, but instead the diabatic circulation (Linz et al. 2017), and they did not calculate trends.

Thank you for this comment. We have expanded this part of the introduction to give a more balanced and detailed synopsis of the literature:

“Past studies have shown substantial spread across models in the mean strength of the residual circulation (e.g. Butchart et al., 2010). Nevertheless, climate and chemistry-climate models (CCMs) consistently simulate a long-term strengthening of the residual circulation with an increase of $\sim 2\%$ decade⁻¹ (e.g. Butchart et al., 2010; Hardiman et al., 2014), though there are differences across models in the relative contribution to trends from resolved and parameterized wave forcing. Reanalysis datasets also suggest a strengthening of the residual circulation over the past several decades of the order 2-5% decade⁻¹ (Abalos et al., 2015; Miyazaki et al., 2016) apart from one (ERA-Interim) which shows a weakening of the deep branch of the BDC (Seviour et al., 2012; Abalos et al., 2015). However, reanalyses are subject to multiple caveats, particularly in their suitability for trend studies, and there are substantial differences in residual circulation trends calculated from the same reanalysis using different methods (Abalos et al., 2015). Given the limitations of reanalyses, evaluating the fidelity of model estimates of residual circulation variability and trends is challenging since there are no direct measurements of the residual circulation.

The only direct estimates of the stratospheric mass circulation come from tracer measurements, which can be used to calculate stratospheric age-of-air (AoA) (Kida, 1983; Schmidt and Khedim 1991; Waugh and Hall, 2002). AoA represents the combined effects of advection by the residual circulation and turbulent mixing processes, and as such cannot be directly related to the residual circulation. While progress has been made in separating the relative effects of these two factors for AoA calculated from models (Garny et al., 2014; Dietmuller et al., 2018, Eichinger et al., 2019, Šácha et al., 2019), using Lagrangian models driven by reanalysis fields (Ploeger et al., 2015a, Ploeger et al., 2015b, Ploeger and Birner, 2016, Ploeger et al., 2019) and comparing these effects in both CCMs and Lagrangian models (Dietmuller et al., 2017), this is more difficult to achieve in observations. Engel et al. (2009) and Stiller et al. (2012) used balloon-borne measurements of stratospheric trace gases and found a statistically non-significant increase in AoA in the middle stratosphere at northern mid-latitudes, which has been corroborated by more recent study using balloon-borne measurements at two midlatitude sites in the Northern Hemisphere (Engel et al., 2017). It has been hypothesised based on analyses of more recent satellite datasets, which have greater spatial and temporal coverage, that the subtropical AoA trends can be explained by a weakening of the mixing barriers at the edge of the tropical pipe (Neu and Plumb, 1999) that is masking the effects of an increase in tropical upwelling on AoA (Stiller et al., 2012; Haenel et al., 2015). In contrast with the AoA trends derived from observations, CCMs forced with observed sea-surface temperatures (SSTs), greenhouse gases and ozone-depleting substances show a decrease in AoA throughout the stratosphere (World Meteorological Organization, 2018; Li et al., 2018; Morgenstern et al.,

2018; Abalos et al., 2019; Polvani et al., 2019). Theoretical approaches based on the tropical leaky pipe model (Neu and Plumb, 1999) have shown promise for bridging the information on the stratospheric circulation derived from observations with outputs from GCMs/CCMs (Ray et al., 2016).

As explained above, it is difficult to infer information on the residual circulation from AoA without carefully accounting for the effects of mixing. More recent theoretical developments offer a means of calculating the diabatic circulation using stratospheric tracers (Linz et al., 2017), which is a promising avenue as this is more closely related to the residual circulation than AoA. The results of Linz et al. (2017) showed consistent estimates of the diabatic circulation in the lower stratosphere, but large uncertainties of up to a factor of two in the mean strength in the upper stratosphere. Targeted measurement strategies to better characterise long-term changes in the observed stratospheric meridional circulation have been proposed (Moore et al., 2014; Ray et al., 2016)."

A more detailed explanation of what nudging is, how it is implemented, why it is used, and the motivation for this work would help the introduction. The current paragraph in the introduction (line 85) is good, and I think that including more specific examples would be very helpful. Similarly, in the conclusion, commentary on what these nudged runs could be useful for would be appreciated in addition to the entirely appropriate cautionary words.

Thank you for this comment. We have expanded this paragraph to include your suggestions above. The relevant part of the manuscript is now (nb: black italic below denote unchanged parts of the manuscript) :

"In an attempt to obtain a closer comparison with observed stratospheric trace species, some studies have used model simulations with meteorological fields nudged or relaxed towards a reanalysis dataset (Jeuken et al., 1996); these include many studies of ozone variability and trends (e.g. van Aalst et al., 2003; Solomon et al., 2016; Hardiman et al., 2017b; Ball et al., 2018), comparisons between models and satellite-based multi-species observational records (Froidevaux et al. 2019) as well as the chemical and climatic effects of volcanic eruptions (Löffler et al., 2016; Solomon et al., 2016; Schmidt et al., 2018). Nudged simulations have also been used to study mechanisms for dynamical coupling between the stratosphere and troposphere (Hitchcock and Simpson, 2014). Nudging involves adding additional tendencies to the model equations by constraining the simulated meteorological fields to reanalysis fields (Kunz et al., 2011). Nudged variables can include horizontal winds (or divergence and vorticity), temperature, surface pressure, latent and sensible heat fluxes. However, vertical winds, which are small residual from horizontal divergence, are not nudged and the underlying model physics can yield quite different results from the reanalysis that they are nudged towards (Telford et al., 2008; Hardiman et al., 2017). A recent study by Orbe et al. (2018) analyzed tropospheric tracers

in nudged CCM simulations and found large differences in the distributions of the tracers, which could be partly traced to differences in the model convection schemes. They urged users to adopt a cautious approach when interpreting tracers in nudged simulations given their dependence on not only large-scale flow but also sub-grid parameterizations. In contrast, critical evaluation of the stratospheric residual circulation in models relaxed to reanalysis fields has been rather limited to date.”

As for the conclusions the now expanded relevant part of the manuscript (second from last paragraph of the manuscript):

“Despite the limitations described here, some success has been reported in studies that used nudged simulations to investigate specific meteorological events, such as Sudden Stratospheric Warmings, and in particular for exploring processes beyond the top of the nudging region in the Mesosphere-Lower Thermosphere (e.g., Tweedy et al., 2013; Chandran and Collins, 2014; Pedatella et al., 2014). In order to reduce discrepancies between nudged and free-running simulations, various nudging techniques have been investigated. The role of gravity waves in the error growth that the nudging introduces over time has been highlighted for a single model (Smith et al., 2017). Constraining just the horizontal winds without the temperature was found to be a good strategy when investigating the aerosol indirect effects without affecting significantly the mean state (Zhang et al., 2014)...

...A dedicated study of the sensitivities within one model to relaxation timescales, nudging parameters, nudging height range, and model resolution would be needed to offer a detailed explanation for these differences.”

Do any of the “free running” models have a nudged QBO, and if so, which ones?

All of the free-running models presented in the study apart from CMAM and MRI-ESM1 impose a QBO “nudging” in their simulations as mentioned in the last paragraph of section 3.4 (Table S8 of supplement of Morgenstern et al., 2017).

Since the focus of this paper is really on the difference between SD and free running models, it might make sense to omit the models that do not include both. I respect the authors choice to have those comparisons in the supplementary information. However, I think I would find the plots more manageable without the additional lines.

Thank you for this recommendation. In response to this comment and also the comments of reviewer 1 we have removed the models which did not perform the SD simulation from the main body of the paper (GEOSCCM, NIWA-UKCA and ULAQ-CCM) and the main text now focuses

on the seven models for which both REF-C1 and REFC1-SD simulations are available. For completeness a subset of diagnostics are shown for those models in the supplementary material.

Line 106: aren't they necessarily inconsistent?

We have corrected this to: *“will add forcings that are inconsistent with the model state.”*

Lines 100-129: excellent discussion.

Thank you for the positive comment!

Line 153: Do you mean Table 2?

Thank you for spotting this, we have now corrected this.

Line 155 etc.: More details here would be good. I don't understand what is meant by T (with wave-0), and I assume that “temperature” is also T (for CMAM), but if that's true then they all nudge temperature so maybe mention that? Including the sentence “Nudging timescales range from 6 - 50 hours.” (or some equivalent) would also emphasize how different these treatments are.

We have added some more details here as per your recommendation. In Table 2, temperature is referred to consistently as T. The wave-0 was intended to refer to the additional nudging of the global mean temperature in EMAC. We now state this explicitly. We have also changed L152-153 in the text to: *“The REF-CISD simulations nudge temperature and other meteorological fields such as horizontal winds, vorticity and divergence and some surface fields (Table 2), while the chemical fields are left to evolve freely. The nudging timescales range from 6 - 50 hours and the height range over which nudging is applied varies (Table 2).”*

Line 159-160: Table 1?

Thank you for spotting this, we have now corrected this.

Line 168: “primary variable”—not sure what is meant by this

We meant prognostic. We have now corrected this.

Line 442: I have trouble seeing this from the plot. Could you provide the numbers? Is the difference in the peak-to-peak amplitude significant?

Yes, due to the congestion of the multiple lines is not that easy to see but the difference is not significant. This sentence now includes the numbers of the peak-to-peak amplitude

“Comparing the MMM annual cycle of the REF-CI runs with the MMM REF-CISD of Figure 7b reveals that on average the REFCI-SD models show a slightly larger peak-to-peak amplitude (0.22 mm s⁻¹ vs. 0.19 mm s⁻¹) in the annual cycle.”

Line 476. . .: How do these results relate to Abalos et al. 2014, who found such a strong dependence of the upwelling on the location of the EPFD?

Thank you for this interesting point. We agree it would be interesting to examine how the patterns of EPFD vary between the free running and nudged simulations and how this relates to fluctuations in upwelling. This work is planned for a follow-up study.

Line 488-90: I found this sentence confusing.

Thank you for spotting this unclear sentence. We have removed this sentence as it was related to an earlier sentence on lines 483-485, which has been edited to:

“This result indicates that although nudging does not constrain the mean residual circulation, it does constrain the interannual variability and produces very similar contributions to variability across models from both resolved and parameterized wave forcing.

Line 498: remove “due to”

Done.

Section 3.6: Nice.

Thank you for the positive comment!

Line 606.: This paragraph is very long and contains a lot of important information. Turn into two or even three shorter paragraphs?

Thank you for this comment, we have now turned this paragraph into three shorter ones to improve the clarity.

Figures: Consider stippling the insignificant part of the plot. I read this on a number of different devices, and on some it looked fine, but on one of them the plots were impossible to see because the colors weren't visible behind the stippling. This wasn't an issue for the last two figures for some reason.

It may take a while for these plots to be loaded properly in a pdf viewer, as they are quite large files. We changed the subtleties of the stippling patterns as seen in the figures and hopefully they are more clear now.

It could be useful to have models that are nudged to the same reanalysis plotted in the same color family to aid interpretation of the complicated line plots.

We have now colour coded the model results nudged towards ERA-I in red(-ish) colours. We hope that without the additional lines of the extra models, the results are more clear.

References:

Linz et al. 2019: <https://www.atmos-chemphys.net/19/5069/2019/acp-19-5069-2019.html>

Randel et al. 1999 <https://acomstaff.acom.ucar.edu/randel/JAS%201999.pdf>

Randel and Wu 1996

<https://journals.ametsoc.org/doi/10.1175/1520-0469%281996%29053%3C2546%3AIOTOQI%3E2.0.CO%3B2>

Ploeger et al. 2019 <https://www.atmos-chem-phys.net/19/6085/2019/>

Ray et al. 2016 <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015JD024447>

Haenel et al. 2015 <https://www.atmos-chem-phys.net/15/13161/2015/acp-15-13161-2015.html>

Linz et al. 2017 <https://www.nature.com/articles/ngeo3013>

Abalos et al. 2014 <https://journals.ametsoc.org/doi/10.1175/JAS-D-13-0366.1>

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.

Abalos, M., Polvani, L., Calvo, N., Kinnison, D., Ploeger, F., Randel, W. and Solomon, S.: New Insights on the Impact of Ozone-Depleting Substances on the Brewer-Dobson Circulation, *J. Geophys. Res. Atmos.*, 124(5), 2435–2451, doi:10.1029/2018JD029301, 2019.

Andrews, D. G., Holton, J. R. and Leovy, C. B.: *Middle Atmosphere Dynamics*, International Geophysical Series, Vol. 40, 1987.

Butchart, N.: The Brewer-Dobson circulation, *Rev. Geophys.*, 52(2), 157–184, doi:10.1002/2013RG000448, 2014.

Butchart, N., Scaife, A. A., Bourqui, M., Grandpré, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D. and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, *Clim. Dyn.*, 27(7–8), 727–741, doi:10.1007/s00382-006-0162-4, 2006.

Butchart, N., Cionni, I., Eyring, V., Shepherd, T. G., Waugh, D. W., Akiyoshi, H., Austin, J., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deckert, R., Dhomse, S., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Li, F., Mancini, E., McLandress, C., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Sassi, F., Scinocca, J. F., Shibata, K., Steil, B. and Tian, W.: Chemistry–Climate Model Simulations of Twenty-First Century Stratospheric Climate and Circulation Changes, *J. Clim.*, 23(20), 5349–5374, doi:10.1175/2010JCLI3404.1, 2010.

Chandran, A. and Collins, R. L.: Stratospheric sudden warming effects on winds and temperature in the middle atmosphere at middle and low latitudes: a study using WACCM, *Ann. Geophys.*, 32(7), 859–874, doi:10.5194/angeo-32-859-2014, 2014.

Dietmüller, S., Garny, H., Plöger, F., Jöckel, P. and Cai, D.: Effects of mixing on resolved and unresolved scales on stratospheric age of air, *Atmos. Chem. Phys.*, 17(12), 7703–7719, doi:10.5194/acp-17-7703-2017, 2017.

Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., Mancini, E., Visioni, D., Stenke, A., Revell, L., Rozanov, E., Plummer, D. A., Scinocca, J., Jöckel, P., Oman, L., Deushi, M., Kiyotaka, S., Kinnison, D. E., Garcia, R., Morgenstern, O., Zeng, G., Stone, K. A. and Schofield, R.: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models, *Atmos. Chem. Phys.*, 18(9), 6699–6720, doi:10.5194/acp-18-6699-2018, 2018.

Engel, A., Möbius, T., Bönišch, 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.

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(7), 4783–4821, doi:10.5194/acp-19-4783-2019, 2019.

Garny, H., Birner, T., Bönisch, H. and Bunzel, F.: The effects of mixing on age of air, *J. Geophys. Res.*, 119(12), 7015–7034, doi:10.1002/2013JD021417, 2014.

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.

Hitchcock, P. and Simpson, I. R.: The Downward Influence of Stratospheric Sudden Warmings*, *J. Atmos. Sci.*, 71(10), 3856–3876, doi:10.1175/jas-d-14-0012.1, 2014.

Kunz, A., Pan, L. L., Konopka, P., Kinnison, D. E. and Tilmes, S.: Chemical and dynamical discontinuity at the extratropical tropopause based on START08 and WACCM analyses, *J. Geophys. Res. Atmos.*, 116(24), 1–15, doi:10.1029/2011JD016686, 2011.

Li, F., Newman, P., Pawson, S. and Perlwitz, J.: Effects of Greenhouse Gas Increase and Stratospheric Ozone Depletion on Stratospheric Mean Age of Air in 1960-2010, *J. Geophys. Res. Atmos.*, (2010), 2098–2110, doi:10.1002/2017JD027562, 2018.

Miyazaki, K., Iwasaki, T., Kawatani, Y., Kobayashi, C., Sugawara, S. and Hegglin, M. I.: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis data sets, *Atmos. Chem. Phys.*, 16(10), 6131–6152, doi:10.5194/acp-16-6131-2016, 2016.

Moore, F. L., Ray, E. A., Rosenlof, K. H., Elkins, J. W., Tans, P., Karion, A. and Sweeney, C.: A Cost-Effective Trace Gas Measurement Program for Long-Term Monitoring of the Stratospheric Circulation, *Bull. Am. Meteorol. Soc.*, 95(1), 147–155, doi:10.1175/BAMS-D-12-00153.1, 2014.

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., Maréchal, 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.

Morgenstern, O., Stone, K. A., Schofield, R., Akiyoshi, H., Yamashita, Y., Kinnison, D. E., Garcia, R. R., Sudo, K., Plummer, D. A., Scinocca, J., Oman, L. D., Manyin, M. E., Zeng, G.,

Rozanov, E., Stenke, A., Revell, L. E., Pitari, G., Mancini, E., Di Genova, G., Visioni, D., Dhomse, S. S. and Chipperfield, M. P.: Ozone sensitivity to varying greenhouse gases and ozone-depleting substances in CCM1 simulations, *Atmos. Chem. Phys.*, 18(2), 1091–1114, doi:10.5194/acp-18-1091-2018, 2018.

Neu, J. L. and Plumb, R. A.: Age of air in a “leaky pipe” model of stratospheric transport, *J. Geophys. Res. Atmos.*, 104(D16), 19243–19255, doi:10.1029/1999JD900251, 1999.

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, *J. Geophys. Res. Sp. Phys.*, 119(2), 1306–1324, doi:10.1002/2013JA019421, 2014.

Ploeger, F. and Birner, T.: Seasonal and inter-annual variability of lower stratospheric age of air spectra, *Atmos. Chem. Phys.*, 16(15), 10195–10213, doi:10.5194/acp-16-10195-2016, 2016.

Ploeger, F., Abalos, M., Birner, T., Konopka, P., Legras, B., Müller, R. and Riese, M.: Quantifying the effects of mixing and residual circulation on trends of stratospheric mean age of air, *Geophys. Res. Lett.*, 42(6), 2047–2054, doi:10.1002/2014GL062927, 2015a.

Ploeger, F., Riese, M., Haenel, F., Konopka, P., Müller, R. and Stiller, G.: Variability of stratospheric mean age of air and of the local effects of residual circulation and eddy mixing, *J. Geophys. Res. Atmos.*, 120(2), 716–733, doi:10.1002/2014JD022468, 2015b.

Ploeger, F., Legras, B., Charlesworth, E., Yan, X., Diallo, M., Konopka, P., Birner, T., Tao, M., Engel, A. and Riese, M.: How robust are stratospheric age of air trends from different reanalyses?, *Atmos. Chem. Phys. Discuss.*, 1–32, doi:10.5194/acp-2018-1281, 2019.

Polvani, L. M., Wang, L., Abalos, M., Butchart, N., Chipperfield, M. P., Dameris, M., Deushi, M., Dhomse, S. S., Jöckel, P., Kinnison, D., Michou, M., Morgenstern, O., Oman, L. D., Plummer, D. A. and Stone, K. A.: Large impacts, past and future, of ozone-depleting substances on Brewer-Dobson circulation trends: A multi-model assessment, *J. Geophys. Res. Atmos.*, 2018JD029516, doi:10.1029/2018JD029516, 2019.

Rosenlof, K. H.: Seasonal cycle of the residual mean meridional circulation in the stratosphere, *J. Geophys. Res.*, 100(D3), 5173, doi:10.1029/94JD03122, 1995.

Šácha, P., Eichinger, R., Garny, H., Pišoft, P., Dietmüller, S., de la Torre, L., Plummer, D. A., Jöckel, P., Morgenstern, O., Zeng, G., Butchart, N. and Añel, J. A.: Extratropical Age of Air

trends and causative factors in climate projection simulations, *Atmos. Chem. Phys. Discuss.*, (2018), 1–37, doi:10.5194/acp-2018-1310, 2019.

Seviour, W. J. M., Butchart, N. and Hardiman, S. C.: The Brewer-Dobson circulation inferred from ERA-Interim, *Q. J. R. Meteorol. Soc.*, 138(665), 878–888, doi:10.1002/qj.966, 2012.

Smith, A. K., Pedatella, N. M., Marsh, D. R. and Matsuo, T.: On the Dynamical Control of the Mesosphere–Lower Thermosphere by the Lower and Middle Atmosphere, *J. Atmos. Sci.*, 74(3), 933–947, doi:10.1175/JAS-D-16-0226.1, 2017.

SPARC: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. V. Eyring, T. Shepherd and D. Waugh (Eds.), SPARC Rep. No. 5, WCRP-30/2010, WMO/TD-No.40, available at www.sparc-climate.org/publications/sp [online] Available from: <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/>, 2010.

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.

Tweedy, O. V., Limpasuvan, V., Orsolini, Y. J., Smith, A. K., Garcia, R. R., Kinnison, D., Randall, C. E., Kvissel, O. K., Stordal, F., Harvey, V. L. and Chandran, A.: Nighttime secondary ozone layer during major stratospheric sudden warmings in specified-dynamics WACCM, *J. Geophys. Res. Atmos.*, 118(15), 8346–8358, doi:10.1002/jgrd.50651, 2013.

World Meteorological Organization (WMO): Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project – Report No. 58, Geneva, Switzerland, 2018., 2018.

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P. L., Rasch, P. J., Neubauer, D. and Lohmann, U.: Technical note: On the use of nudging for aerosol-climate model intercomparison studies, *Atmos. Chem. Phys.*, 14(16), 8631–8645, doi:10.5194/acp-14-8631-2014, 2014.

The effect of atmospheric nudging on the stratospheric residual circulation in chemistry-climate models

5 Andreas Chrysanthou¹, Amanda C. Maycock¹, Martyn P. Chipperfield¹, Sandip Dhomse¹, Hella Garny^{2,3}, Douglas Kinnison⁴, Hideharu Akiyoshi⁵, Makoto Deushi⁶, Rolando R. Garcia⁴, Patrick Jöckel², Oliver Kirner⁷, ~~Olaf Morgenstern⁸~~, Giovanni Pitari^{8,9}, David A. Plummer^{9,10}, Laura Revell^{10,11}, Eugene Rozanov^{11,12,123}, Andrea Stenke¹¹², Taichu Y. Tanaka⁶, Daniele Visioni¹³⁴ and ~~Yousuke Yamashita^{145,156}~~ and ~~Guang Zeng⁸~~

- 10 ¹ School of Earth and Environment, University of Leeds, Leeds, UK
² Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
³ Ludwig Maximilians University of Munich, Meteorological Institute Munich, Munich, Germany
⁴ National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA
⁵ National Institute of Environmental Studies (NIES), Tsukuba, Japan
⁶ Meteorological Research Institute (MRI), Tsukuba, Japan
15 ⁷ Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Karlsruhe, Germany
⁸ ~~Department of Physical and Chemical Sciences, Università dell'Aquila, L'Aquila, Italy~~ ~~National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand~~
⁹ ~~Department of Physical and Chemical Sciences, Università dell'Aquila, L'Aquila, Italy~~
^{9,10} Environment and Climate Change Canada, Climate Research Division, Montréal, QC, Canada
20 ^{10,11} School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand
¹¹² Institute for Atmospheric and Climate Science, ETH Zürich (ETHZ), Zürich, Switzerland
¹²³ Physical-Meteorological Observatory/World Radiation Center, Davos, Switzerland
¹³⁴ Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, USA
¹⁴⁵ Climate Modelling and Analysis Section, Center for Global Environmental Research, National Institute for Environmental
25 ~~Studies, Tsukuba, Japan~~
¹⁵⁶ Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan

Correspondence: Andreas Chrysanthou (eeac@leeds.ac.uk)

Abstract

We perform the first multi-model **inter**comparison of the impact of nudged meteorology on the
30 stratospheric residual circulation using hindcast simulations from the Chemistry-Climate Model
Initiative (CCMI). We examine simulations over the period 1980-2009 from **75** models in which the
meteorological fields are nudged towards a reanalysis dataset –and compare **these** with **their** equivalent
free-running simulations **and the reanalyses themselves.** ~~from 9 models.~~ We show that **for the current**
implementations, nudging meteorology does not constrain the mean strength of the stratospheric
35 residual circulation and the inter-model spread is similar, or even larger, than in the free-running
simulations. The nudged ~~simulations~~ **models generally show slightly** stronger upwelling in the tropical
lower stratosphere **compared to the free-running versions and exhibit marked differences compared to**
the **directly estimated** residual circulation ~~estimated directly from~~ **from** the reanalysis **dataset** they are
nudged towards. Downward control calculations **applied to the nudged simulations** reveal substantial
40 differences between the **climatological** lower stratospheric tropical upward mass flux (TUMF)
computed from the modelled wave forcing and that calculated directly from the residual circulation.
This explicitly shows that nudging decouples the wave forcing and the residual circulation, so that the
divergence of the angular momentum flux due to the mean motion is not balanced by eddy motions, as
would typically be expected in the time mean. Overall, nudging meteorological fields leads to increased

45 inter-model spread for most of the measures of the mean climatological stratospheric residual
circulation assessed in this study. ~~Although the mean circulation is poorly constrained~~In contrast, the
nudged simulations show a high degree of consistency in the interannual variability of the TUMF in the
lower stratosphere, which is primarily related to the contribution to variability from the resolved wave
forcing. The more consistent interannual variability in TUMF in the nudged models also compares more
50 closely with the variability found in the reanalyses, particularly in boreal winter. We apply a multiple
linear regression (MLR) model to separate the drivers of interannual and long-term variations in the
simulated TUMF; this ~~The MLR model~~ explains up to ~75% of the variance in TUMF in the nudged
simulations. ~~and~~The MLR model reveals a statistically significant positive trend in TUMF for most
models ~~in TUMF~~ over the period 1980-2009. The TUMF trend magnitude is generally larger in the
55 nudged models compared to their free-running counterparts, but the intermodel range of trends doubles
from around a factor of 2 to a factor of 4 due to nudging. Furthermore, the nudged models generally do
not match the TUMF trends in the reanalysis they are nudged toward for trends over different periods in
the interval 1980-2009. Hence, we conclude that nudging does not strongly constrain chemistry-climate
model (CCM) simulated long-term trends in the residual circulation. ~~Overall, nudging meteorological
60 fields leads to increased inter-model spread for most of the measures of the mean climatological
stratospheric residual circulation assessed in this study.~~Our findings show that while nudged
simulations may, by construction, produce accurate temperatures and realistic representations of fast
horizontal transport, this is not necessarily typically the case for the slower zonal mean vertical
transport in the stratosphere. Consequently, caution is required when using nudged simulations to
65 interpret the behaviour of long-lived stratospheric tracers that are ~~controlled~~ affected by the residual
circulation.

1 Introduction

The Brewer-Dobson circulation (BDC) is characterized by upwelling of air in the tropics, poleward
flow in the stratosphere, and downwelling at mid and high latitudes. The circulation can be separated
70 into two branches: the shallow branch in the lower stratosphere and the deep branch in the middle and
upper stratosphere (Plumb, 2002; Birner and Bönisch, 2011). The BDC affects the distribution of trace
species in the stratosphere, such as ozone, and its strength partly determines the lifetimes of long-lived
gases such as chlorofluorocarbons (CFCs; Butchart and Scaife, 2001). It also determines stratosphere to
troposphere exchange of ozone (Hegglin and Shepherd, 2009), which is important for the tropospheric
75 ozone budget (Wild, 2007). In the tropical lower stratosphere, where the photochemical lifetime of
ozone is long, variations and trends in the strength of the BDC are the main drivers of ozone within the
annual cycle (Weber et al., 2011), for interannual and longer term variability (Randel and Thompson,
2011) and in response to climate change (e.g. Keeble et al., 2017). ~~It is important to note that the overall
tracer transport in the stratosphere also accounts for the effect of turbulent eddy mixing, which has been~~

80 ~~evaluated separately in previous modelling studies~~. Here we focus on the advective part of the BDC, or the residual circulation, which is driven by wave breaking in the stratosphere from planetary scale Rossby waves and gravity waves (Holton et al., 1995). It is important to note that the overall tracer transport in the stratosphere is also affected by turbulent eddy mixing, which has been evaluated separately in previous studies (Garny et al., 2014; Ploeger et al., 2015a, 2015b; Dietmüller et al., 2018; 85 Eichinger et al., 2019; Šácha et al., 2019). The residual circulation is commonly evaluated in model (Butchart et al., 2010, 2011) and reanalysis (Abalos et al., 2015; Kobayashi and Iwasaki, 2016) studies using the Transformed Eulerian Mean circulation (TEM; Andrews and McIntyre, 1976, 1978; Andrews et al., 1987).

Past studies have shown substantial spread across **chemistry-climate** models (CCMs) in the mean 90 strength of the residual circulation (e.g. Butchart et al., 2010). Nevertheless, ~~climate and chemistry-climate models~~ (CCMs) consistently simulate a long-term strengthening of the residual circulation with an increase of $\sim 2\%$ decade⁻¹ (e.g. Butchart et al., 2010; Hardiman et al., 2014), though there are differences across models in the relative contribution to trends from resolved and parameterized wave forcings. ~~While subject to multiple caveats, Reanalysis~~ **reanalysis** datasets also suggest a strengthening 95 of the residual circulation over the past ~~several~~ **few** decades of the order 2-5% decade⁻¹ (Abalos et al., 2015; Miyazaki et al., 2016) apart from one (ERA-Interim) which shows a weakening ~~trend in~~ of the deep branch of the BDC (Seviour et al., 2012; Abalos et al., 2015). However, reanalyses are subject to multiple caveats, particularly in their suitability for trend studies, and there can be substantial differences in residual circulation trends calculated from the same reanalysis using different methods 100 (Abalos et al., 2015).

Given the limitations of reanalyses, evaluating the fidelity of model estimates of residual circulation variability and trends is challenging since there are no direct measurements of the residual circulation. The only direct estimates of the stratospheric mass circulation come from tracer measurements, which can be used to calculate stratospheric age-of-air (AoA) (Kida, 1983; Schmidt and Khedim, 1991; 105 Waugh and Hall, 2002). AoA represents the combined effects of advection and mixing processes, and as such cannot be directly related to the residual circulation. While progress has been made in separating the relative effects of advection and mixing for AoA calculated from models (Garny et al., 2014; Dietmüller et al., 2018; Eichinger et al., 2019; Šácha et al., 2019), from Lagrangian models driven by reanalysis data (Ploeger et al., 2015a, 2015b, 2019; Ploeger and Birner, 2016), and comparing the effects in both CCMs and Lagrangian models (Dietmüller et al., 2017), this is more difficult to achieve 110 in observations. ~~Comparing model and reanalysis estimates of BDC variability and trends with observations is challenging since there are no direct measurements of the residual circulation and there are limited tracer measurements from which age of air (AoA) can be estimated (Waugh and Hall, 2002).~~ Engel et al. (2009) ~~and Stiller et al. (2012)~~ used balloon-borne measurements of stratospheric 115 trace gases and found ~~find~~ a statistically ~~non~~-significant increase in AoA in the middle stratosphere at

northern midlatitudes; this has been corroborated in a more recent study using longer –measurement records at two midlatitude sites in the Northern Hemisphere (Engel et al., 2017). ~~which could suggest the deep branch of the residual circulation is weakening. In contrast, CCMs forced with observed sea-surface temperatures (SSTs) show a decrease in AoA throughout the depth of the stratosphere in northern midlatitudes (World Meteorological Organization, 2018).~~ It has been hypothesized based on analyses of recent satellite tracer datasets, which have greater spatial and temporal coverage, that subtropical AoA trends can be explained by a weakening of the mixing barriers at the edge of the tropical pipe (Neu and Plumb, 1999) that is masking the effects of an increase in tropical upwelling on AoA (Stiller et al., 2012; Haenel et al., 2015). In contrast with AoA trends derived from observations, CCMs forced with observed sea-surface temperatures (SSTs), greenhouse gases and ozone-depleting substances show a decrease in AoA throughout the stratosphere (Karpechko and Maycock, 2018; Li et al., 2018; Morgenstern et al., 2018; Abalos et al., 2019; Polvani et al., 2019). Theoretical approaches based on the tropical leaky pipe model (Neu and Plumb, 1999) have shown promise for bridging the information on the stratospheric circulation derived from observations with outputs from general circulation models (GCMs)/CCMs (Ray et al., 2016), but differences remain (Karpechko and Maycock, 2018).

More recent theoretical developments offer a means of calculating the diabatic circulation using stratospheric tracers (Linz et al., 2017), which is a promising avenue as this is more closely related to the residual circulation than AoA. Linz et al. (2017) showed consistent estimates of the diabatic circulation in the lower stratosphere based on two independent satellite tracer datasets but identified large uncertainties of up to a factor of two in the mean circulation strength in the upper stratosphere. Hence, the available tracer datasets are not yet suitable for characterizing trends in the diabatic circulation using these methods. Targeted measurement strategies to better characterize long-term changes in the stratospheric meridional circulation have been proposed (Moore et al., 2014; Ray et al., 2016).

In an attempt to obtain a closer comparison with observed stratospheric trace species, some studies have used model simulations with meteorological fields nudged or relaxed towards analysis or reanalysis datasets (Jeuken et al., 1996).; These include ~~many~~ studies of stratospheric ozone variability and trends (e.g. van Aalst et al., 2004; Solomon et al., 2016; Hardiman et al., 2017b; Ball et al., 2018), comparisons between models and satellite-based multi-species observational records (Froidevaux et al., 2019) and in particular, focusing on specific meteorological events such as the Sudden Stratospheric Warming in the 2009/2010 winter (Akiyoshi et al., 2016), ~~and~~ as well as the chemical and climatic effects of volcanic eruptions (Löffler et al., 2016; Solomon et al., 2016; Schmidt et al., 2018). Nudged simulations have also been used to study mechanisms for dynamical coupling between the stratosphere and troposphere (Hitchcock and Simpson, 2014) and to examine the effects of different regions on atmospheric predictability (e.g. Douville, 2009; Jung et al., 2010). Nudging involves adding additional

tendencies to the model equations to constrain the modeled variables ~~to reanalysis fields~~. Nudged variables can include horizontal winds (or divergence and vorticity), temperature, surface pressure, latent and sensible heat fluxes. However, vertical winds, which are a small residual from horizontal divergence, are not nudged and the underlying model physics can yield quite different results from the reanalysis datasets that they are nudged towards (Telford et al., 2008; Hardiman et al., 2017). ~~A recent study by Orbe et al. (2018) analyzed tropospheric tracers in nudged CCM simulations and found large differences in the distributions of the tracers, which could be partly traced to differences in the model convection schemes. They urged users to adopt a cautious approach when interpreting tracers in nudged simulations given their dependence on not only large-scale flow but also sub-grid parameterizations.~~

The approach of nudging a CCM towards reanalysis data follows a similar philosophy to traditional off-line chemical transport models (CTM), although there are fundamental differences between these types of model in terms of their tracer advection. CTMs need to match the mass transport with the evolution of the pressure field. This can be done exactly in isobaric coordinates (often used in the stratosphere) but requires a correction in regions where grid box mass changes (e.g. as surface pressure changes). CCMs are less affected by this mass-wind inconsistency than CTMs (Jöckel et al., 2001), but nudging will add forcings which may that are be inconsistent with the model state. CTMs use the full 3-D circulation from the (re-)analyses directly and have been widely developed and used over the past few decades (e.g. Rood et al., 1988; Chipperfield et al., 1994; Lefèvre et al., 1994). They have proven to be very successful at simulating stratospheric tracers on a range of timescales (Chipperfield, 1999), including decadal changes (Mahieu et al., 2014). However, this success has been built on extensive testing of the optimum way to use the reanalysis data to force the CTMs. For example, Chipperfield (2006) showed how different approaches to calculating the vertical velocity in the TOMCAT/SLIMCAT model could lead to very different distributions of stratospheric age of air, while Monge-Sanz et al.; (2013a) compared the performance of different European Centre for Medium-Range Weather Forecasts (ECMWF) analyses within the same CTM framework. Krol et al.; (2018) recently provided a summary of how current CTMs intercompare for tracer calculations. Monge-Sanz et al.; (2013b) compared the approaches of using ECMWF analyses directly in a CTM with the ECMWF CCM nudged using the same analyses. They found that the CTM and nudged CCM were consistent in showing ~~athe~~ degraded performance when using older ERA-40 reanalysis compared to the later ERA-Interim. However, they also showed some differences between CTM and nudged-CCM tracers using the same analyses, with the nudged CCM showing stronger upward motion in the tropical stratosphere. ~~Furthermore,~~ Therefore, with regards to the slow residual circulation, one cannot assume that a nudged CCM will behave in a similar way to a CTM even when using the same meteorological analyses.

Recently, Ball et al.; (2018) showed 2 nudged CCMs which failed to capture the observed variations in the lower stratospheric ozone as measured by satellite observations, while Chipperfield et al.; (2018) using the TOMCAT CTM simulated a better agreement of modeled ozone variations with the

observations. Overall, the success of some CTM simulations in simulating long-lived stratospheric tracers has been built on many years of model development and testing. In contrast, nudged CCMs are much newer tools and have not yet been evaluated to the same extent. A recent study by Orbe et al. (2018) analyzed tropospheric tracers in nudged CCM simulations and found large differences in the distributions of the tracers, which could be partly traced to differences in the model convection schemes. They urged users to adopt a cautious approach when interpreting tracers in nudged simulations given their dependence on not only large-scale flow but also sub-grid parameterizations. However, a critical evaluation of the stratospheric residual circulation in nudged CCM simulations has been lacking to date. Furthermore, with regards to the slow residual circulation, one cannot assume that a nudged CCM will behave in a similar way to a CTM even when using the same meteorological analyses.

To examine the effect of nudging on the stratospheric residual circulation this study we compares hindcast simulations from free-running and nudged versions of the same models that participated in the phase 1 of the Chemistry-Climate Model Initiative (CCMI; Morgenstern et al., 2017). Nudged experiments were not performed in previous chemistry-climate multi-model comparisons (Chemistry-Climate Model Validation Activity 2; CCMVal-2), so CCMI offers a timely opportunity to evaluate the effect of nudging on simulated mean biases, variability and long-term trends in the residual circulation. For completeness, we also present a comparison between the nudged simulations and the reanalysis datasets the models are nudged towards. The manuscript is laid out as follows: Section 2 describes the CCMI and the reanalysis data used in the present study along with the diagnostics for the residual circulation, section 3 presents results covering the mean circulation, annual cycle, interannual variability and trends, and section 4 summarizes the results and discusses the implications for using nudged simulations to study aspects of the observational record.

2 Data and Methods

2.1 Models and experiments

CCMI is the successor activity to CCMVal-2 and the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP; Lamarque et al., 2013). We use the hindcast free-running simulations, REF-C1-(C1), and the nudged specified dynamics (SD)-simulations, REF-C1SD, which cover the periods 1960-2009+0 and 1980-2009+0, respectively. Here we analyze the common 30-year period 1980-2009 that was run by all models for both experiments Both experiments are run with prescribed observed SSTs and sea ice concentrations. The CCMI data were downloaded from the British Atmospheric Data Centre (Hegglin and Lamarque, 2015). For an extensive overview of the CCMI models see Morgenstern et al. (2017). We analyze those CCMI models that output the necessary TEM diagnostics. At a minimum, this requires the residual vertical velocity (\bar{w}^*) and the residual meridional velocity (\bar{v}^*) (Andrews et al., 1987); where available we also use the resolved and

parametrized wave forcing fields from the models. ~~We use~~ This gives results from a total of 9-10 models which differ from one another in various aspects such as their horizontal resolution, ~~which ranges~~ from 1.9° to 5.6°, their vertical resolution as well as their sub-grid parameterizations (see Table 1). The main text concentrates on the 7 out of 10 models that performed both the REF-C1 and REF-C1SD experiments (Table 2). However, the broad conclusions drawn in the main text for the characteristics of the 7-member REF-C1 ensemble are consistent with the behavior for all 10 models. Hence the 3 models that only performed the REF-C1 experiment (GEOSCCM, NIWA-UKCA and ULAQ-CCM) are not discussed further, but for completeness a subset of diagnostics from those models is shown in the supplement (Supplementary Figures S1-S5).

For the REF-C1 simulations we analyze between 1-5 ensemble members (depending on what was available) and for REF-C1SD the one realization submitted from each model. The REF-C1SD simulations nudge temperature and other meteorological fields such as horizontal winds, vorticity and divergence and some surface fields (Table 2), while the chemical fields are left to evolve freely. The nudging timescales range from 6 - 50 hours and the height range over which nudging is applied varies (Table 2). ~~The REF-C1SD simulations nudge meteorological fields such as horizontal winds and temperature, while the chemical fields are left to evolve freely (Table 1).~~ The TEM and related diagnostics that were available from each model are shown in Table 3. The models use different reanalysis fields for nudging taken from ERA-Interim (Dee et al., 2011), JRA-55 (Ebita et al., 2011; Kobayashi et al., 2015) or MERRA (Rienecker et al., 2011). The differences in the residual circulation diagnosed from ~~the~~ reanalyses have been identified and documented in previous studies (e.g. Abalos et al., 2015).

2.2. Model diagnostics

2.2.1 TEM residual circulation

The TEM velocities (\bar{v}^* , \bar{w}^*) are defined as (Andrews et al., 1987):

$$\bar{v}^* = -\frac{1}{\rho_0 \cdot a \cdot \cos\phi} \frac{\partial \bar{\Psi}^*}{\partial z}, \quad \bar{w}^* = \frac{1}{\rho_0 \cdot a \cdot \cos\phi} \frac{\partial \bar{\Psi}^*}{\partial \phi}, \quad (1)$$

where $\bar{\Psi}^*(\phi, z)$ is the residual meridional mass streamfunction, ρ_0 is ~~the~~ log-pressure density, a is ~~the~~ Earth's radius and ϕ is ~~the~~ latitude. As most of the models analyzed here use a hybrid-pressure vertical coordinate, the ~~primary prognostic~~ variable is the pressure vertical velocity, $\bar{\omega}^*$ calculated in $Pa s^{-1}$, which must be converted to $m s^{-1}$ in order to get the residual vertical velocity, \bar{w}^* . ~~The Converting conversion of ω to w in isobaric coordinates~~ is given by the following equation:

$$\omega = \frac{dp}{dt} = \frac{\partial z}{\partial t} \frac{\partial p}{\partial z} = w \frac{-pg}{RT} = w \frac{-p}{H}, \quad (2)$$

where p is pressure, $R = 287 \text{ J K}^{-1} \text{ kg}^{-1}$ is the gas constant for dry air and H is a fixed scale height. Both
255 TEM velocity components were submitted as monthly mean fields to the CCMI data archive (~~Hegglin
and Lamarque, 2015~~). Upon close examination of the CCMI model output, some discrepancies were
found in the way that the residual vertical velocity was calculated among the models. Although a fixed
scale height of $H = 6950 \text{ m}$ was recommended in the CCMI data request (Eyring et al., 2013; Hegglin
and Lamarque, 2015), the ~~TEM model~~ output from some models (EMAC and ~~SOCOL~~~~SOCOL3~~) was
260 calculated incorrectly using a temperature-dependent density, $\rho_0 = p/RT$, instead of the log-pressure
~~definition of the~~ density $\rho_0 = \rho_s \cdot e^{-z/H}$, such that z has a unique 1:1 correspondence with p . This
methodological error leads to artificial spread in the model \bar{w}^* fields (~~see Dietmüller et al., 2018~~). We
note that previous multi-model comparisons of the residual circulation that use \bar{w}^* taken directly from
models may have been subject to the same issue (e.g. Butchart et al., 2010; SPARC, 2010), though we
265 cannot confirm this (~~e.g. Butchart et al., 2010; SPARC, 2010~~). To avoid this methodological
inconsistency, Dietmüller et al., (2018) recalculated \bar{w}^* from \bar{v}^* using the continuity equation, which
~~consists of~~ requires a vertical integration and a derivative along the meridional direction. The
recalculation of \bar{w}^* from \bar{v}^* ~~was also been~~ explored for this study, but it was found to introduce
additional errors affecting the latitudinal structure of \bar{w}^* (not shown), specifically because of the reduced
270 number of CCMI requested pressure levels compared to the native model levels. We were able to
overcome ~~this~~ discrepancy in the submitted \bar{w}^* fields for the EMAC simulations by reconverting high
frequency $\bar{\omega}^*$ output to \bar{w}^* using the log-pressure density ~~as in equation 2, using the log-pressure
density~~. However, for ~~SOCOL3~~~~SOCOL3~~ the required output for ~~this~~ ~~was not possible~~ was not
available and hence we use the submitted \bar{w}^* for which the absolute values ~~for that model~~ should be
275 treated with caution. For ~~the other the~~ models ~~other than EMAC~~, the results presented in this study are
based on the original diagnostics submitted to the CCMI data archive, which we have verified were
calculated in the correct way.

We ~~also~~ compute the mass flux across a given pressure surface as (Rosenlof, 1995):

$$2\pi \int_{\phi}^{\text{pole}} \rho_0 a^2 \cos\phi \bar{w}^* d\phi = 2\pi a \Psi^*(\phi), \quad (3)$$

280 using the boundary condition that $\Psi = 0$ at the poles. By finding at each pressure level the latitude at
which Ψ_{max} and Ψ_{min} occur, which correspond to the height-dependent turnaround (TA) latitudes, we
can calculate the net downward mass flux in each hemisphere. The net tropical upward mass flux, equal
to the sum of the downward mass fluxes in each hemisphere, can then be expressed as (Rosenlof, 1995):

$$\text{Tropical Upward Mass Flux (TUMF)} = 2\pi a (\Psi_{max}^* - \Psi_{min}^*) \quad (4)$$

285

The TUMF has been used widely as a measure of the strength of the BDC (e.g. Rosenlof, 1995; Butchart and Scaife, 2001; Butchart et al., 2006, 2010, 2011; Butchart, 2014 and references therein; Seviour et al., 2012), so its use here enables a direct comparison with earlier studies. Arguably, the strength of the TUMF is a first-order metric for evaluating changes in the stratospheric mass circulation as a consequence of nudging. As mentioned above, by calculating the annual means of TUMF accounting for the seasonal cycle of the TA latitudes we capture the correct evolution of the intraseasonal (not shown) and interannual variability in the TUMF.

2.2.2 Downward control principle calculations

Under steady-state conditions, the $\bar{\Psi}^*(\phi, z)$ at a specified latitude ϕ and at a log-(pressure)-height z is given by the vertically integrated eddy-induced total zonal forces, \bar{F} , above that level (Haynes et al., 1991):

$$\Psi^*(\phi, z) = \int_z^\infty \left\{ \frac{\rho_0 a^2 \bar{F} \cos^2 \phi}{\bar{m}_\phi} \right\}_{\phi=\phi(z)} dz, \quad (5)$$

where in the quasi-geostrophic limit $\bar{m}_\phi \approx -2\Omega a^2 \sin\phi \cos\phi$. The above integration occurs along lines of constant zonal-mean absolute angular momentum per unit mass, $\bar{m} = a \cos\phi (\bar{u} + a\Omega \cos\phi)$ where \bar{u} is the zonal mean zonal wind and Ω corresponds to Earth's rotation rate, with boundary conditions of $\Psi \rightarrow 0$ and $\rho_0 \bar{w}^* \rightarrow 0$ as $z \rightarrow \infty$. These lines of constant angular momentum are almost-approximately vertical apart from except near the equator (up to $\sim \pm 20^\circ$) such that we can approximate-calculate the solution of the above integral using constant ϕ for the limits of the integral (Haynes et al., 1991). In climate model-simulations, \bar{F} corresponds to has contributions from resolved waves due to the divergence-of-Eliassen-Palm flux divergence (EPFD) and/or contributions from parameterized gravity wave drag due to sub-grid scale waves that originate from orography, convection and frontal instabilities. This enables us to estimate the contribution to the tropical upward mass flux of both resolved planetary wave driving (EPFD) along-with and the orographic (OGWD) and non-orographic (NOGWD) parameterized gravity wave drag from the CCM1 model output (Table 2) to the tropical upward mass flux and compare with the direct estimates derived from the residual vertical velocity \bar{w}^* .

Applying the downward control principle (DCP Haynes et al., 1991) can provide useful insights into the driving mechanisms of the stratospheric residual circulation and therefore explain part of the inter-model spread found in both REF-C1 and REF-C1SD simulations. While the downward control principle DCP enables the contributions of EPFD and OGWD/NOGWD to TUMF to be calculated under

various assumptions (Haynes et al., 1991), one has to keep in mind that the different wave forcings can interact and thus are not independent of each other (Cohen et al., 2013).

It is important to note some possible limitations of the diagnostic approaches chosen for this study. Both the direct and downward control principle methods rely on the applicability of quasi-geostrophic theory to interpret the results. In addition to the two approaches used here, the residual circulation can also be estimated using the thermodynamic equation. Studies have shown that the estimates from the different methods for evaluating the residual circulation can differ (Seviour et al., 2012; Abalos et al., 2015; Linz et al., 2019), particularly in reanalyses where standard global conservation laws (e.g. conservation of mass) are generally not required to be met. Similar issues are likely to beset the nudged model simulations owing to the additional tendencies included in the model equations. The differences between the calculation methods for the residual circulation can be as large as, or larger than, the differences between reanalysis datasets for the same diagnostic (Abalos et al., 2015; Linz et al., 2019), and may further depend on choices around averaging between fixed latitudes or the TA latitudes (Linz et al., 2019), so it is important to bear this in mind in interpretation of the results presented here. Unfortunately, heating rates were not available from all CCMI model simulations to perform the thermodynamic equation calculation. Nevertheless, we compute the direct and downward control principle diagnostics for the residual circulation in a self-consistent manner in the models and reanalyses to enable comparison with earlier multi-model studies (Butchart et al., 2006, 2010; SPARC, 2010).

2.3 Multiple linear regression model

To investigate the drivers of interannual variability in the residual circulation we apply a multiple linear regression (MLR) model (equation 6) to the annual mean timeseries of TUMF. The model includes terms for known drivers of variations in tropical lower stratospheric upwelling: major volcanic eruptions (Pitari and Rizi, 1993), El Nino Southern Oscillation (ENSO) (García-Herrera et al., 2006; Marsh and Garcia, 2007; Randel et al., 2009), the Quasi-Biennial Oscillation (QBO) (Baldwin et al., 2001) and a long-term linear trend (Calvo et al., 2010).

$$TUMF(t) = \beta_0 + \beta_{VOL} \cdot x_{VOLC}(t) + \beta_{ENSO} \cdot x_{ENSO}(t) + \beta_{TREND} \cdot x_{TREND}(t) + \beta_{QBO1} \cdot x_{QBO1}(t) + \beta_{QBO2} \cdot x_{QBO2}(t) + \varepsilon(t),$$

(6)

where β_0 is a constant, β_i is the regression coefficient for basis function x_i and $\varepsilon(t)$ is the residual. Following Maycock et al., (2018), the volcanic basis function is defined as the tropical lower stratospheric average volcanic surface area density (SAD), the ENSO term-basis function is the timeseries of based-on-east-central equatorial Pacific Ocean SST timeseries-anomalies (Niño 3.4 index; 5°S to 5°N; 170°W to 120°W), the two orthogonal-QBO terms are the first two principal component

timeseries from an empirical orthogonal function (EOF) analysis of the zonal mean zonal winds between 10°S-10°N and 70 to 5 hPa, and a linear trend. The first three regressors, volcanic, ENSO and the linear trend are identical for both REF-C1 and REF-C1SD runs, while the QBO terms are calculated using the model winds for each model and experiment. For the REF-C1 runs, CMAM does not include a QBO, hence when we apply the MLR to the CMAM REF-C1 simulation the QBO terms are omitted. We opted not to include an equivalent effective stratospheric chlorine (EESC) MLR term to account for the changes in ozone depleting substances (Abalos et al., 2019; Morgenstern et al., 2018; Polvani et al., 2018, 2019) as the period considered in the study is may not be sufficiently long for the linear trend to be separated properly from EESC. Since we are regressing annual mean TUMF we do not consider a seasonal cycle term or any lag in the terms. The results in section 3.5 focus on the first ensemble member (in the rip-nomenclature, where r stands for realization, i for initialization and p for physics - r1i1p1), but where applicable the results from the MLR model for the rest of the ensemble members of the REF-C1 runs are presented in the supplement (Supplementary Figures S67-S12S9).

2.4 Reanalysis Data

In order to compare the REF-C1 and REF-C1SD simulations against the reanalysis datasets used for the nudging, we use the SPARC Reanalysis Intercomparison Project (S-RIP) dataset (Martineau, 2017; Martineau et al., 2018). This provides a common gridded version of the reanalysis TEM fields on a 2.5° × 2.5° grid up to 1 hPa. The pressure vertical velocity, \bar{w}^* , is converted to the residual vertical velocity, \bar{w}^* , using equation 2. A detailed comparison of the stratospheric residual circulation in reanalysis datasets has been given by Abalos et al., (2015).

375 3 Results

3.1 Climatological residual circulation: \bar{w}^*

Figure 1 shows a latitude-pressure cross-sections of the climatological (1980-2009) multi-model mean (MMM) annual mean \bar{w}^* for the REF-C1 (Fig. 1a) and REF-C1SD (Fig. 1b) simulations and their absolute differences (Fig. 1c). In Figure 1c, absolute differences are computed, so positive values indicate where the magnitude of the circulation in REF-C1SD (whether upwelling or downwelling) is larger than in REF-C1. As expected, the climatologies show upwelling occurs in the tropics between around 30°S to 30°N and downwelling at higher latitudes. In the lowermost stratosphere (100-80 hPa), within the region of tropical upwelling, the REF-C1SD runs-MMM generally shows larger \bar{w}^* values in the low to mid stratosphere in the subtropics and smaller values at the equator compared to REF-C1, indicating a tendency for a more double peaked \bar{w}^* structure in the tropics in the lowermost stratosphere (Ming et al., 2016a). Above this, between ~70-4 hPa, the REF-C1SD MMM shows on average stronger

upwelling at the equator compared to REF-C1, indicating a less pronounced double peaked \bar{w}^* structure in the REF-C1SD experiments in the lower to middle stratosphere. ~~In the upper stratosphere between 4-2 hPa, REF-C1 shows a slightly narrower and more peaked region of upwelling than REF-C1SD, and vice versa near 1 hPa.~~ Between 1-2 hPa, the REF-C1SD MMM shows larger interhemispheric asymmetry than in REF-C1, with stronger upwelling in the northern tropics (Figure 1b). ~~In the~~ At southern mid-latitudes, between $\sim 30-65^\circ\text{S}$, the REF-C1SD runs-MMM exhibits on average slightly ~~stronger~~-weaker downwelling than in ~~the~~ REF-C1 simulations, with the largest magnitude differences found in the upper stratosphere. ~~except in the upper stratosphere.~~ In the polar and subpolar regions of ~~both hemispheres~~ Arctic throughout the stratosphere, the picture reverses as the REF-C1SD runs-MMM shows significantly stronger downwelling over the poles than in REF-C1. In the Antarctic the picture is more complex, with the REF-C1SD MMM showing weaker downwelling right at the pole in the upper stratosphere (2-10 hPa), but stronger downwelling between around $75-88^\circ\text{S}$ ~~and weaker downwelling near the edge of the polar region ($60-70^\circ$).~~ In the middle stratosphere, from 50-10 hPa, the REF-C1SD MMM shows stronger downwelling between $60-80^\circ\text{S}$.

We note these differences are rather similar when the MMM for the REF-C1 simulations is calculated using only the same 6 simulations as used in the REF-C1SD calculation (Supplement Figure S1) with some differences in the tropical upper stratosphere within the upwelling region (not shown). To show the differences in the transition between regions of upwelling and downwelling motion, Figure 2 shows vertical profiles of the climatological annual mean turnaround (TA) latitudes in each hemisphere for the REF-C1 and REF-C1SD MMM and the three reanalysis datasets used for nudging. ~~Additionally, the climatological TA latitudes, when calculated for the above mentioned subset of REF-C1 MMM, reveal significant differences throughout the depth of the stratosphere when compared with the MMM for the REF-C1 and REF-C1SD runs respectively (Supplement Figure S2).~~ Note that since five of the REF-C1SD models were nudged towards ERA-I, the REF-C1SD MMM may be more weighted towards ERA-I than the other reanalyses. In the Northern hemisphere (NH), ~~Specifically, the~~ the REF-C1SD runs MMM simulate on average shows a wider more poleward Northern Hemisphere (NH) TA latitude compared to both REF-C1 and the reanalyses throughout almost all the entire depth of the stratosphere (Figure 2b). ~~than the REF-C1 runs while in the Southern Hemisphere (SH) TA latitude~~ A more poleward TA latitude for REF-C1SD than in both REF-C1 and the reanalyses is also found in the Southern hemisphere (SH) at pressures greater than 30 hPa (Figure 2a). Hence the nudged simulations show, on average, a wider region of tropical upwelling in the lower stratosphere compared to their free-running counterparts by up to around 5° latitude. In the middle and upper stratosphere the, ~~the picture reverses in the middle and upper stratosphere (Supplement Figure S2).~~ REF-C1SD MMM shows a narrower upwelling region in the SH. Interestingly, above 10 hPa in the SH (Figure 2a) the REF-C1SD does not show a progressive widening of the upwelling region with decreasing pressure as seen in the reanalyses. This is reflected in the structural differences in \bar{w}^* in the SH upper stratosphere found in

some models (Supplementary Figure S10). It should be noted though that the differences in TA latitudes between the REF-C1 and REF-C1SD MMMs are comparable to the differences found between the three reanalysis datasets. Hence, nudging meteorology significantly affects the strength and structure of the climatological residual circulation throughout the stratosphere

Focusing on the lower stratosphere, Figure 32 shows the climatological annual mean \bar{w}^* at 70 hPa in the individual models for the (a) REF-C1 and (b) REF-C1SD simulations and (c) their differences. Also plotted in Figure 32b are \bar{w}^* estimates from the three reanalysis datasets reanalyses used by the nudged and Figure 3d shows the difference between each REF-C1SD simulation and the reanalysis it was nudged towards models. The nudged runs on average show a wider region of tropical upwelling especially in the lower stratosphere compared to their free-running counterparts (Supplement Figure S2). Within the upwelling region, the vast majority of all the models show a clear double peaked \bar{w}^* structure in the tropics with the exception of the CCSRNIES-MIROC3.2 and MRI-ESM1r1 models in the REF-C1SD experiment. In those two cases, CCSRNIES-MIROC3.2 simulates a tri-modal \bar{w}^* structure while MRI-ESM-r1 shows a relatively constant \bar{w}^* across the tropics. For the REF-C1 experiment, with ULAQ-CCM being a notable outlier as it exhibits a rather different and single-peaked maximum and NIWA-UKCA showing a relative flat and not clearly defined structure. Both EMAC simulations, along with CMAM and SOCOL3SOCOL3 models, show a narrower double-peaked structure, with EMAC-L47 exhibiting a rather pronounced NH subtropical-~~local~~ maximum. Conversely, The remainder of the REF-C1 simulations exhibit a double-peaked structure as well, but to a certain degree more symmetric, albeit with various amplitudes. CESM1-WACCM simulates a the broadest ~~double-peaked structure~~ region of tropical upwelling in the lower stratosphere, with the SH ~~local~~-subtropical maximum occurring at higher latitudes compared with the rest of the models. The other REF-C1 simulations also exhibit a double peaked \bar{w}^* structure, which is generally more hemispherically symmetric, but with varying amplitudes.

A double-peaked \bar{w}^* structure in the lower stratosphere has previously been shown in reanalysis datasets (Abalos et al., 2015; Ming et al., 2016) and some CCMs (Butchart et al., 2006, 2010). This can also be seen in Figure 32b for the three reanalysis datasets (ERA-I, JRA-55 and MERRA), where ERA-I and JRA-55 show an asymmetrical double-peaked structure with stronger upwelling in the NH ~~local~~ subtropics compared to the SH maximum. As documented by Abalos et al. (2015), based on the direct calculation from of the residual circulation definition, MERRA exhibits downwelling at the equator, an issue which was highlighted in Abalos et al., (2015) and linked manifested as with a negative cell in the streamfunction.

Figure 32c shows the absolute differences in ~~70 hPa~~ \bar{w}^* at 70 hPa between the REF-C1SD and REF-C1 experiments for the 6 models that ran both experiments. Positive values show where the ~~As in Figure 1c~~ positive differences indicate that the magnitude of the circulation in REF-C1SD is larger than

in REF-C1. The largest differences are generally found within the inner tropics, where CCSRNIES-MIROC3.2 and CMAM and MRI-ESM1r1 exhibit significantly stronger upwelling (up to 3 times more for CMAM) near the local \bar{w}^* minimum at the equator. Similarly, for MRI-ESM1r1, which was nudged towards JRA-55, the differences near the equator reflect a reduction in the double-peaked structure of \bar{w}^* in the REF-C1SD experiment. Similarly, EMAC-L90 differences are a manifestation of having a more double-peaked structure in their REF-C1 simulation. EMAC-L47 and CESM1-WACCM although nudged towards ERA-I and MERRA, respectively, show similar and relatively small differences between REF-C1SD and REF-C1 both within the upwelling region and the downwelling regions in both hemispheres. There are also larger differences in many models near edges of the upwelling region in the SH mid-latitudes (30°S - 40°S), which REF-C1 shows stronger downwelling in the mid-latitudes (up to 50° in both hemispheres) compared to the nudged models reflect differences in the width of the tropical pipe between the free running and nudged simulations (-Figure 2 and Section 3.3). Around the subpolar and. In the subpolar and polar latitudes of the SH, the majority of the nudged REF-C1 models simulate stronger downwelling than their nudged free-running counterparts, while in the NH extratropics no consistent picture emerges across the models. EMAC-L47 and EMAC-L90 show markedly different behaviours despite the fact they are nudged towards the same reanalysis (ERA-I) and differ only in their vertical resolution. This indicates that the effect of nudging on the mean residual circulation is likely to be sensitive to a great many factors that vary from model to model.

Another interesting key result from Figure 32 is that the inter-model spread in \bar{w}^* for both experiments is larger in the NH downwelling region than in the equivalent region of the SH. Specifically, the inter-model spread is 0.145 mm s⁻¹ for the REF-C1 runs for all points between 30°S - 80°S and 0.22 mm s⁻¹ for points between 30°N - 80°N, while for REF-C1SD the values are 0.124 mm s⁻¹ and 0.198 mm s⁻¹, respectively. This also demonstrates that the inter-model spread in \bar{w}^* in the REF-C1SD simulations is comparable to that in REF-C1 exhibitslightly larger inter-model spread in \bar{w}^* in at the extratropical latitudes and is also true when taking into account the same 6 simulations as used in the REF-C1SD experiment. In contrast, in the tropics between 30°S - 30°N the REF-C1SD simulations exhibit a slightly larger inter-model spread than the free-running simulations (0.09 mm s⁻¹ vs. 0.07 mm s⁻¹). All the above show that nudging, as applied in these simulations, rather weakly constrains the mean amplitude and structure of the residual circulation.

Comparing the SD models against the reanalysis datasets they are nudged towards reveals some distinct differences. For example, CESM1-WACCM is nudged towards MERRA as shown in Figure 2b, but it does not show downwelling around the equator as the direct MERRA estimate, while it does resemble the reanalysis in the northern subtropics quite closely. MRI-ESM1r1 in the REF-C1SD experiment is nudged towards JRA-55 and shows a rather flat latitudinal \bar{w}^* structure with no clearly defined local maxima and does not resemble the reanalysis in the inner tropics. As for the models

nudged towards ERA-I, CCSRNIES MIROC3.2 simulates a largely dissimilar structure to the reanalysis, exhibiting a single peaked structure without capturing the tropical maxima asymmetry, which is a feature of EMAC-L47 and CMAM simulations of the REF-C1SD simulations to a lesser degree. Figure 3d shows the absolute differences in \bar{w}^* between the REF-C1SD simulations and the respective reanalysis dataset used for nudging. In the upwelling region, the REF-C1SD experiments generally show stronger upwelling near the equator than in the reanalyses. Although CESM1-WACCM is nudged towards MERRA, it does not simulate downwelling at the equator as seen in the MERRA direct estimate. The relatively larger \bar{w}^* differences near 10--15°N in CCSRNIES MIROC3.2, EMAC-L90 and MRI-ESM-r1 reflect a lack of inter-hemispheric asymmetry in the double peaked \bar{w}^* structure in the REF-C1SD experiment compared to the reanalyses. Outside of the tropics, the REF-C1SD experiments generally show weaker downwelling in the NH mid-latitudes, while at polar latitudes (>65°) the REF-C1SD runs consistently show stronger downwelling than in the reanalyses. The difference in \bar{w}^* at high latitudes between the REFC1-SD and reanalysis datasets extends throughout the depth of the stratosphere (see Supplementary Figure S10). More generally, Figure 3b shows that the different models that all nudge towards ERA-I (CCSRNIES MIROC3.2, CMAM, EMAC-L47/90, SOCOL3) produce very different mean residual circulations.

In summary, we conclude based on the results in Figures 1 to 3 that nudging meteorology does affect the strength and structure of the climatological residual circulation throughout the stratosphere. However, as implemented in these simulations (Table 2), nudging does not strongly constrain the mean amplitude and structure of the residual circulation nor does it produce circulations that closely resemble the direct estimates from the reanalyses.

~~In summary, nudging does not strongly constrain the mean strength of the residual circulation and in most cases the REF-C1SD simulations exhibit some notable differences compared to the residual circulation directly estimated from the reanalysis RA that they were nudged towards.~~

3.2 Climatological residual circulation: Tropical upward mass flux climatology

Figure 43 shows the vertical profiles of the climatological TUMF between 100 and 3 hPa calculated from annual means of \bar{w}^* from 100 hPa to 3 hPa for the (a) REF-C1, (b) REF-C1SD experiments and (c) their difference, calculated from annual means of \bar{w}^* . Note the logarithmic x-axis scale and that the CCMI and S-RIP fields have been interpolated from their native model levels to a set of predefined common pressure levels, which are rather sparse in the upper stratosphere; hence the TUMF calculation could be different if it was performed on the native model grid of both CCMI models and the reanalyses.

In terms of the differences between the REF-C1SD and REF-C1 simulations (Figure 4c), there is no consistent picture of the effect of nudging on the TUMF at different stratospheric levels. In the lowermost stratosphere between 70-100 hPa, most models ~~eases~~ (apart from EMAC-L47 and EMAC-

L90), ~~the REF-C1SD runs simulate stronger TUMF in the REF-C1SD runs than the equivalent free-running in REF-C1 simulations up to 10hPa (Figure 3e).~~ The largest TUMF differences in TUMF in the lower stratosphere due to nudging ~~are found occur in the CMAM and CCSRNIES MIROC3.2 EMAC-L90 and SOCOL3 nudged simulations,~~ which show larger values differences at 90 hPa ~~by more than of~~ around -20% and +25%, respectively. In the middle lower stratosphere, between 10-7100-30 hPa, some models show almost no difference in TUMF due to nudging (MRI-ESMr1), some show a stronger mass flux (CCSRNIES-MIROC3.2, CESM1-WACCM, CMAM) and others a weaker mass flux (EMAC-L47, SOCOL3). ~~In CESM1-WACCM, the differences in TUMF between the REF-C1SD and REF-C1 simulations are generally positive throughout the stratosphere and reach 20% at 20 hPa and ~25% at 3 hPa. The MRI-ESMr1 model generally shows the smallest differences in TUMF between the two experiments.~~ In the upper stratosphere (above 10 hPa) the picture is also mixed as half of the models show higher TUMF in the nudged experiments (CESM1-WACCM, EMAC-L47/L90) and the others show weaker TUMF (CCSRNIES-MIROC3.2, MRI-ESMr1, SOCOL3), ~~including EMAC-L47 and EMAC-L90 which show opposite sign differences at higher altitudes pressures.~~ CMAM shows the smallest change in TUMF in the upper stratosphere due to nudging. CESM1-WACCM is the only model to show a consistent sign of the TUMF differences between REF-C1SD and REF-C1 at all levels, with higher TUMF found throughout the stratosphere. There is no apparently simple relationship between the free-running model TUMF climatologies (Figure 4a) and the effect of nudging (Figure 4c).

We now ~~compared~~ the TUMF in each REF-C1SD experiment ~~with to the reanalysis- reanalysis that it they were~~ nudged towards (Figures ~~43b and 44d~~). Taking at first a broad view of the entire profiles, there is a resemblance between the profiles of TUMF differences in EMAC-L47 and SOCOL3 as compared to ERA-I, which may be related to the similarities in the implementation of nudging in these models; for example, vorticity and divergence were nudged with the same relaxation parameters (see Table 2). The CESM1-WACCM REF-C1SD simulation generally shows larger TUMF values than MERRA by up to 10-15% apart from in the upper stratosphere where they start to converge. MRI-ESMr1 exhibits relatively better agreement of TUMF with JRA-55 throughout the stratosphere. Looking across the models, ~~three most of the out of the four REF-C1SD simulations that were nudged towards ERA-I (CCSRNIES MIROC3.2 and EMAC-L47/L90) simulate stronger upwelling than their respective reanalysis RAreanalysis in the upper stratosphere,~~ with differences reaching up to 30-3550% in the two EMAC models. In fact, EMAC-L47/L90 show a high degree of similarity in the vertical structure of the TUMF differences between REFC1-SD and ERA-I at pressures less than 30 hPa, despite showing substantial differences in the lower stratosphere. ~~This could be partly due to the fact that both the CCMI and S-RIP fields have been interpolated from their native model levels to a set of predefined common pressure levels which are rather sparse in that vicinity, hence the TUMF calculation could be different if it was performed on the native model grid of both CCMI models and the reanalysis.~~ Additionally, for both EMAC REF-C1SD simulations the differences from the reanalysis in the TUMF

above 10 hPa can also be explained by the fact that because in EMAC the nudging is only imposed strongly up to 10 hPa, while higher model layers have weakening nudging coefficients as they serve as transition layers. In the middle stratosphere (50-20 hPa), most of the REF-C1SD models simulate a lower TUMF compared to the reanalysis. CESM1-WACCM generally shows slightly larger TUMF values than MERRA apart from the upper stratosphere where they start to converge. MRI-ESM1r1 exhibits relatively better agreement of TUMF with JRA-55 throughout the depth of the stratosphere, while CMAM which is nudged towards ERA-I follows closely the reanalysis especially above 10 hPa but is generally biased low. In terms of the spread found in both sets of experiments, in the lower stratosphere (100-30 hPa) where the maximum spread is located, the nudged simulations and their free-running counterparts show a total spread of $3.26 \times 10^9 \text{ kg s}^{-1}$ and $3.1 \times 10^9 \text{ kg s}^{-1}$ respectively, though note there are fewer REF-C1SD simulations to consider. Again, a key message is that the nudged REF-C1SD simulations show if not a slightly larger, a comparable, if not a slightly larger, spread in the climatological TUMF compared to the free-running REF-C1 simulations, throughout almost all the depth of the stratosphere.

To understand the dynamical factors that contribute to the modelled climatological residual circulation and its spread, Figure 54 shows the annual mean TUMF at 70 hPa along with the downward control calculations (section 2.2.2) to quantify the contribution of resolved and parameterized wave forcing to the TUMF. The black bars on the left show the TUMF diagnosed from \bar{w}^* and the grey bars on the right show the estimated contribution to TUMF from the Eliassen-Palm flux divergence (EPFD, dark grey), the orographic (mid-grey) and non-orographic (light grey) gravity wave drag. Note that SOCOL3 and ULAQ-CCM did not provide any wave forcing fields (Table 3), and NIWA-UKCA only provided the EPFD so we cannot perform the downward control calculations for that model.

In the free-running REF-C1 simulations (Fig. 54a), the estimated TUMF from the total wave forcing for the majority of the models (apart from CESM1-WACCM and EMAC-L90), slightly exceeds the TUMF calculated directly from \bar{w}^* ; this was not the case though for the CCMVal-2 models in SPARC (2010) (Figure 4.10, p.121). Since these simulations are internally consistent, the imperfect match indicates that the downward control principle as applied here relies on the close but inexact 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 REF-C1 inter-model range in TUMF at 70 hPa is 5.74×10^9 to $6.62 \times 10^9 \text{ kg s}^{-1}$ (inter-model standard deviation = $0.29 \times 10^9 \text{ kg s}^{-1}$). Comparing the CCMI results in Figure 5a with the that figure results from CCMVal-2 models (see Figure 4.10; SPARC, 2010) with the results in Figure 4a, the MMM TUMF at 70 hPa ($5.9 \times 10^9 \text{ kg s}^{-1}$) for the ten-seven REF-C1 model simulations analysed here ($6.05 \times 10^9 \text{ kg s}^{-1}$) is in very close agreement within the inter-model range of the when compared to the MMM of the fourteen CCMVal-2 models, which show a MMM TUMF around 4% weaker ($5.8 \times 10^9 \text{ kg s}^{-1}$) (SPARC, 2010). In terms of the contribution of the resolved wave

600 forcing to the TUMF in the free running simulations, there appears to be a decreased inter-model range
(3.26×10^9 to 5.33×10^9 kg s⁻¹) in the present study compared with the CCMVal-2 models, albeit that
study included more models (1.5×10^9 to 5.5×10^9 kg s⁻¹) (SPARC, 2010). Some CCMI models have
increased their horizontal resolution by up to a factor of two (CMAM, MRI-ESM1r1,
SOCOL3 and ULAQ-CCM) and also their vertical resolution up to 80 vertical levels (MRI-
605 ESM1r1) compared with CCMVal-2 models (Dietmüller et al., 2018), which could improve their ability
to simulate resolved wave forcing. There is a notable feature of CMAM which shows that the NOGWD
contributes negatively to TUMF (indicated with two red horizontal lines on Figure 54 and
Supplementary Figure S113); this was also found for CMAM in CCMVal-2 (Figure 4.10; SPARC,
2010).

610 The MMM TUMF at 70 hPa in the REF-C1SD simulations (Figure 54b) is 6.32×10^9 kg s⁻¹ or around
5% higher than in REF-C1. The REF-C1SD model range is larger than in REF-C1 being 5.39×10^9 to
 7.08×10^9 kg s⁻¹ (inter-model standard deviation = 0.51×10^9 kg s⁻¹). A notable feature is that the
contribution from the individual and total wave forcing contributions shows reduced inter-model spread
in the REF-C1SD simulations (Figure 5b, darker grey bars). For example, the inter-model standard
615 deviation of the EPFD contribution to TUMF at 70 hPa is around 40% smaller than in REF-C1
(0.44×10^9 kg s⁻¹ and 0.72×10^9 kg s⁻¹, respectively). Interestingly, for the single simulations that were
nudged towards MERRA and JRA-55 (CESM1-WACCM and MRI-ESM1r1, respectively) the TUMF
in the REF-C1SD runs is closer to the estimates from the reanalyses they are nudged towards. This may
simply be a coincidence given that there remain substantial differences in the structure of \bar{w}^* between
620 the REF-C1SD simulations and reanalyses (Figure 2b) and this does not apply necessarily for all models
that were nudged towards ERA-I. Another notable feature is that the contribution from the individual
and total wave forcing contributions shows reduced inter-model spread in the REF-C1SD simulations
(Figure 4b, grey bars). For example, the inter-model standard deviation of the EPFD contribution to
TUMF at 70 hPa is 43% smaller than in REF-C1 (0.43×10^9 kg s⁻¹ and 0.76×10^9 kg s⁻¹, respectively).

625 ~~However~~ Nonetheless, the residuals (i.e. the difference between the directly calculated TUMF and the
total downward control estimated contribution from the wave forcing) are substantially larger and
positive (except for EMAC-L90) in the REF-C1SD experiment than in REF-C1. This shows that
nudging adds an additional non-physical tendency in the model equations which acts to decouple the
wave forcing from the residual circulation; this means the physical constraint that the divergence of the
630 angular momentum flux due to the mean motion is balanced over some sufficient time average by that
of all eddy motions does not apply in the nudged models (Haynes et al., 1991). 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. Comparison of
the TUMF at 10 hPa for the REF-C1SD experiment (see Supplementary Figure S11b) also reveals
635 substantial differences in some models between the direct and downward control TUMF estimates in

the middle stratosphere. Variations in the residuals as a function of height may indicate differences in the effect of nudging on the connection between the climatological wave forcing and the shallow and deep branches of the circulation (Birner and Bönisch, 2011). However, the inter-model ranges in the directly calculated TUMF at 10 hPa are more comparable in the two experiments than was found at 70 hPa (1.45×10^9 to 1.70×10^9 kg s⁻¹ and 1.51×10^9 to 1.72×10^9 kg s⁻¹ for REF-C1 and REF-C1SD, respectively) (Supplementary Figure S11b).

Interestingly, for the single simulations that were nudged towards MERRA and JRA-55 (CESM1-WACCM and MRI-ESM1r1, respectively) the TUMF at 70 hPa in the REF-C1SD runs appear to be close to the estimates from the reanalyses they are nudged towards (compare black bars in Figure 5b). This may simply be a coincidence given that there are substantial differences in the structure of \bar{w}^* between the REF-C1SD simulations for those models and the reanalyses (Figures 3b and 3d), and this is not found for all 5 models that were nudged towards ERA-I. Indeed, given there is substantial spread in TUMF amongst the 5 REFC1-SD models nudged to ERA-I, it is likely that the differences between the REFC1-SD and reanalysis datasets 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 (Table 2). Despite this, the lower TUMF calculated directly from \bar{w}^* in EMAC-L90 compared to EMAC-L47 seen in both the REF-C1 and REF-C1SD experiments, is consistent with the results of Revell et al. (2015b) who also find that an increase in the model vertical resolution for ~~SOCOL3~~ results in a slowdown of the BDC.

~~Comparison of the TUMF at 10 hPa for the REF-C1SD experiment (see Supplement Figure S23b), reveal that the residuals are smaller in the middle stratosphere and as stated previously for both EMAC simulations, 10 hPa is the maximum level that the nudging is applied. Nevertheless, the rest of the REF-C1SD models were nudged even above that level also show smaller residuals and this may indicate differences in the effect of nudging on the shallow versus the deep branch of the circulation (Birner and Bönisch, 2011).~~ In summary, the results from Figures 4 and 5 reflect further demonstrate the fact that nudging imparts an external and non-physical tendency in the model equations, which in turn might cause violations of the normal constraints on the global circulation, such as conservation of momentum and energy. This is found to alter the residual circulation but in a manner that cannot be understood from and appears to limit the ability to a closure of the circulation through the integrated wave forcing as would ordinarily apply in the downward control principle (Haynes et al., 1991).

3.3 Annual cycle

We now evaluate the representation of the annual cycle in the residual circulation. Figure 65 shows the MMM climatological annual cycle of \bar{w}^* at 70 hPa for the REF-C1 and REF-C1SD simulations and their difference. ~~Note there are no significant variations in the results when the MMM of the REF-C1~~

~~experiment is computed only for those models used in the REF-C1SD analysis (Supplement Figure S4).~~

Both experiments show similar broad features in the annual cycle, with stronger tropical upwelling in boreal winter, a latitudinal asymmetry in the region of upwelling with the ~~turnaround-TA~~ latitude being further poleward in the summer hemisphere, and stronger downwelling over the winter pole. These features resemble the annual cycle found in other multi-model studies (e.g. Hardiman et al., 2014). Figure 65c shows that on average the nudged models simulate stronger upwelling in the sub-tropics, particularly in the NH in boreal winter with a few exceptions; the most prominent one being the narrow band between the equator and 10°N where the REF-C1 simulations exhibit stronger upwelling in austral winter. Consequently, the nudged models simulate substantially stronger downwelling in the midlatitudes in winter. In the NH mid-latitudes in the summer months nudged runs show ~~stronger~~ weaker downwelling, which reverses for the SH mid-latitudes in the austral winter. At polar latitudes there is a distinct seasonality to the differences between the REF-C1SD and REF-C1 simulations, with the nudged models simulating stronger downwelling in boreal winter and weaker downwelling in the Arctic during the rest of the year, corresponding to an amplified annual cycle. Conversely in the Antarctic, the REF-C1SD simulations generally simulate weaker downwelling, particularly during austral summer and spring.

To compare the annual cycle in residual circulation ~~across-in~~ the individual models, Figures 76a and 76b show the mean tropical (30°S - 30°N) \bar{w}^* at 70 hPa for the REF-C1 and REF-C1SD simulations, respectively. Comparing the MMM annual cycle of the REF-C1 runs (Figure 7a) with the MMM REF-C1SD (Figure 7b) reveals that on average the nudged models show a slightly larger peak-to-peak annual cycle amplitude (0.16 mm s⁻¹ vs. 0.13 mm s⁻¹). In general, the amplitude ~~and phasing~~ of the annual cycle in tropical mean \bar{w}^* is ~~rather weakly~~ slightly more constrained across the REF-C1SD simulations with the ~~intermodel~~ spread in peak-to-peak amplitude, as measured by the inter-model standard deviation, being around 2025% ~~higher-smaller~~ (0.062 mm s⁻¹ vs. 0.05 mm s⁻¹) ~~across all months compared to than~~ in the REF-C1 ($\sigma = 0.015$ mm s⁻¹ vs. 0.020 mm s⁻¹, respectively) ~~runs. Comparing the MMM annual cycle of the REF-C1 runs with the MMM REF-C1SD of Figure 6b reveals that on average the SD models show a slightly larger peak-to-peak amplitude in the annual cycle.~~ In terms of seasonal mean behaviour, the nudging appears to constrain the tropical mean \bar{w}^* in boreal summer (JJA), which exhibits ~20% less spread than in the REF-C1 experiments, but it does not constrain the tropical mean \bar{w}^* in boreal winter (DJF), which shows a factor of two larger spread than the free running models. Furthermore, ~~t~~he differences in tropical mean \bar{w}^* between the REF-C1SD runs and the respective reanalysis they are nudged towards are generally larger in boreal winter than in boreal summer for most models. In terms of spatially-resolved differences in \bar{w}^* between REF-C1SD and the reanalyses (Supplementary Figure S12), some consistent features include the REF-C1SD simulations showing stronger downwelling in the Arctic in boreal winter compared to the reanalyses and showing weaker

upwelling in the northern subtropics in boreal summer and autumn. Overall, the REF-C1SD minus reanalysis differences for the individual models highlight a wide variety in both the magnitude and the spatial patterns of their absolute differences, with no consistent picture emerging even for the models nudged towards the same reanalysis dataset.

710 Figures 76c and 76d show the climatological annual cycle in the ~~turnaround~~-TA latitudes at 70 hPa for the REF-C1 and REF-C1SD runs, respectively. This further breaks down the MMM annual mean perspective shown in Figure 2 by model and by season. In the SH, the spread in seasonal mean TA latitude across models, as measured by the intermodel standard deviation, is increased in the REF-C1SD experiment in all seasons by up to 30% compared to REF-C1. Conversely in the NH, the spread in
715 seasonal mean TA latitude is decreased for REF-C1SD in all seasons except boreal spring (MAM) where it is increased. There are also substantial differences between the TA latitudes in the REF-C1SD experiment and the reanalyses in all months, which shows that nudging does not produce consistent structures of regions of upwelling and downwelling to those in the reanalysis. ~~The inter-model spread in the REF-C1SD runs is also slightly higher in this metric (+6% for the NH TA latitude), as seen from~~
720 ~~Figure 6d, while the reanalyses also exhibit some differences. Note for this measure we exclude ULAQ-CCM when calculating the inter-model spread for the REF-C1 runs because it lacks a realistic seasonal evolution in its TA latitudes. The REF-C1SD runs show a weaker annual cycle in the SH TA latitude, with the models showing a consistently more poleward TA latitude in the SH in austral winter (JJA). In the NH, the TA latitudes in the REF-C1SD runs show a slightly smaller annual cycle compared to the~~
725 ~~REF-C1 runs. In summary, the amplitude and phasing of the annual cycle in tropical lower stratospheric \bar{w}^* appears to be more consistent across the nudged REF-C1SD simulations compared to the free-running REF-C1.~~ To summarize the results of Figure 7, there is substantial inter-model spread in the ~~turnaround~~-TA latitudes and in the amplitude of the annual cycle in \bar{w}^* highlighting significant interhemispheric differences in the upwelling region ~~in-between~~ both sets of simulations as well
730 ~~between the nudged experiment and the reanalyses. which is, if anything, higher in the nudged simulations than in the free running models.~~

3.4 Interannual variability of the tropical upward mass flux

Figure 87 shows timeseries over 1980-2009 for the annual, December-January-February (DJF), and
735 June-July-August (JJA) mean TUMF at 70 hPa for the REF-C1 (left column) and REF-C1SD (right column) simulations. As expected, the TUMF is larger in DJF compared to the annual and JJA means in both the REF-C1 and REF-C1SD runs because the average tropical upwelling is stronger in boreal winter. The individual REF-C1SD simulations show remarkably similar temporal variability in contrast to REF-C1 ~~where the modeled interannual variability is very diverse despite the models all being~~
740 forced with observed SSTs. Hence, although nudging does not constrain the mean TUMF in the lower

stratosphere, it does constrain the interannual variability; this is even more apparent for the DJF and JJA seasonal means (Figures 8d,f). Additionally, the REF-C1SD runs-simulations show exhibit a close relatively high agreement in their temporal variability to the reanalysis datasets they were nudged towards, albeit with magnitude-differences in magnitude and different-trends in-at the beginning of the 21st century where ERA-I and MERRA show a negative-behaviour-decrease in TUMF. Nevertheless, the inter-model spread is higher for the REF-C1SD-runs especially in the annual means for the whole time period of the TUMF timeseries in the lower stratosphere, with the inter-model spread for the nudged experiment being 25% higher (standard deviation of $0.54 \times 10^9 \text{ kg s}^{-1}$ vs. $0.4 \times 10^9 \text{ kg s}^{-1}$ respectively) than their free-running counterparts.

To investigate the cause of the high temporal correlation-coherence of the REF-C1SD annual-mean TUMF timeseries, in-Figure 98 we-presents the annual mean TUMF anomalies at 70 hPa along with the relative contributions to-the-interannual-variations-in-TUMF-from EPFD, OGWD, NOGWD and the total parameterized wave forcing (from top to bottom panels) for REF-C1 (left column) and REF-C1SD (right column), respectively. Figure 98b shows again the remarkably similar temporal variability in TUMF across the REF-C1SD runs, which can be contrasted against the weak interannual coherence in the REF-C1 runs (Figure 98a). Figure 98d and 98j show that both the EPFD and the total parametrized wave forcing contributions to the TUMF show a high degree of temporal coherence in the REF-C1SD simulations. The fact that the individual OGWD and NOGWD terms do not show such a strong inter-model agreement, while the total parametrized wave forcing does, could suggest there is some compensation occurring between the different parameterised wave forcing components (e.g. Cohen et al., 2013). This result indicates that although nudging does not constrain the mean residual circulation, it does constrain the interannual variability through influencing both the resolved and parametrized waves.

It should be noted that the reanalyses have been shown to exhibit strong similarities in their resolved EP fluxes as shown by the linear correlation in the timeseries of tropical upwelling at the 70 hPa level when considering the momentum balance estimates of \bar{w}^* (Abalos et al., 2015). This result indicates that although nudging does not constrain the mean residual circulation, it does constrain the interannual variability and produces very-similar contributions to variability across models from both resolved and parameterized wave forcing. Nonetheless, nudging only u , v (in most cases) and T (in some cases) in the SD models leads to very similar inferred contributions from resolved and parametrized wave foreing to the TUMF although these fields are not directly nudged. In contrast, The fact that the individual OGWD and NOGWD terms do not show such a strong inter-model agreement, while the total parametrized wave forcing does, could suggest there is some compensation between the resolved and parameterised wave foreing occurring (e.g. Cohen et al., 2013). The REF-C1 simulations show a highly variable pattern of the estimated TUMF anomalies contributed-by-from EPFD and OGWD-parameterized wave forcing (Figures 98c, 98ie), despite the fact they use the same observed SSTs and some nudge the phase of the QBO (CCSRNIES-MIRCO3.2, CESM1-WACCM, EMAC47/L90 and, SOCOL3SOCOL3-and

780 ~~ULAQ-CCM~~). In summary, the ~~source of the~~ remarkably coherent interannual variability in the annual TUMF timeseries in the REF-C1SD simulations ~~is due to the interannual variability of both the~~ is due to ~~both the~~ resolved and parametrized wave forcing being constrained by nudging; this is in strong contrast to ~~more tightly than~~ the climatological strength of the TUMF where there were large differences between the directly calculated TUMF and that due to wave forces (Figure 54b). The reasons for the difference in the effect of nudging on the behaviour of the residual circulation between the long-term mean and interannual variability is unclear.

785 3.5 Multiple Linear Regression analysis

790 Figures 109 and 110 show ~~present~~ timeseries of annual TUMF anomalies at 70 hPa attributed to each of the basis functions in the MLR model described in section 2.3 and the regression residuals for the REF-C1 and REF-C1SD runs, respectively. Also shown in the supplementary Figures S13 and S14 are the regression coefficients for each term and for each model along with their uncertainties. Figure 109a shows a large spread in the diagnosed signal of volcanic eruptions in the TUMF timeseries. The majority of the REF-C1 simulations analyzed here show a negative TUMF anomaly around the time of the El Chichón (1982) and Mount Pinatubo (1991) eruptions; however, the magnitude is within the estimated uncertainty range for all models except SOCOL3 (Supplementary Figure S13). In contrast to the REF-C1 results, most REF-C1SD simulations (except EMAC L47/L90 - see above discussion) show a positive anomaly in TUMF attributed to volcanic eruptions (Figure 11), consistent with earlier studies (Garcia et al., 2010; Diallo et al., 2017). However, there is still a considerable range of amplitudes and only the CESM1-WACCM and MRI-ESM1r1 regression coefficients are highly significant (Supplementary Figure S14). ~~a result that can be affected by how well aerosol heating is represented in the various models. Consequently, it highlights that in a free-running climate simulation, internal variability can be larger than the response to a transient forcing.~~ The issue of establishing a robust response of the TUMF to volcanic forcing over a short period is further demonstrated by the range in ~~the~~ amplitudes of the volcanic regressors for among different REF-C1 ensemble members from the same model (see Supplementary Figures S7S6-S12S9). This highlights that in a free-running climate simulation, internal variability can overwhelm the response to forcing over short timescales. The “true” volcanic signal in TUMF will also depend on the representation of stratospheric heating due to aerosol in the various models. We note that the EMAC L47/L90 models contained a unit conversion error where the extinction of stratospheric aerosols was set too low by a factor of ~500 (see Appendix B4 of Morgenstern et al., 2017), hence the stratospheric dynamical effects of the eruptions were not properly represented in the EMAC simulations (Jöckel et al., 2016).

810 ~~In the REF-C1SD runs (Figure 10), most models show a positive anomaly in TUMF attributed to volcanic eruptions, consistent with earlier studies, but still with a considerable range in amplitudes with~~

MRI-ESM1r1 and CESM1-WACCM showing the largest volcanic responses. The reason that the EMAC-L47/L90 models show a negative TUMF anomaly to volcanic forcing in both REF-C1 and REF-C1SD runs is documented in Appendix B4 of Morgenstern et al. (2017) and relates to a unit conversion error where the extinction of stratospheric aerosols was too low, by a factor of ~500, hence the stratospheric dynamical effects of those eruptions were not represented (Jöckel et al., 2016). The REF-C1 models all show a positive best estimate regression coefficient for the TUMF response to ENSO (Figure 10), which is quite consistent in amplitude, but it is only strongly statistically significant in CCSRNIES-MIROC3.2 and SOCOL3 (Figure S5). This is in contrast to the REF-C1SD models which all show a larger and more significant positive ENSO regression coefficient. ~~re are differences amongst models in the amplitude of the variance in the TUMF attributed to ENSO and the linear trend, but they are all consistent in sign (i.e. positive TUMF anomaly for El Niño and positive long-term trend). Despite the inter-model spread, for each model that performed both REF-C1 and REF-C1SD experiments the ENSO and linear trend contributions to the TUMF anomalies are quite similar, although the magnitude varies.~~ The linear trend regression coefficient over 1980-2009 is positive in all REF-C1 models and is statistically significantly different from zero at the 95% confidence level in five out of the seven models. The magnitude of the linear trend term varies by around a factor of 2 for REF-C1. In REF-C1SD, the amplitude of the linear trend regression coefficient increases in all models, but the intermodel spread increases to around a factor of 4. Hence, in these simulations nudging increases the disparity across models in the magnitude of the long-term TUMF trend.

As expected, the variations in TUMF attributed to the QBO are quite different in the REF-C1 and REF-C1SD runs for those models that do not nudge the QBO in REF-C1, as shown in Figures 109 and 110. The nudging of zonal winds in REF-C1SD constrains the phase of the QBO, and hence there is strikingly similar variability in the TUMF anomalies attributed to the QBO in the REF-C1SD runs. ~~The linear trend coefficient is statistically significantly different from zero at the 95% confidence level in five out of the six REF-C1SD models over the period 1980-2009, in contrast to being significant in half four of the REF-C1 simulations.~~

The overall R^2 values from the MLR model for the REF-C1 simulations vary between 0.16 (CMAM) and 0.67 (CESM1-WACCM). REF-C1SD runs generally give more consistent R^2 values across the models ranging from 0.62 (CCSRNIES-MIROC3.2) to 0.77 (EMAC-L47). This means there is still a substantial fraction (>23%) of unexplained variance in the annual TUMF timeseries in the REFC1-SD simulations after applying the MLR model, ~~which and the residuals exhibits~~ a remarkable degree of temporal correlation. In contrast, the MLR residuals in the REF-C1 runs (Figure 109f) show much less temporal coherence apart from a drop around 1989. ~~, which is also apparent in the REF-C1SD runs (bottom panel Figure 10). In contrast, the~~ The residuals in the REF-C1SD simulations (Figure 110f) show a high degree of coherent interannual variability, another manifestation of the fact that the nudged runs do reproduce a much more consistent inter-annual variability. This makes a substantial

contribution to the coherence of the TUMF timeseries in Figure 98b, but it cannot be attributed to any of the terms included in the MLR model.

850 For completeness, the MLR model was also applied to the reanalysis TUMF at 70 hPa (Supplementary Figures S15 and S16). This highlights significant discrepancies in attributing the variance in TUMF in the different reanalysis datasets to the various basis functions in the MLR model. Both the volcanic activity and ENSO contributions to the variance in the TUMF is rather weak 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 reanalyses and the REF-C1SD models. The negative trend in ERA-I found in the TUMF in the 855 lower stratosphere over 1980-2009 corroborates the findings of Abalos et al. (2015) who showed a negative trend in the direct \bar{w}^* estimate in ERA-I over 1979-2012. Despite this difference in the representation of the long-term TUMF trend, ERA-I shows the highest percentage of TUMF variance explained by the MLR model (66%), with MERRA showing a substantially lower R^2 (0.3) compared to 860 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 broadly similar on inter-decadal timescales. However, the regression residuals in the reanalyses show a different temporal behavior from those in the REF-C1SD simulations (Figure 11) (note that the y-axis scale for the residuals in Supplementary Figure S15 is double that for the CCMI models in Figures 10 and 11). In 865 summary, although nudging constrains the interannual variability in the TUMF at 70 hPa, the attribution to some specific drivers differs across the models and in comparison, to the reanalyses they were nudged towards.

870 3.6 Trend sensitivity analysis

Following ~~from~~ the results of the MLR analysis ~~described~~ in section 3.5, which showed a statistically significant ~~positive~~ linear trend in ~~most-some~~ REF-C1 and REF-C1SD models for the 30-year period 1980-2009, we now explore the sensitivity of the linear trend to the time period considered. We apply the same MLR model as discussed in section 3.5 to the annual mean 70 hPa TUMF timeseries of the 875 first ensemble member for both REF-C1 and REF-C1SD runs ~~as well as the reanalyses~~, but systematically vary~~ing~~ the start and end dates to cover all time periods in the window 1980-2009 that are at least ten years in length. We then extract the linear trend coefficient ~~from the MLR model~~ and its associated p-value. Figures 12~~+~~ and 13~~2~~ present the linear trend calculations for the REF-C1 and REF-C1SD runs, respectively, as a function of trend start and end date. ~~The same trend sensitivity analysis~~ 880 ~~for the reanalyses is presented in the supplement (Supplementary Figure S17)~~. Statistically significant trends at the 95% confidence level are marked with black stippling.

None of the periods considered in either the REF-C1 or REF-C1SD experiments shows a significant negative **TUMF** trend. A statistically significant positive trend emerges in almost all of the SD models for trends beginning in the mid-1980s to early 1990s extending to the mid-2000s. The trends are mainly significant for periods of 20 years or more and no less than around 12 years. This result broadly corroborates the findings of Hardiman et al. (2017a) who used a control run to estimate the period required to detect a BDC trend with an amplitude of 2% ~~per decade⁻¹ trend~~ against the background internal variability. There is range of different structures in the diagnosed trends among models, particularly for the REF-C1 simulations where a consistent pattern of positive trends only emerges across most models for the entire time period. This is because ~~of the differences in~~ internal variability ~~amongst the models that~~ can mask BDC trends over short periods (Hardiman et al., 2017a). However, the REF-C1SD runs simulate more consistent variations in **TUMF** trends as a function of time period, but generally show ~~weaker~~ **stronger** positive trends ~~in lower stratospheric TUMF~~ than their free-running counterparts. Interestingly, the reanalysis 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 reanalysis datasets exhibit a wide range of different trends from one another (Supplementary Figure S17) and differences compared to the trends in the REF-C1SD simulations (Figure 13). For example, none of the REFC1-SD models simulate a statistically non-significant negative trend in **TUMF** starting around mid-1990s up to 2009, as seen in all the reanalyses. However, it should also be noted that any trend combination starting around the end of 1990s in almost all cases of both REF-C1 and REF-C1SD runs exhibit no statistical significance possibly pointing towards the role of declining ozone depleting substances (ODS; Polvani et al., 2018) due to the implementation of the Montreal Protocol (Polvani et al., 2018).

905 4. Conclusions

This study ~~provides~~ **has performed** the first multi-model **intercomparison** of the impact of nudged meteorology on the representation of the stratospheric residual circulation. We use hindcast ~~CCMI runs~~ **simulations over 1980-2009 from CCMI** with identical prescribed ~~natural~~ **external** forcings in two configurations: REF-C1SD with meteorological fields nudged towards reanalysis data (specified dynamics, SD) and REF-C1 that is free-running ~~over the period 1980-2009~~. The nudged simulations use ~~one of three different~~ **various** reanalysis datasets (ERA-Interim, JRA-55, MERRA), nudge different variables (u, v, T, vorticity, **divergence**, surface pressure), and use different time constants to impose the additional nudging tendencies in the model equations.

The key findings of this study are:

- 915 1. Nudging ~~large-scale~~ meteorology does not ~~strongly~~ constrain the mean strength of the residual circulation compared to free-running ~~model~~ simulations. In fact, for most of the metrics of the

climatological ~~mean~~-residual circulation examined, including residual vertical velocities and mass fluxes, the inter-model spread ~~for fewer simulations~~ is comparable or ~~even in some cases~~ larger in the REF-C1SD ~~runs simulations~~ than in ~~the free-running~~ REF-C1 ~~simulations~~.

1. Nudging ~~generally~~ leads to ~~REF-C1SD runs~~ the models simulating on average ~~slightly~~ stronger upwelling at the equator in the lower to middle stratosphere and a wider tropical pipe in the lower stratosphere. In most cases, ~~and do not~~ the magnitude and structure of the climatological residual circulation in the REF-C1SD experiments differs markedly from that estimated for the reanalysis they are nudged towards. ~~quite resemble the direct from the reanalysis they were nudged towards.~~

2.

3. In ~~the most of the~~ nudged ~~simulations~~-models there are larger differences of up to 25% ~~among~~ between the directly ~~calculated~~ ~~simulated~~-tropical upward mass flux in the lower stratosphere ~~compared to their free-running counterparts which~~ and that calculated from ~~can be explained~~ ~~from~~ the diagnosed total wave forcing using the downward control principle (Haynes et al., 1991). ~~Nevertheless~~ However, the spread in the contributions from the resolved and parametrized wave forcing to the tropical mass flux is ~~slightly~~ reduced in the ~~nudged~~-REF-C1SD simulations ~~compared to REF-C1~~.

2.4. Despite the lack of ~~consistency~~ agreement in the mean circulation, nudging tightly constrains the interannual variability in the tropical upward mass flux (TUMF) in the lower stratosphere ~~compared to the free-running simulations~~. This ~~is~~ is associated with constraints to the contributions from both the resolved and parametrized wave forcing despite the fact the models use different reanalysis datasets for nudging. ~~The reanalysis datasets themselves exhibit broadly similar interannual variability in TUMF in the lower stratosphere, albeit with different long-term trends.~~

3. A multiple linear regression (MLR) analysis showed ~~ed~~ that up to 77% (~~67%~~) of the interannual variance ~~of in~~ the lower stratospheric ~~tropical upward mass flux~~ TUMF ~~timeseries~~ in the ~~lower stratosphere~~ REF-C1SD (REF-C1) experiments can be explained by ~~factors including~~ volcanic eruptions, ENSO, the QBO and a linear trend. The remaining unexplained TUMF variance ~~in the nudged models~~ shows a high degree of a temporal ~~covariance~~ coherence, ~~amongst models in the nudged simulations~~ but ~~this is not the case~~ for the free-running simulations.

5.

6. ~~However, applying the~~ The results of the MLR analysis applied to the TUMF in the reanalyses show differences in the total ~~same in the reanalysis datasets highlights the fact that they exhibit their own inherent uncertainties as seen from their differences in the~~ variance explained of the ~~mass flux in the lower stratosphere~~ and the attribution of variance to the different physical proxies ~~attributed to the same proxies~~. There are also marked differences between the

individual regression coefficients derived for the REF-C1SD models compared to the reanalysis dataset used for nudging.

- 955 7. Most nudged simulations show a statistically significant positive trend in 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 analyzed were nudged towards ERA-Interim, which shows a negative long-term trend in TUMF (see also Abalos et al., 2015), while JRA-55 and MERRA show a positive trend. However, the magnitude of the TUMF trend varies by up to a factor of 4 across the nudged models, which is larger than the spread in the free-running simulations. This is an important limitation for using nudged CCM simulations to interpret long-term changes in stratospheric tracers.

960 ~~—Nudged simulations show a statistically significant positive linear trend in tropical mass flux in the lower stratosphere over the period 1980-2009 despite the ERA-Interim reanalysis not showing a positive trend in upwelling. A linear trend sensitivity analysis for of the time period over for calculating lower stratospheric TUMF trends which the trend is calculated has showsn~~
965 that a statistically significant (at the 95% confidence level) ~~robust~~ positive ~~linear~~ trend in ~~the tropical upward mass flux~~ TUMF takes at least 12 years and in most cases around 20 years to emerge in the REF-C1SD runs. Despite the three reanalysis datasets showing different 30-year trends (1980-2009) they show a striking agreement in the statistically non-significant negative trends starting from the late 1990s up to 2009.

4.8.

970 Our findings highlight ~~the fact~~ that nudging strongly affects the representation of the stratospheric residual circulation in chemistry-climate model simulations, but it does not necessarily lead to improvements in the circulation. Similar disagreement in the characteristics of tropospheric transport in the CCMI nudged simulations has also been reported (Orbe et al., 2018). The differences found in the nudged runs compared with the free-running simulations suggest that although nudging horizontal fields can remove a-model biases ~~of~~in, for example, ~~the~~temperature and horizontal wind fields (Hardiman et al., 2017b), the simulated vertical wind field will not necessarily be similar to the reanalysis. ~~It is~~A particularly interesting finding of our study is that ~~while~~the nudging does not constrain the mean strength ~~of the circulation and the trend~~of the residual circulation, ~~but~~ it does constrain the interannual variability.

975 The reasons for ~~the~~is distinct effects of nudging on the residual circulation across these ~~different~~ ~~between~~ timescales ~~is~~are currently unknown.

980 Multiple factors are likely to determine the effect of nudging on the residual circulation in a given model including model biases, nudging timescales, nudging parameters, nudging height range, and model resolution. The differences in the stratospheric residual circulation between the REF-C1SD and the REF-C1 runs may not arise solely from the dynamics, but can also be partly influenced by the

indirect effects of nudging the temperatures which in turn affects the diabatic heating (Ming et al., 2016a, 2016b). In addition to nudging the horizontal winds (mechanical nudging), nudging the temperature (thermal nudging) might be systematically creating a spurious heat source in the model, which leads to a stronger BDC in the lower stratosphere as suggested by Miyazaki et al., (2005) with MRI GCM. Our results highlight that the method by which the large-scale flow is specified and more specifically the choice of the reanalysis fields, the relaxation timescale and the vertical grid (pressure level versus model level) in which the nudging is applied needs to be better understood and evaluated for their influence on the stratospheric circulation. Discrepancies between the vertical grid of the models and the reanalysis pressure levels they are interpolated onto or unbalanced dynamics are possible explanations for the differences found between the directly inferred circulation and that diagnosed from the wave forcing in the nudged simulations. Nudging would either violate continuity, or if continuity is maintained, it will come at the expense of the vertical fluxes, which are not nudged. The interesting aspect here seems to be that this results in substantial change to the net fluxes across a range of timescales, i.e. it does not only increase numerical noise in the \bar{w}^* component.

In order to reduce discrepancies between nudged and free-running simulations, various nudging techniques have been investigated. The role of gravity waves in the error growth that the nudging introduces over time has been highlighted for a single model (Smith et al., 2017). Constraining just the horizontal winds without the temperature was found to be a good strategy when investigating the aerosol indirect effects without affecting significantly the mean state (Zhang et al., 2014). The relaxation timescale when applying the nudging has been found to play an important role in single model studies (Merryfield et al., 2013), but there is no general consensus for the value of the relaxation constant, which is model-specific for the simulations considered here (Morgenstern et al., 2017). Given the varying implementations of nudging in the models analysed here, our study is ill-suited to investigate in detail the mechanisms for how nudging affects the residual circulation. A dedicated study of the sensitivities within one model to relaxation timescales, nudging parameters, nudging height range, and vertical resolution ~~etc~~ would ~~be needed~~ help to offer a detailed explanation for these differences.

The large spread in climatological residual circulation in nudged CCM simulations is an important limitation for those wishing to use them to examine tracer transport, for example stratospheric ozone trends (Solomon et al., 2016), volcanic aerosols (Schmidt et al., 2018), and diagnostics for age-of-air (Dietmüller et al., 2018). Despite the limitations for transport within the stratosphere described here, some success has been reported in studies that used nudged simulations to investigate specific meteorological events such as Sudden Stratospheric Warmings, and in particular for exploring processes beyond the top of the nudging region in the Mesosphere-Lower Thermosphere (e.g., Tweedy et al., 2013; Chandran and Collins, 2014; Pedatella et al., 2014). In conclusion, owing to the limitations of the current techniques for nudging models highlighted here, we urge caution in drawing quantitative

1025 comparisons of **stratospheric tracers affected by the residual circulation in nudged CCM-simulations**
against stratospheric observational data.

Data availability

1030 The majority of CCMI-1 used in this study can be obtained through the British Atmospheric Data
Centre (BADC) archive (<ftp://ftp.ceda.ac.uk>, last access: June 2018). For instructions for access to the
archive, see <http://blogs.reading.ac.uk/ccmi/badc-data-access>, last access: July 2019). The correctly
calculated EMAC-L47MA/EMAC-L90MA TEM model output for both REF-C1 and REF-C1SD were
obtained directly from Hella Garny. The SOCOL3 REF-C1SD TEM model output was obtained from
Andrea Stenke. The S-RIP data used in this study can be obtained through the through the British
Atmospheric Data Centre (BADC) archive (<ftp://ftp.ceda.ac.uk>, last access: September 2018), see
1035 <https://catalogue.ceda.ac.uk/uuid/b241a7f536a244749662360bd7839312>

Author contributions

AC performed the analysis and wrote the article.; ACM and MPC ~~made substantial contributions to the
conception and~~ designed ~~of~~ the study and **made substantial contributions to the** interpretation of the
data. Moreover, they participated in drafting and revising the article. HG provided the correctly
1040 calculated EMAC data and SD contributed to the discussion ~~on~~ of the content. **AS provided the SOCOL
REF-C1SD data.** The other co-authors contributed information pertaining to their individual models and
helped ~~edit revise the is~~ paper.

Competing Interests

The authors declare that they have no conflict of interest

Acknowledgments

AC was supported by a University of Leeds Anniversary **Postgraduate** Scholarship. ACM was
supported by a NERC Independent Research Fellowship (grant NE/M018199/1). MPC and SD
acknowledge support through the NERC SISLAC grant NE/R001782/1. We acknowledge the modelling
groups for making their simulations available for this analysis, the joint WCRP SPARC/IGAC
1050 Chemistry–Climate Model Initiative (CCMI) for organizing and coordinating the model data analysis
activity and the British Atmospheric Data Centre (BADC) for collecting and archiving the CCMI model
output. The EMAC simulations have been performed at the German Climate Computing Centre
(DKRZ) through support from the Bundesministerium für Bildung und Forschung (BMBF). DKRZ and
its scientific steering committee are gratefully acknowledged for providing the HPC and data archiving

1055 resources for this consortial project ESCiMo (Earth System Chemistry Integrated Modelling). ~~Olaf Morgenstern and Guang Zeng acknowledge the UK Met Office for use of the Met Office Unified Model (MetUM). This research was supported by the New Zealand Government's Strategic Science Investment Fund (SSIF) through the NIWA programme CACV. Olaf Morgenstern acknowledges funding by the New Zealand Royal Society Marsden Fund (grant 12-NIW-006) and by the Deep South~~
1060 ~~National Science Challenge (<http://www.deepsouthchallenge.co.nz>). The authors wish to acknowledge the contribution of New Zealand eScience Infrastructure (NeSI) high-performance computing facilities to the results of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme (<https://www.nesi.org.nz>). CCSRNIES research was supported by~~
1065 ~~the Environment Research and Technology Development Fund (2-1303 and 2-1709) of the Ministry of the Environment, Japan, and a grant-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (16H01183 and 18KK0289), and computations were performed on NEC-SX9/A(ECO) and NEC-SXACE computers at the CGER, NIES. The GEOSCCM is supported by the NASA MAP program and the high-performance computing~~
1070 ~~resources were provided by the NASA Center for Climate Simulations (NCCS). The analysis and visualization of the study has been performed using NCAR Command Language (NCL). The authors wish to acknowledge the contribution of Olaf Morgenstern and Guang Zeng in the discussion of the manuscript specifically for the model results of NIWA-UKCA in this research.~~

1075 **References**

- van Aalst, M. K., van den Broek, M. M. P., Bregman, A., Brühl, C., Steil, B., Toon, G. C., Garcelon, S., Hansford, G. M., Jones, R. L., Gardiner, T. D., Roelofs, G. J., Lelieveld, J. and Crutzen, P. J.: Trace gas transport in the 1999/2000 Arctic winter: comparison of nudged GCM runs with observations, *Atmos. Chem. Phys.*, 4(1), 81–93, doi:10.5194/acp-4-81-2004, 2004.
- 1080 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.*, 120(15), 7534–7554, doi:10.1002/2015JD023182, 2015.
- Abalos, M., Polvani, L., Calvo, N., Kinnison, D., Ploeger, F., Randel, W. and Solomon, S.: New Insights on the Impact of Ozone-Depleting Substances on the Brewer-Dobson Circulation, *J. Geophys. Res. Atmos.*, 124(5), 2435–2451, doi:10.1029/2018JD029301, 2019.
- 1085 Akiyoshi, H., Nakamura, T., Miyasaka, T., Shiotani, M. and Suzuki, M.: A nudged chemistry-climate model simulation of chemical constituent distribution at northern high-latitude stratosphere observed by SMILES and MLS during the 2009/2010 stratospheric sudden warming, *J. Geophys. Res. Atmos.*, 121(3), 1361–1380, doi:10.1002/2015JD023334, 2016.

- 1090 Andrews, D. G. and McIntyre, M. E.: An exact theory of nonlinear waves on a Lagrangian-mean flow, *J. Fluid Mech.*, 89(4), 609–646, doi:10.1017/S0022112078002773, 1978.
- Andrews, D. G. and McIntyre, M. E.: Planetary Waves in Horizontal and Vertical Shear: The Generalized Eliassen-Palm Relation and the Mean Zonal Acceleration, *J. Atmos. Sci.*, 33(11), 2031–2048, doi:10.1175/1520-0469(1976)033<2031:pwihav>2.0.co;2, 2002.
- 1095 Andrews, D. G., Holton, J. R. and Leovy, C. B.: *Middle Atmosphere Dynamics*, International Geophysical Series, Vol. 40, 1987.
- Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K. and Takahashi, M.: The quasi-biennial oscillation, *Rev. Geophys.*, 39(2), 179–229, doi:10.1029/1999RG000073, 2001.
- 1100 Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R. and Rozanov, E. V.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, *Atmos. Chem. Phys.*, 18(2), 1379–1394, doi:10.5194/acp-18-1379-2018, 2018.
- 1105 Beres, J. H., Garcia, R. R., Boville, B. A. and Sassi, F.: Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the Whole Atmosphere Community Climate Model (WACCM), *J. Geophys. Res. D Atmos.*, 110(10), 1–13, doi:10.1029/2004JD005504, 2005.
- 1110 Birner, T. and Bönisch, H.: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere, *Atmos. Chem. Phys.*, 11(2), 817–827, doi:10.5194/acp-11-817-2011, 2011.
- Butchart, N.: The Brewer-Dobson circulation, *Rev. Geophys.*, 52(2), 157–184, doi:10.1002/2013RG000448, 2014.
- Butchart, N. and Scaife, A. A.: Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate, *Nature*, 410(6830), 799–802, doi:10.1038/35071047, 2001.
- 1115 Butchart, N., Scaife, A. A., Bourqui, M., Grandpré, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D. and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, *Clim. Dyn.*, 27(7–8), 727–741, doi:10.1007/s00382-006-0162-4, 2006.
- 1120 Butchart, N., Cionni, I., Eyring, V., Shepherd, T. G., Waugh, D. W., Akiyoshi, H., Austin, J., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deckert, R., Dhomse, S., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Li, F., Mancini, E., McLandress, C., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Sassi, F., Scinocca, J. F., Shibata, K., Steil, B. and Tian, W.: Chemistry-climate model simulations of twenty-first century stratospheric climate and circulation
- 1125

changes, *J. Clim.*, 23(20), 5349–5374, doi:10.1175/2010JCLI3404.1, 2010.

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., Kushner, P. J., Newman, P. A., Osprey, S. M., Perlwitz, J., Sigmond, M., Wang, L., Akiyoshi, H., Austin, J., Bekki, S., Baumgaertner, A., Braesicke, P., Brhl, C., Chipperfield, M., Dameris, M., Dhomse, S., Eyring, V., Garcia, R., Garny, H., Jöckel, P., Lamarque, J. F., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Pawson, S., Plummer, D., Pyle, J., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssède, H., Tian, W., Waugh, D. and Yamashita, Y.: Multimodel climate and variability of the stratosphere, *J. Geophys. Res. Atmos.*, 116(5), 1–21, doi:10.1029/2010JD014995, 2011.

Calvo, N., Garcia, R. R., Randel, W. J. and Marsh, D. R.: Dynamical Mechanism for the Increase in Tropical Upwelling in the Lowermost Tropical Stratosphere during Warm ENSO Events, *J. Atmos. Sci.*, 67(7), 2331–2340, doi:10.1175/2010jas3433.1, 2010.

Chandran, A. and Collins, R. L.: Stratospheric sudden warming effects on winds and temperature in the middle atmosphere at middle and low latitudes: A study using WACCM, *Ann. Geophys.*, 32(7), 859–874, doi:10.5194/angeo-32-859-2014, 2014.

Chipperfield, M. P.: Multiannual simulations with a three-dimensional chemical transport model, *J. Geophys. Res. Atmos.*, 104(D1), 1781–1805, doi:10.1029/98JD02597, 1999.

Chipperfield, M. P.: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments, *Q. J. R. Meteorol. Soc.*, 132(617), 1179–1203, doi:10.1256/qj.05.51, 2006.

Chipperfield, M. P., Cariolle, D. and Simon, P.: A 3D transport model study of chlorine activation during EASOE, *Geophys. Res. Lett.*, 21(13), 1467–1470, doi:10.1029/93GL01679, 1994.

Chipperfield, M. P., Dhomse, S., Hossaini, R., Feng, W., Santee, M. L., Weber, M., Burrows, J. P., Wild, J. D., Loyola, D. and Coldewey-Egbers, M.: On the Cause of Recent Variations in Lower Stratospheric Ozone, *Geophys. Res. Lett.*, 45(11), 5718–5726, doi:10.1029/2018GL078071, 2018.

Cohen, N. Y. and Gerber, and Oliver Bühler, E. P.: Compensation between Resolved and Unresolved Wave Driving in the Stratosphere: Implications for Downward Control, *J. Atmos. Sci.*, 70(12), 3780–3798, doi:10.1175/jas-d-12-0346.1, 2013.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137(656), 553–597, doi:10.1002/qj.828, 2011.

Deushi, M. and Shibata, K.: Development of a Meteorological Research Institute chemistry-climate

model version 2 for the study of tropospheric and stratospheric chemistry, *Pap. Meteorol. Geophys.*, 62, 1–46, doi:10.2467/mripapers.62.1, 2011.

1165 Diallo, M., Ploeger, F., Konopka, P., Birner, T., Müller, R., Riese, M., Garny, H., Legras, B., Ray, E., Berthet, G. and Jegou, F.: Significant Contributions of Volcanic Aerosols to Decadal Changes in the Stratospheric Circulation, *Geophys. Res. Lett.*, 44(20), 10,780–10,791, doi:10.1002/2017GL074662, 2017.

1170 Dietmüller, S., Garny, H., Plöger, F., Jöckel, P. and Cai, D.: Effects of mixing on resolved and unresolved scales on stratospheric age of air, *Atmos. Chem. Phys.*, 17(12), 7703–7719, doi:10.5194/acp-17-7703-2017, 2017.

1175 Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., Mancini, E., Visioni, D., Stenke, A., Revell, L., Rozanov, E., Plummer, D. A., Scinocca, J., Jöckel, P., Oman, L., Deushi, M., Kiyotaka, S., Kinnison, D. E., Garcia, R., Morgenstern, O., Zeng, G., Stone, K. A. and Schofield, R.: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCM1-1 models, *Atmos. Chem. Phys.*, 18(9), 6699–6720, doi:10.5194/acp-18-6699-2018, 2018.

Douville, H.: Stratospheric polar vortex influence on Northern Hemisphere winter climate variability, *Geophys. Res. Lett.*, 36(18), 1–5, doi:10.1029/2009GL039334, 2009.

1180 Ebata, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y. and Ishimizu, T.: The Japanese 55-year Reanalysis “JRA-55”: An Interim Report, *Sola*, 7, 149–152, doi:10.2151/sola.2011-038, 2011.

1185 Eichinger, R., Dietmüller, S., Garny, H., Šácha, P., Birner, T., Bönisch, H., Pitari, G., Visioni, D., Stenke, A., Rozanov, E., Revell, L., Plummer, D. A., Jöckel, P., Oman, L., Deushi, M., Kinnison, D. E., Garcia, R., Morgenstern, O., Zeng, G., Adam Stone, K. and Schofield, R.: The influence of mixing on the stratospheric age of air changes in the 21st century, *Atmos. Chem. Phys.*, 19(2), 921–940, doi:10.5194/acp-19-921-2019, 2019.

1190 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.

Engel, A., Bönisch, H., Ullrich, M., Sitals, R., Membrive, O., Danis, F. and Crevoisier, C.: Mean age of stratospheric air derived from AirCore observations, *Atmos. Chem. Phys.*, 17(11), 6825–6838, doi:10.5194/acp-17-6825-2017, 2017.

1195 Eyring, V., Lamarque, J.-F., Cionni, I., Duncan, B., Fiore, A., Gettel-Man, A., Hegglin, M., Hess, P., Nagashima, T., Ryerson, T., Shepherd, T., Shindell, D., Waugh, D. and Young, P.: Report on the IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) 2013 Science Workshop. [online] Available from: www.sparc-climate.org/publications/newsletter/ (Accessed 15 July 2019), 2013.

- 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(7), 4783–4821, doi:10.5194/acp-19-4783-2019, 2019.
- García-Herrera, R., Calvo, N., Garcia, R. R. and Giorgetta, M. A.: Propagation of ENSO temperature signals into the middle atmosphere: A comparison of two general circulation models and ERA-40 reanalysis data, *J. Geophys. Res. Atmos.*, 111(6), 1–14, doi:10.1029/2005JD006061, 2006.
- 1205 Garcia, R. R. and Boville, B. A.: “Downward Control” of the Mean Meridional Circulation and Temperature Distribution of the Polar Winter Stratosphere, *J. Atmos. Sci.*, 51(15), 2238–2245, doi:10.1175/1520-0469(1994)051<2238:cotmmc>2.0.co;2, 2002.
- Garcia, R. R., Randel, W. J. and Kinnison, D. E.: On the Determination of Age of Air Trends from Atmospheric Trace Species, *J. Atmos. Sci.*, 68(1), 139–154, doi:10.1175/2010jas3527.1, 2010.
- 1210 Garcia, R. R., Smith, A. K., Kinnison, D. E., Cámara, Á. de la and Murphy, D. J.: Modification of the Gravity Wave Parameterization in the Whole Atmosphere Community Climate Model: Motivation and Results, *J. Atmos. Sci.*, 74(1), 275–291, doi:10.1175/jas-d-16-0104.1, 2016.
- Garny, H., Birner, T., Bönisch, H. and Bunzel, F.: The effects of mixing on age of air, *J. Geophys. Res.*, 119(12), 7015–7034, doi:10.1002/2013JD021417, 2014.
- 1215 Haenel, F. J., Stiller, G. P., Von Clarmann, T., Funke, B., Eckert, E., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A. and Reddmann, T.: Reassessment of MIPAS age of air trends and variability, *Atmos. Chem. Phys.*, 15(22), 13161–13176, doi:10.5194/acp-15-13161-2015, 2015.
- Hardiman, S. C., Butchart, N. and Calvo, N.: The morphology of the Brewer-Dobson circulation and its response to climate change in CMIP5 simulations, *Q. J. R. Meteorol. Soc.*, 140(683), 1958–1965, doi:10.1002/qj.2258, 2014.
- 1220 Hardiman, S. C., Lin, P., Scaife, A. A., Dunstone, N. J. and Ren, H. L.: The influence of dynamical variability on the observed Brewer-Dobson circulation trend, *Geophys. Res. Lett.*, 44(6), 2885–2892, doi:10.1002/2017GL072706, 2017a.
- Hardiman, S. C., Butchart, N., O’Connor, F. M. and Rumbold, S. T.: The Met Office HadGEM3-ES chemistry-climate model: Evaluation of stratospheric dynamics and its impact on ozone, *Geosci. Model Dev.*, 10(3), 1209–1232, doi:10.5194/gmd-10-1209-2017, 2017b.
- 1225 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.
- 1230 Hegglin, M.I.; and Lamarque, J. F.: The IGAC/SPARC Chemistry-Climate Model Initiative Phase-1 (CCMI-1) model data output. NCAS British Atmospheric Data Centre, [online] Available from: <http://catalogue.ceda.ac.uk/uuid/9cc6b94df0f4469d8066d69b5df879d5> (Accessed 6 June 2018), 2015.

- Hegglin, M. I. and Shepherd, T. G.: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux, *Nat. Geosci.*, 2(10), 687–691, doi:10.1038/ngeo604, 2009.
- Hines, C. O.: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation, *J. Atmos. Solar-Terrestrial Phys.*, 59(4), 371–386, doi:10.1016/S1364-6826(96)00079-X, 1997a.
- Hines, C. O.: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 2: Broad and quasi monochromatic spectra, and implementation, *J. Atmos. Solar-Terrestrial Phys.*, 59(4), 387–400, doi:10.1016/s1364-6826(96)00080-6, 1997b.
- Hitchcock, P. and Simpson, I. R.: The Downward Influence of Stratospheric Sudden Warmings*, *J. Atmos. Sci.*, 71(10), 3856–3876, doi:10.1175/jas-d-14-0012.1, 2014.
- Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B. and Pfister, L.: Stratosphere-troposphere exchange, *Rev. Geophys.*, 33(4), 403–439, doi:10.1029/95RG02097, 1995.
- Imai, K., Manago, N., Mitsuda, C., Naito, Y., Nishimoto, E., Sakazaki, T., Fujiwara, M., Froidevaux, L., Von Clarmann, T., Stiller, G. P., Murtagh, D. P., Rong, P. P., Mlynarczyk, M. G., Walker, K. A., Kinnison, D. E., Akiyoshi, H., Nakamura, T., Miyasaka, T., Nishibori, T., Mizobuchi, S., Kikuchi, K. I., Ozeki, H., Takahashi, C., Hayashi, H., Sano, T., Suzuki, M., Takayanagi, M. and Shiotani, M.: Validation of ozone data from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), *J. Geophys. Res. Atmos.*, 118(11), 5750–5769, doi:10.1002/jgrd.50434, 2013.
- Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J. and Bengtsson, L.: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation, *J. Geophys. Res. Atmos.*, 101(D12), 16939–16950, doi:10.1029/96JD01218, 1996.
- Jöckel, P., von Kuhlmann, R., Lawrence, M. G., Steil, B., Brenninkmeijer, C. A. M., Crutzen, P. J., Rasch, P. J. and Eaton, B.: On a fundamental problem in implementing flux-form advection schemes for tracer transport in 3-dimensional general circulation and chemistry transport models, *Q. J. R. Meteorol. Soc.*, 127(573), 1035–1052, doi:10.1256/smsqj.57317, 2001.
- Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S. and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), *Geosci. Model Dev.*, 3(2), 717–752, doi:10.5194/gmd-3-717-2010, 2010.
- Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K. D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D. and Zahn, A.: Earth System Chemistry integrated Modelling (ESCiMo) with the Modular Earth Submodel System (MESSy) version 2.51, *Geosci. Model Dev.*, 9(3), 1153–1200, doi:10.5194/gmd-9-1153-2016, 2016.
- Jonsson, A. I., de Grandpré, J., Fomichev, V. I., McConnell, J. C. and Beagley, S. R.: Doubled CO₂-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback,

- 1270 J. Geophys. Res. D Atmos., 109(24), 1–18, doi:10.1029/2004JD005093, 2004.
- Jung, T., Miller, M. J. and Palmer, T. N.: Diagnosing the Origin of Extended-Range Forecast Errors, *Mon. Weather Rev.*, 138(6), 2434–2446, doi:10.1175/2010mwr3255.1, 2010.
- Karpechko, A.Yu and A.C. Maycock (Lead Authors), M. Abalos, H. Akiyoshi, J.M. Arblaster, C.I. Garfinkel, K.H. Rosenlof and M. Sigmund, Stratospheric Ozone Changes and Climate, Chapter 5 in
 1275 Scientific Assessment of Ozone Depletion: 2018, Global Ozone Rese., 2018.
- Keeble, J., Bednarz, E. M., Banerjee, A., Luke Abraham, N., Harris, N. R. P., Maycock, A. C. and Pyle, J. A.: Diagnosing the radiative and chemical contributions to future changes in tropical column ozone with the UM-UKCA chemistry-climate model, *Atmos. Chem. Phys.*, 17(22), 13801–13818, doi:10.5194/acp-17-13801-2017, 2017.
- 1280 Kida, H.: General Circulation of Air Parcels and Transport Characteristics Derived from a hemispheric GCM, *J. Meteorol. Soc. Japan. Ser. II*, 61(2), 171–187, doi:10.2151/jmsj1965.61.2_171, 1983.
- Kobayashi, C. and Iwasaki, T.: Brewer-Dobson circulation diagnosed from JRA-55, *J. Geophys. Res.*, 121(4), 1493–1510, doi:10.1002/2015JD023476, 2016.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H.,
 1285 Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, *J. Meteorol. Soc. Japan. Ser. II*, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.
- Krol, M., De Bruine, M., Killaars, L., Ouwersloot, H., Pozzer, A., Yin, Y., Chevallier, F., Bousquet, P., Patra, P., Belikov, D., Maksyutov, S., Dhomse, S., Feng, W. and Chipperfield, M. P.: Age of air as a
 1290 diagnostic for transport timescales in global models, *Geosci. Model Dev.*, 11(8), 3109–3130, doi:10.5194/gmd-11-3109-2018, 2018.
- Lamarque, J. F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J. and Tyndall, G. K.: CAM-chem: Description and evaluation of interactive atmospheric chemistry in the Community Earth System
 1295 Model, *Geosci. Model Dev.*, 5(2), 369–411, doi:10.5194/gmd-5-369-2012, 2012.
- Lamarque, J. F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A. and
 1300 Zeng, G.: The atmospheric chemistry and climate model intercomparison Project (ACCMIP): Overview and description of models, simulations and climate diagnostics, *Geosci. Model Dev.*, 6(1), 179–206, doi:10.5194/gmd-6-179-2013, 2013.
- Lefevre, F., Brasseur, G. P., Folkins, I., Smith, A. K. and Simon, P.: Chemistry of the 1991-1992 stratospheric winter: three-dimensional model simulations, *J. Geophys. Res.*, 99(D4), 8183–8195,
 1305 doi:10.1029/93JD03476, 1994.

- Li, F., Newman, P., Pawson, S. and Perlwitz, J.: Effects of Greenhouse Gas Increase and Stratospheric Ozone Depletion on Stratospheric Mean Age of Air in 1960–2010, *J. Geophys. Res. Atmos.*, 123(4), 2098–2110, doi:10.1002/2017JD027562, 2018.
- 1310 Linz, M., Plumb, R. A., Gerber, E. P., Haenel, F. J., Stiller, G., Kinnison, D. E., Ming, A. and Neu, J. L.: The strength of the meridional overturning circulation of the stratosphere, *Nat. Geosci.*, 10(9), 663–667, doi:10.1038/ngeo3013, 2017.
- Linz, M., Abalos, M., Sasha Glanville, A., Kinnison, D. E., Ming, A. and Neu, J. L.: The global diabatic circulation of the stratosphere as a metric for the brewer-dobson circulation, *Atmos. Chem. Phys.*, 19(7), 5069–5090, doi:10.5194/acp-19-5069-2019, 2019.
- 1315 Löffler, M., Brinkop, S. and Jöckel, P.: Impact of major volcanic eruptions on stratospheric water vapour, *Atmos. Chem. Phys.*, 16(10), 6547–6562, doi:10.5194/acp-16-6547-2016, 2016.
- Mahieu, E., Chipperfield, M. P., Notholt, J., Reddman, T., Anderson, J., Bernath, P. F., Blumenstock, T., Coffey, M. T., Dhomse, S. S., Feng, W., Franco, B., Froidevaux, L., Griffith, D. W. T., Hannigan, J. W., Hase, F., Hossaini, R., Jones, N. B., Morino, I., Murata, I., Nakajima, H., Palm, M., Paton-Walsh, C., Russell, J. M., Schneider, M., Servais, C., Smale, D. and Walker, K. A.: Recent Northern Hemisphere stratospheric HCl increase due to atmospheric circulation changes, *Nature*, 515(7525), 104–107, doi:10.1038/nature13857, 2014.
- 1320 Marsh, D. R. and Garcia, R. R.: Attribution of decadal variability in lower-stratospheric tropical ozone, *Geophys. Res. Lett.*, 34(21), 1–5, doi:10.1029/2007GL030935, 2007.
- 1325 Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J. F., Calvo, N. and Polvani, L. M.: Climate change from 1850 to 2005 simulated in CESM1(WACCM), *J. Clim.*, 26(19), 7372–7391, doi:10.1175/JCLI-D-12-00558.1, 2013.
- Martineau, P.: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels. Centre for Environmental Data Analysis, doi:10.5285/b241a7f536a244749662360bd7839312, 2017.
- 1330 Martineau, P., Wright, J. S., Zhu, N. and Fujiwara, M.: Zonal-mean data set of global atmospheric reanalyses on pressure levels, *Earth Syst. Sci. Data*, 10(4), 1925–1941, doi:10.5194/essd-10-1925-2018, 2018.
- 1335 Maycock, A. C., Randel, W. J., Steiner, A. K., Karpechko, A. Y., Cristy, J., Saunders, R., Thompson, D. W. J., Zou, C.-Z., Chrysanthou, A., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Butchart, N., Chipperfield, M., Dameris, M., Deushi, M., Dhomse, S., Di Genova, G., Jöckel, P., Kinnison, D. E., Kirner, O., Ladstädter, F., Michou, M., Morgenstern, O., O’Connor, F., Oman, L., Pitari, G., Plummer, D. A., Revell, L. E., Rozanov, E., Stenke, A., Visionsi, D., Yamashita, Y. and Zeng, G.: Revisiting the mystery of recent stratospheric temperature trends, *Geophys. Res. Lett.*, 1–15, doi:10.1029/2018GL078035, 2018.
- 1340 McLandress, C., Scinocca, J. F., Shepherd, T. G., Reader, M. C. and Manney, G. L.: Dynamical Control

of the Mesosphere by Orographic and Nonorographic Gravity Wave Drag during the Extended Northern Winters of 2006 and 2009, *J. Atmos. Sci.*, 70(7), 2152–2169, doi:10.1175/jas-d-12-0297.1, 2012.

1345 Merryfield, W. J., Lee, W. S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R. S., Fyfe, J. C., Tang, Y. and Polavarapu, S.: The canadian seasonal to interannual prediction system. part I: Models and initialization, *Mon. Weather Rev.*, 141(8), 2910–2945, doi:10.1175/MWR-D-12-00216.1, 2013.

Ming, A., Hitchcock, P. and Haynes, P.: The Double Peak in Upwelling and Heating in the Tropical Lower Stratosphere, *J. Atmos. Sci.*, 73(5), 1889–1901, doi:10.1175/jas-d-15-0293.1, 2016a.

Ming, A., Hitchcock, P. and Haynes, P.: The Response of the Lower Stratosphere to Zonally Symmetric Thermal and Mechanical Forcing, *J. Atmos. Sci.*, 73(5), 1903–1922, doi:10.1175/jas-d-15-0294.1, 2016b.

Miyazaki, K., Iwasaki, T., Shibata, K., Deushi, M. and Sekiyama, T. T.: The Impact of Changing Meteorological Variables to Be Assimilated into GCM on Ozone Simulation with MRI CTM, *J. Meteorol. Soc. Japan. Ser. II*, 83(5), 909–918, doi:10.2151/jmsj.83.909, 2005.

Miyazaki, K., Iwasaki, T., Kawatani, Y., Kobayashi, C., Sugawara, S. and Hegglin, M. I.: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis data sets, *Atmos. Chem. Phys.*, 16(10), 6131–6152, doi:10.5194/acp-16-6131-2016, 2016.

1360 Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S. and Eichmann, A.: The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna. [online] Available from: <https://ntrs.nasa.gov/search.jsp?R=20120011790>, 2012.

Molod, A., Takacs, L., Suarez, M. and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2, *Geosci. Model Dev.*, 8(5), 1339–1356, doi:10.5194/gmd-8-1339-2015, 2015.

Monge-Sanz, B. M., Chipperfield, M. P., Dee, D. P., Simmons, A. J. and Uppalac, S. M.: Improvements in the stratospheric transport achieved by a chemistry transport model with ECMWF (re)analyses: Identifying effects and remaining challenges, *Q. J. R. Meteorol. Soc.*, 139(672), 654–673, doi:10.1002/qj.1996, 2013a.

1370 Monge-Sanz, B. M., Chipperfield, M. P., Untch, A., Morcrette, J. J., Rap, A. and Simmons, A. J.: On the uses of a new linear scheme for stratospheric methane in global models: Water source, transport tracer and radiative forcing, *Atmos. Chem. Phys.*, 13(18), 9641–9660, doi:10.5194/acp-13-9641-2013, 2013b.

Moore, F. L., Ray, E. A., Rosenlof, K. H., Elkins, J. W., Tans, P., Karion, A. and Sweeney, C.: A cost-effective trace gas measurement program for long-term monitoring of the stratospheric circulation, *Bull. Am. Meteorol. Soc.*, 95(1), 147–155, doi:10.1175/BAMS-D-12-00153.1, 2014.

Morgenstern, O., Braesicke, P., O’Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M. and

Pyle, J. A.: Evaluation of the new UKCA climate-composition model-Part 1: The stratosphere, *Geosci. Model Dev.*, 2(1), 43–57, doi:10.5194/gmd-2-43-2009, 2009.

1380 Morgenstern, O., Zeng, G., Abraham, N. L., Telford, P. J., Braesicke, P., Pyle, J. A., Hardiman, S. C., O’connor, F. M. and Johnson, C. E.: Impacts of climate change, ozone recovery, and increasing methane on surface ozone and the tropospheric oxidizing capacity, *J. Geophys. Res. Atmos.*, 118(2), 1028–1041, doi:10.1029/2012JD018382, 2013.

1385 Morgenstern, O., Hegglin, M., Rozanov, E., O’Connor, F., Luke Abraham, N., Akiyoshi, H., Archibald, A., Bekki, S., Butchart, N., Chipperfield, M., Deushi, M., Dhomse, S., Garcia, R., Hardiman, S., Horowitz, L., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M., Marchand, M., Marécal, V., Michou, M., Oman, L., Pitari, G., Plummer, D., Revell, L., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T., 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.

1390 Morgenstern, O., Stone, K. A., Schofield, R., Akiyoshi, H., Yamashita, Y., Kinnison, D. E., Garcia, R. R., Sudo, K., Plummer, D. A., Scinocca, J., Oman, L. D., Manyin, M. E., Zeng, G., Rozanov, E., Stenke, A., Revell, L. E., Pitari, G., Mancini, E., DI Genova, G., Visioni, D., Dhomse, S. S. and Chipperfield, M. P.: Ozone sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations, *Atmos. Chem. Phys.*, 18(2), 1091–1114, doi:10.5194/acp-18-1091-2018, 2018.

Neu, J. L. and Plumb, R. A.: Age of air in a “leaky pipe” model of stratospheric transport, *J. Geophys. Res. Atmos.*, 104(D16), 19243–19255, doi:10.1029/1999JD900251, 1999.

1400 Oman, L. D., Ziemke, J. R., Douglass, A. R., Waugh, D. W., Lang, C., Rodriguez, J. M. and Nielsen, J. E.: The response of tropical tropospheric ozone to ENSO, *Geophys. Res. Lett.*, 38(13), doi:10.1029/2011GL047865, 2011.

Oman, L. D., Douglass, A. R., Ziemke, J. R., Rodriguez, J. M., Waugh, D. W. and Nielsen, J. E.: The ozone response to enso in aura satellite measurements and a chemistry-climate simulation, *J. Geophys. Res. Atmos.*, 118(2), 965–976, doi:10.1029/2012JD018546, 2013.

1405 Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., Lamarque, J.-F., Tilmes, S., Plummer, D. A., Scinocca, J. F., Josse, B., Marecal, V., Jöckel, P., Oman, L. D., Strahan, S. E., Deushi, M., Tanaka, T. Y., Yoshida, K., Akiyoshi, H., Yamashita, Y., Stenke, A., Revell, L., Sukhodolov, T., Rozanov, E., Pitari, G., Visioni, D., Stone, K. A., Schofield, R. and Banerjee, A.: Large-scale tropospheric transport in the Chemistry–Climate Model Initiative (CCMI) simulations, *Atmos. Chem. Phys.*, 18(10), 7217–7235, doi:10.5194/acp-18-7217-2018, 2018.

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, *J. Geophys.*

Res. Sp. Phys., 119(2), 1306–1324, doi:10.1002/2013JA019421, 2014.

- 1415 Pitari, G. and Rizi, V.: Pitari, Rizi - 1993.pdf, *J. Atmos. Sci.*, 50(19), 3260–3276, doi:10.1175/1520-0469(1993)050<3260:AEOTCA>2.0.CO;2, 1993.
- Pitari, G., Aquila, V., Kravitz, B., Robock, A., Watanabe, S., Cionni, I., Luca, N., Genova, G., Mancini, E. and Tilmes, S.: Stratospheric ozone response to sulfate geoengineering: Results from the geoengineering model intercomparison project (GeoMip), *J. Geophys. Res.*, 119(5), 2629–2653, doi:10.1002/2013JD020566, 2014.
- 1420 Ploeger, F. and Birner, T.: Seasonal and inter-annual variability of lower stratospheric age of air spectra, *Atmos. Chem. Phys.*, 16(15), 10195–10213, doi:10.5194/acp-16-10195-2016, 2016.
- Ploeger, F., Abalos, M., Birner, T., Konopka, P., Legras, B., Müller, R. and Riese, M.: Quantifying the effects of mixing and residual circulation on trends of stratospheric mean age of air, *Geophys. Res. Lett.*, 42(6), 2047–2054, doi:10.1002/2014GL062927, 2015a.
- 1425 Ploeger, F., Riese, M., Haenel, F., Konopka, P., Müller, R. and Stiller, G.: Variability of stratospheric mean age of air and of the local effects of residual circulation and eddy mixing, *J. Geophys. Res.*, 120(2), 716–733, doi:10.1002/2014JD022468, 2015b.
- Ploeger, F., Legras, B., Charlesworth, E., Yan, X., Diallo, M., Konopka, P., Birner, T., Tao, M., Engel, A. and Riese, M.: How robust are stratospheric age of air trends from different reanalyses?, *Atmos. Chem. Phys.*, 19(9), 6085–6105, doi:10.5194/acp-19-6085-2019, 2019.
- 1430 PLUMB, R. A.: Stratospheric Transport, *J. Meteorol. Soc. Japan. Ser. II*, 80(4B), 793–809, doi:10.2151/jmsj.80.793, 2004.
- Polvani, L. M., Abalos, M., Garcia, R., Kinnison, D. and Randel, W. J.: Significant Weakening of Brewer-Dobson Circulation Trends Over the 21st Century as a Consequence of the Montreal Protocol, *Geophys. Res. Lett.*, 45(1), 401–409, doi:10.1002/2017GL075345, 2018.
- 1435 Polvani, L. M., Wang, L., Abalos, M., Butchart, N., Chipperfield, M. P., Dameris, M., Deushi, M., Dhomse, S. S., Jöckel, P., Kinnison, D., Michou, M., Morgenstern, O., Oman, L. D., Plummer, D. A. and Stone, K. A.: Large impacts, past and future, of ozone-depleting substances on Brewer-Dobson circulation trends: A multi-model assessment, *J. Geophys. Res. Atmos.*, 2018JD029516, doi:10.1029/2018JD029516, 2019.
- Randel, W. J. and Thompson, A. M.: Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes, *J. Geophys. Res. Atmos.*, 116(7), 1–9, doi:10.1029/2010JD015195, 2011.
- 1445 Randel, W. J., Garcia, R. R., Calvo, N. and Marsh, D.: ENSO influence on zonal mean temperature and ozone in the tropical lower stratosphere, *Geophys. Res. Lett.*, 36(15), 1–5, doi:10.1029/2009GL039343, 2009.
- Ray, E. A., Moore, F. L., Rosenlof, K. H., Plummer, D. A., Kolonjari, F. and Walker, K. A.: An idealized stratospheric model useful for understanding differences between long-lived trace gas

- 1450 measurements and global chemistry-climate model output, *J. Geophys. Res.*, 121(10), 5356–5367, doi:10.1002/2015JD024447, 2016.
- 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, 2015a.
- 1455 Revell, L. E., Tummon, F., Salawitch, R. J., Stenke, A. and Peter, T.: The changing ozone depletion potential of N₂O in a future climate, *Geophys. Res. Lett.*, 42(22), 10047–10055, doi:10.1002/2015GL065702, 2015b.
- 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(1), 136–156, doi:10.1175/2009jas3112.1, 2010.
- 1460 Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M. and Woollen, J.: MERRA: NASA’s modern-era retrospective analysis for research and applications, *J. Clim.*, 24(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.
- Rood, R. B., Allen, D. J., Baker, W. E., Lamich, D. J. and Kaye, J. A.: The Use of Assimilated Stratospheric Data in Constituent Transport Calculations, *J. Atmos. Sci.*, 46(5), 687–702, doi:10.1175/1520-0469(1989)046<0687:tuoasd>2.0.co;2, 2002.
- 1470 Rosenlof, K. H.: Seasonal cycle of the residual mean meridional circulation in the stratosphere, *J. Geophys. Res.*, 100(D3), 5173–5191, doi:10.1029/94JD03122, 1995.
- Šácha, P., Eichinger, R., Garny, H., Pišoft, P., Dietmüller, S., de la Torre, L., Plummer, D. A., Jöckel, P., Morgenstern, O., Zeng, G., Butchart, N. and Añel, J. A.: Extratropical age of air trends and causative factors in climate projection simulations, *Atmos. Chem. Phys.*, 19(11), 7627–7647, doi:10.5194/acp-19-7627-2019, 2019.
- 1475 Scaife, A. A., Butchart, N., Warner, C. D. and Swinbank, R.: Impact of a Spectral Gravity Wave Parameterization on the Stratosphere in the Met Office Unified Model, *J. Atmos. Sci.*, 59(9), 1473–1489, doi:10.1175/1520-0469(2002)059<1473:ioasgw>2.0.co;2, 2002.
- 1480 Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., Bardeen, C. G., Conley, A., Forster, P. M., Gettelman, A., Portmann, R. W., Solomon, S. and Toon, O. B.: Volcanic Radiative Forcing From 1979 to 2015, *J. Geophys. Res. Atmos.*, 123(22), 12,491–12,508, doi:10.1029/2018JD028776, 2018.
- 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(4), 763–766,
- 1485

doi:10.1029/91GL00022, 1991.

Scinocca, J. F.: An Accurate Spectral Nonorographic Gravity Wave Drag Parameterization for General Circulation Models, *J. Atmos. Sci.*, 60(4), 667–682, doi:10.1175/1520-0469(2003)060<0667:aasngw>2.0.co;2, 2003.

1490 Scinocca, J. F., McFarlane, N. A., Lazare, M., Li, J. and Plummer, D.: Technical note: The CCCma third generation AGCM and its extension into the middle atmosphere, *Atmos. Chem. Phys.*, 8(23), 7055–7074, doi:10.5194/acp-8-7055-2008, 2008.

Seviour, W. J. M., Butchart, N. and Hardiman, S. C.: The Brewer-Dobson circulation inferred from ERA-Interim, *Q. J. R. Meteorol. Soc.*, 138(665), 878–888, doi:10.1002/qj.966, 2012.

1495 Smith, A. K., Pedatella, N. M., Marsh, D. R. and Matsuo, T.: On the Dynamical Control of the Mesosphere–Lower Thermosphere by the Lower and Middle Atmosphere, *J. Atmos. Sci.*, 74(3), 933–947, doi:10.1175/jas-d-16-0226.1, 2017.

Solomon, S., Kinnison, D., Bandoro, J. and Garcia, R.: Simulation of polar ozone depletion: An update, *J. Geophys. Res.*, 120(15), 7958–7974, doi:10.1002/2015JD023365, 2015.

1500 Solomon, S., Kinnison, D., Garcia, R. R., Bandoro, J., Mills, M., Wilka, C., Neely, R. R., Schmidt, A., Barnes, J. E., Vernier, J. P. and Höpfner, M.: Monsoon circulations and tropical heterogeneous chlorine chemistry in the stratosphere, *Geophys. Res. Lett.*, 43(24), 12,624–12,633, doi:10.1002/2016GL071778, 2016.

1505 SPARC: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. V. Eyring, T. Shepherd and D. Waugh (Eds.), SPARC Rep. No. 5, WCRP-30/2010, WMO/TD-No.40, available at www.sparc-climate.org/publications/sp [online] Available from: <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/>, 2010.

Stenke, A., Schraner, M., Rozanov, E., Egorova, T., Luo, B. and Peter, T.: The SOCOL version 3.0 chemistry-climate model: Description, evaluation, and implications from an advanced transport algorithm, *Geosci. Model Dev.*, 6(5), 1407–1427, doi:10.5194/gmd-6-1407-2013, 2013.

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.

1515 Stone, K. A., Morgenstern, O., Karoly, D. J., Klekociuk, A. R., French, W. J., Abraham, N. L. and Schofield, R.: Evaluation of the ACCESS - Chemistry-climate model for the Southern Hemisphere, *Atmos. Chem. Phys.*, 16(4), 2401–2415, doi:10.5194/acp-16-2401-2016, 2016.

Telford, P. J., Braesicke, P., Morgenstern, O. and Pyle, J. A.: Technical note: Description and assessment of a nudged version of the new dynamics Unified Model, *Atmos. Chem. Phys.*, 8(6), 1701–1712, doi:10.5194/acp-8-1701-2008, 2008.

Tweedy, O. V., Limpasuvan, V., Orsolini, Y. J., Smith, A. K., Garcia, R. R., Kinnison, D., Randall, C.

E., Kvissel, O. K., Stordal, F., Harvey, V. L. and Chandran, A.: Nighttime secondary ozone layer during major stratospheric sudden warmings in specified-dynamics WACCM, *J. Geophys. Res. Atmos.*, 118(15), 8346–8358, doi:10.1002/jgrd.50651, 2013.

1525 UCAR/NCAR/CISL/TDD: The NCAR Command Language (NCL), Version 6.3.0, , doi:10.5065/D6WD3XH5, 2015.

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.

1530 Weber, M., Dikty, S., Burrows, J. P., Garny, H., Dameris, M., Kubin, A., Abalichin, J. and Langematz, U.: The Brewer-Dobson circulation and total ozone from seasonal to decadal time scales, *Atmos. Chem. Phys.*, 11(21), 11221–11235, doi:10.5194/acp-11-11221-2011, 2011.

Wild, O.: Modelling the global tropospheric ozone budget: Exploring the variability in current models, *Atmos. Chem. Phys.*, 7(10), 2643–2660, doi:10.5194/acp-7-2643-2007, 2007.

1535 World Meteorological Organization (WMO): WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project – Report No. 58, Geneva, Switzerland, 2018., 2018.

1540 Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., Tanaka, T. Y., Deushi, M., Obata, A., Nakano, H., Adachi, Y., Shindo, E., Yabu, S., Ose, T. and Kitoh, A.: Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1) - Model Description., 2011.

1545 Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., Mizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T. and Kitoh, A.: A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 ^|^mdash;Model Description and Basic Performance^|^mdash;, *J. Meteorol. Soc. Japan*, 90A(0), 23–64, doi:10.2151/jmsj.2012-A02, 2012.

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P. L., Rasch, P. J., Neubauer, D. and Lohmann, U.: Technical note: On the use of nudging for aerosol-climate model intercomparison studies, *Atmos. Chem. Phys.*, 14(16), 8631–8645, doi:10.5194/acp-14-8631-2014, 2014.

1550

Model Name	Reference(s)	Resolution	Top Level	REF-C1 ensemble members	Coord. Sys.	NOGWD Reference
CCSRNIES MIROC3.2	Imai et al., (2013), Akiyoshi et al., (2016)	T42, L34	1.2 Pa	3	TP	Hines (1997)
CESM1 WACCM	Marsh et al., (2013), Solomon et al., (2015), Garcia et al., (2016)	1.9° × 2.5°, L66	140 km	5	TP	Beres et al., (2005), Richter et al., (2010)
CMAM	Jonsson et al., (2004), Scinocca et al., (2008)	T47, L71	0.08 Pa	3	TP	Scinocca (2003)
EMAC (L47/L90)	Jöckel et al., (2010, 2016)	T42, L47/90	1 Pa	2	TP	Hines (1997a, b)
GEOSCCM	Molod et al., (2012, 2015), Oman et al., (2011, 2013)	~2° × 2°, L72	1.5 Pa	1	TP	Garcia and Boville, (1994)
MRI-ESM1r1	Yukimoto et al., (2011, 2012), Deushi and Shibata, (2011)	TL159, L80	1 Pa	1	TP	Hines (1997b)
NIWA-UKCA	Morgenstern et al., (2009, 2013), Stone et al., (2016)	3.75° × 2.5°, CP60	84 km	3	TA	Scaife et al., (2002)
SOCOL3	Stenke et al., (2013), Revell et al., (2015)	T42, L39	1 Pa	4	TP	Hines (1997a, b)
ULAQ CCM	Pitari et al., (2014)	T21, CP126	4 Pa	3	NTP	no OGWD

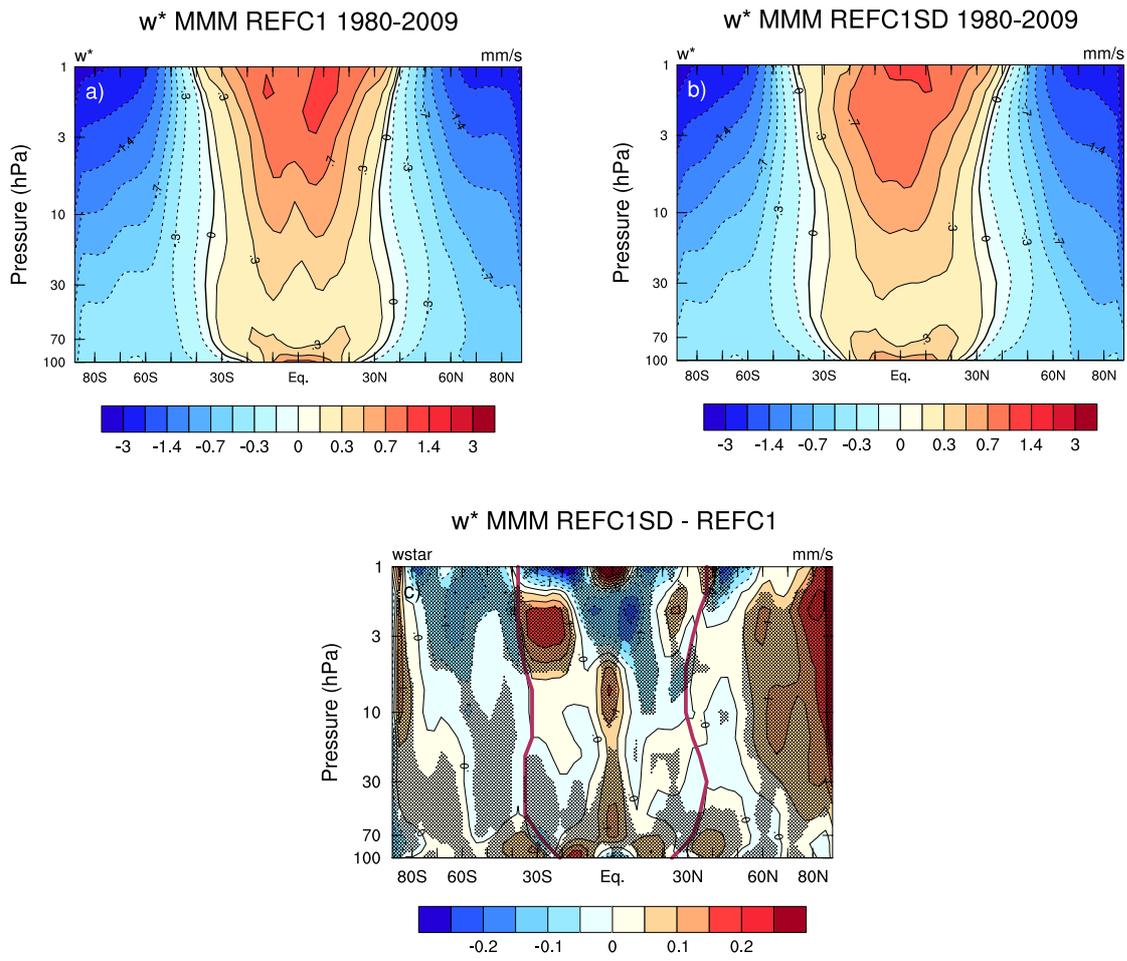
1555 **Table 1.** CCM1 models that provided TEM diagnostics model output used in this study. CP is Charney–Phillips; T21 $\approx 5.6^\circ \times 5.6^\circ$; T42 $\approx 2.8^\circ \times 2.8^\circ$; T47 $\approx 2.5^\circ \times 2.5^\circ$; TL159 $\approx 1.125^\circ \times 1.125^\circ$; TA is hybrid terrain-following altitude; TP is hybrid terrain-following pressure; NTP is non-terrain-following pressure.

Model Name	Pressure/height range of nudging	Newtonian relaxation timescale	Spectral nudging (Y/N)	Nudged Variables	Source of nudging data	Reference
CCSRNIES MIROC3.2	1000-1 hPa 1-0.01 hPa	1 day 1 day	Y N Y N	u, v, T, zonal-mean u and T zonal mean	ERA-I CIRA	Akiyoshi et al., (2016)
CESM1 WACCM	Surface – 50km (transition 40 – 50km)	50 h	N	u, v, T, surface pressure, surface stress, latent/sensible heat flux	MERRA	Lamarque et al., (2012)
CMAM	Surface – 1hPa	24 h	Y	Divergence, vorticity, temperature	ERA-I	McLandress et al., (2013)
EMAC (L47/L90)	920 – 780 hPa (transition) 710 – 10 hPa (full) 10 – 6 hPa (transition)	48 h 6 h 24 h 24 h	Y	Divergence, vorticity, T (with wave-0), (logarithm of) surface pressure	ERA-I	Jöckel et al., (2016)
MRI-ESM1r1	870 -1 hPa	24 h (870 -40 hPa) 24-∞ h (40-1 hPa)	N	u, v, T	JRA55	Deushi and Shibata, (2011)
SOCOL3	Surface to 0.01 hPa	48 h 6 h 24 h 24 h	Y	Divergence, vorticity, T, (logarithm of) surface pressure	ERA-I	

Table 2: Details of nudging in the CCMI REF-C1SD simulations that provided TEM diagnostics model output used in this study. ERA-I = ERA-Interim; CIRA = Cooperative Institute for Research in the Atmosphere; MERRA = Modern Era Retrospective ReAnalysis; JRA-55 = Japanese 55- year ReAnalysis. **T (with wave-0) for EMAC refers to the additional nudging of the global mean temperature.**

Model Name	REF-C1	REF-C1SD
CCSRNIES MIROC3.2	✓❖+★▲■	✓❖+★▲■
CESM1 WACCM	✓❖+▲■	✓❖+▲■
CMAM	✓❖+▲■	✓❖+▲■
EMAC (L47/L90)	✓❖+★▲■	✓❖+★▲■
GEOSCCM	✓❖+★▲■	
MRI-ESM1r1	✓❖+★▲■	✓❖+★▲■
NIWA-UKCA	✓❖+★	
SOCOL3	✓❖	✓❖
ULAQ CCM	✓❖	

Table 3. Available TEM-related model output for each model from the CCMI-1 archive: \bar{w}^* (✓), \bar{v}^* (❖), EPFD (+), GWD (OGWD+NOGWD) (★), OGWD (▲), NOGWD (■).



1575

Figure 1. Latitude vs. pressure climatology (1980-2009) of MMM annual mean \bar{w}^* for (a) REF-C1 simulations, (b) REF-C1SD simulations and (c) the REF-C1SD – REF-C1 absolute differences. Stippling denotes statistical significance at the 95% confidence level and the red lines in (c) denote the climatological turnaround latitudes in REF-C1SD.

1580

CCMI vs SRIP TA Latitudes 1980-2009

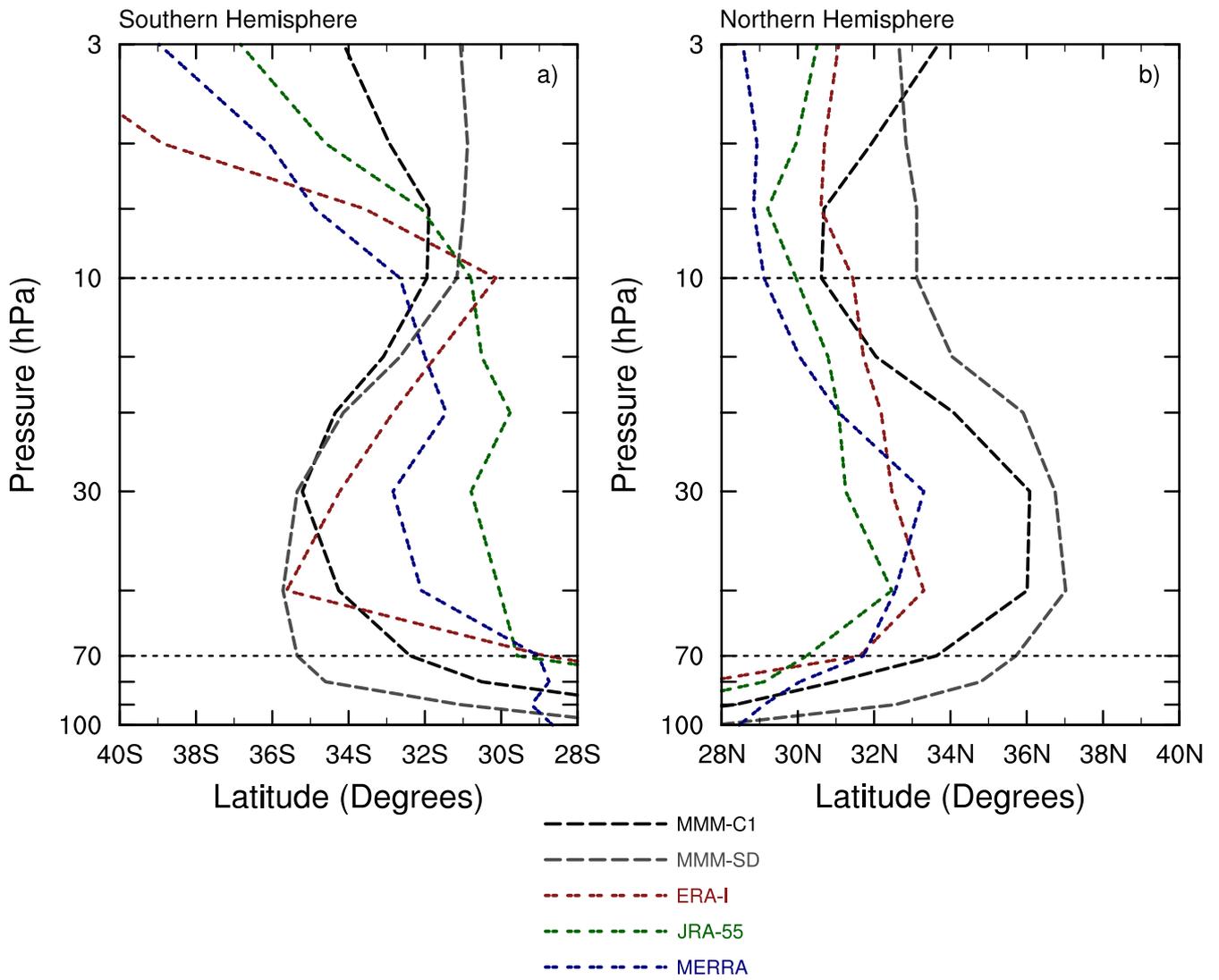


Figure 2: Vertical profiles of the climatological turnaround latitudes in the stratosphere for the MMM of the REF-C1 runs (black dashed), the MMM of the REF-C1SD runs (grey dashed) and the S-RIP reanalysis datasets (ERA-I, JRA-55, MERRA) for the (a) Southern Hemisphere and (b) Northern Hemisphere.

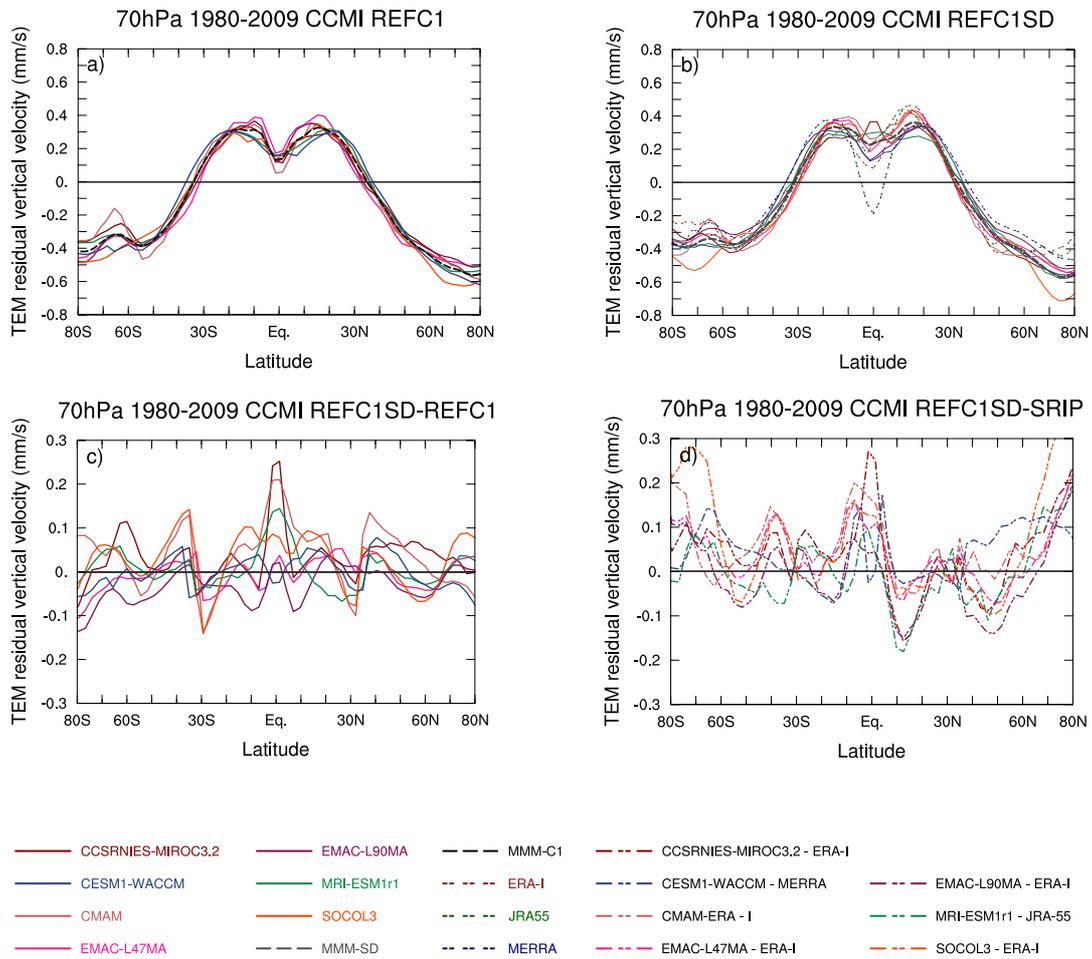


Figure 3. Mean strength of annual mean \bar{w}^* [mm s^{-1}] at 70 hPa for (a) REF-C1 free-running models, (b) REF-C1SD nudged models, (c) absolute differences between the REF-C1SD – REF-C1 experiment for each model and (d) absolute differences between each REF-C1SD simulation and the respective reanalysis used for nudging.

1595

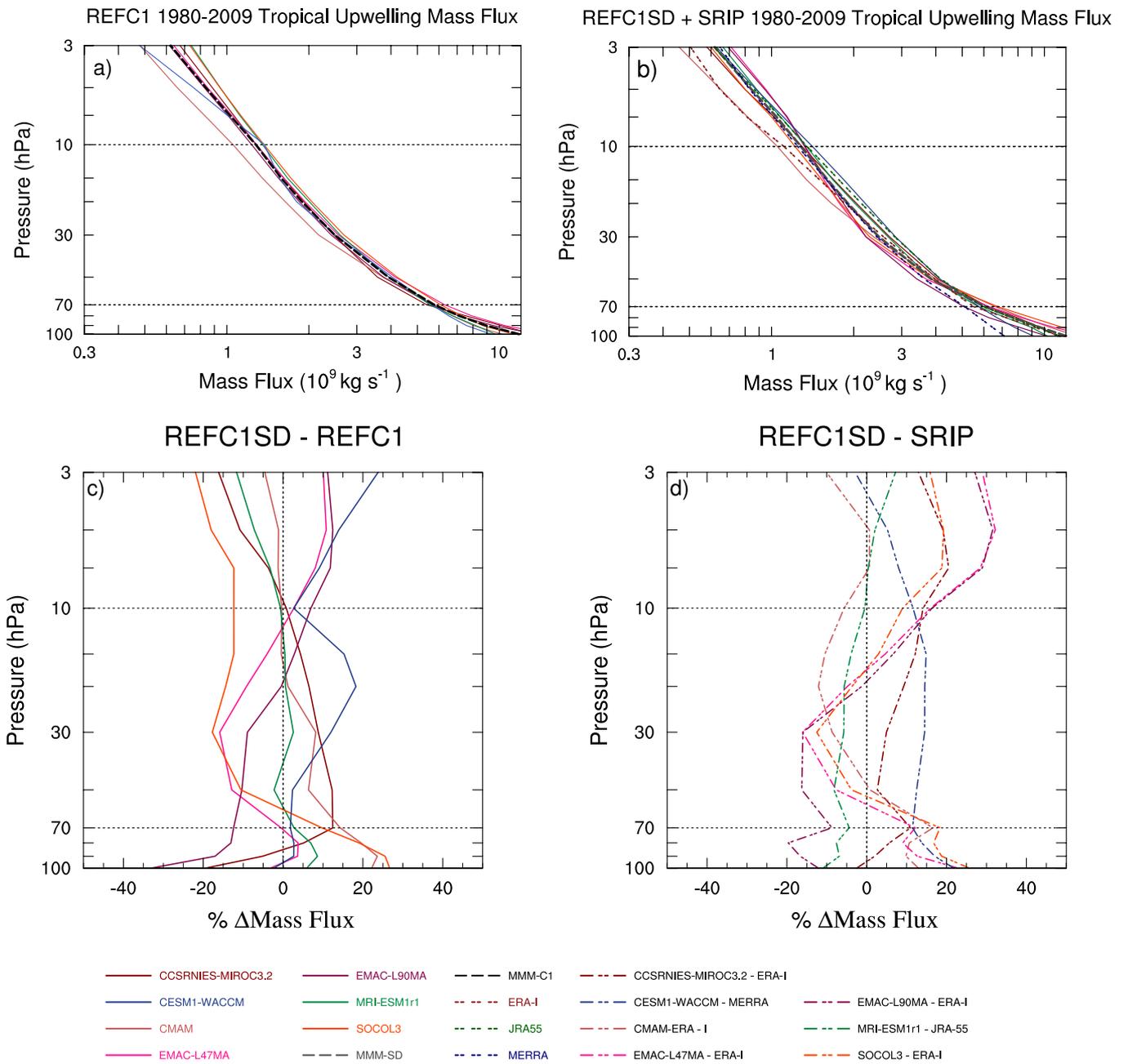
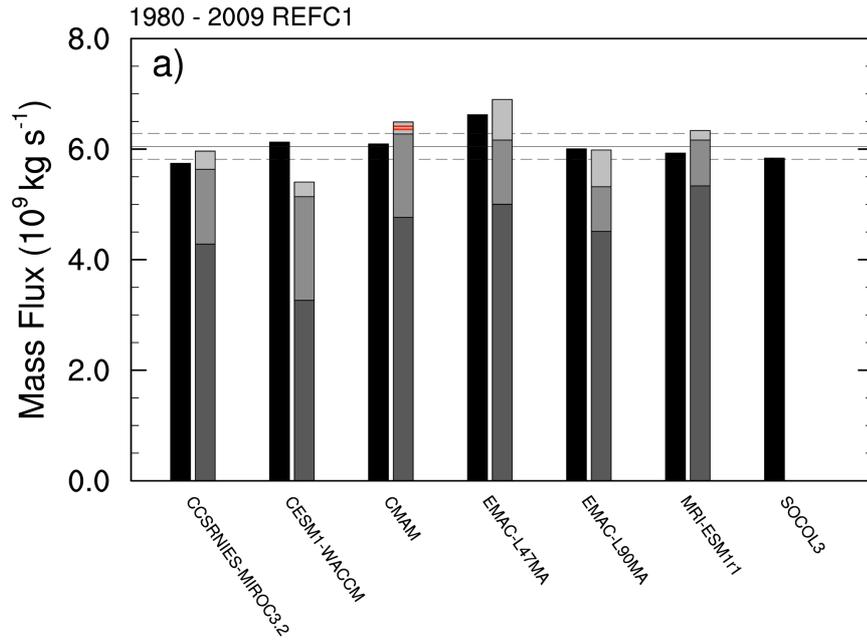


Figure 4. Vertical profiles of climatological (1980-2009) tropical upward mass flux [$\times 10^9 \text{ kg s}^{-1}$] averaged between the turnaround latitudes for (a) REF-C1 and (b) REF-C1SD, (c) % differences between REF-C1SD and REF-C1 and (d) % differences between REF-C1SD and the respective reanalysis used for nudging. Note the logarithmic x-axes in panels (a) and (b).

Annual mean upward mass flux at 70hPa



Annual mean upward mass flux at 70hPa

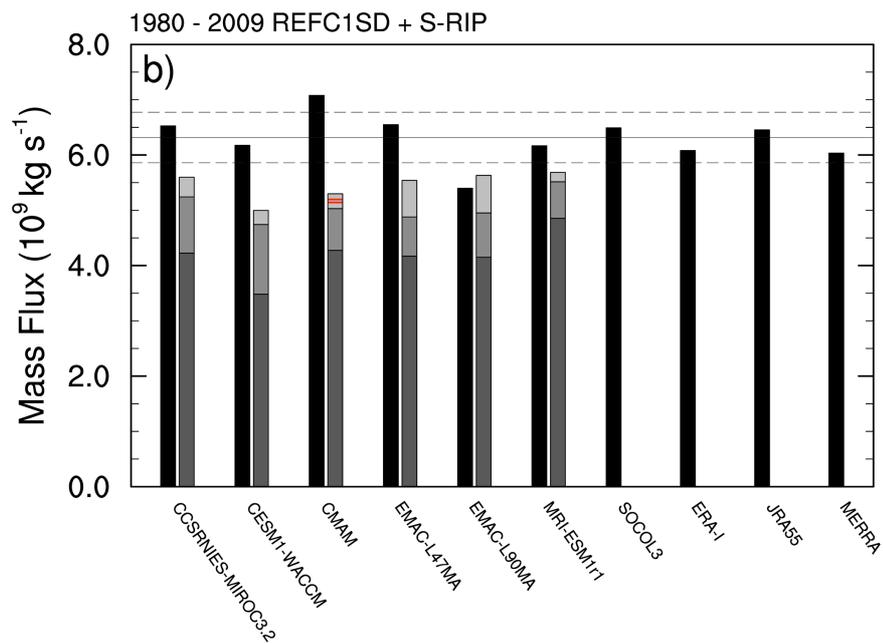


Figure 5. Tropical upward mass flux at 70 hPa (left bars) along with downward control calculations (right bars) showing contributions from EPFD (dark grey), OGWD (mid-grey), and NOGWD (light grey) for (a) REF-C1 and (b) REF-C1SD and the reanalyses. For CMAM the NOGWD contributes negatively to TUMF and is indicated with two red horizontal lines inside the lighter grey bar.

1615

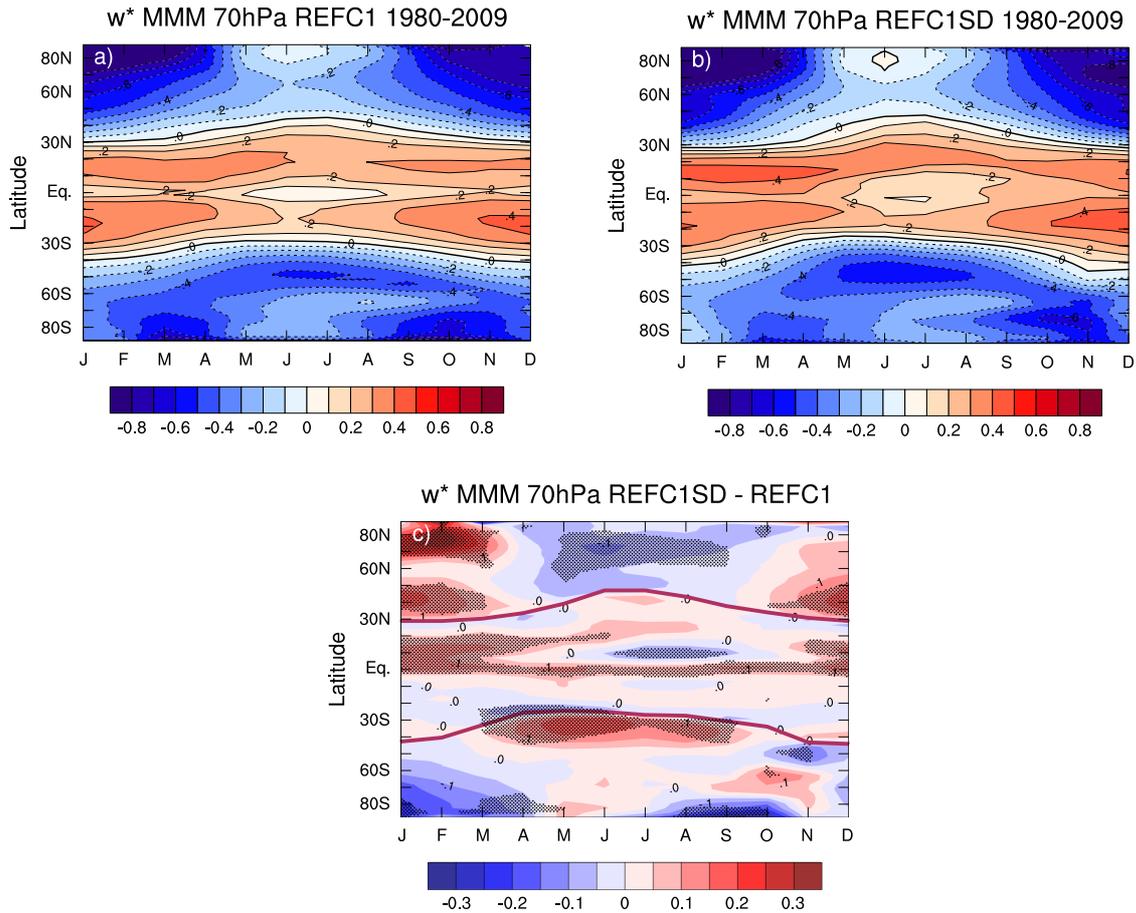


Figure 6. Climatological MMM annual cycle in \bar{w}^* [mm s^{-1}] at 70 hPa for (a) REF-C1 simulations, (b) REF-C1SD simulations and (c) the REF-C1SD minus REF-C1 absolute differences. The stipling in (c) denotes regions where the differences are statistically significant above 95% using a two-tailed Student's t-test where stipling denotes statistical significance above 95%. The turnaround latitudes ($\bar{w}^* = 0$) are shown by the thick black lines in panels 5a and 5b (a) and (b) and by the thick red purple lines for the MMM of REF-C1SD MMM in simulations in panel 5 (c).

1620

1625

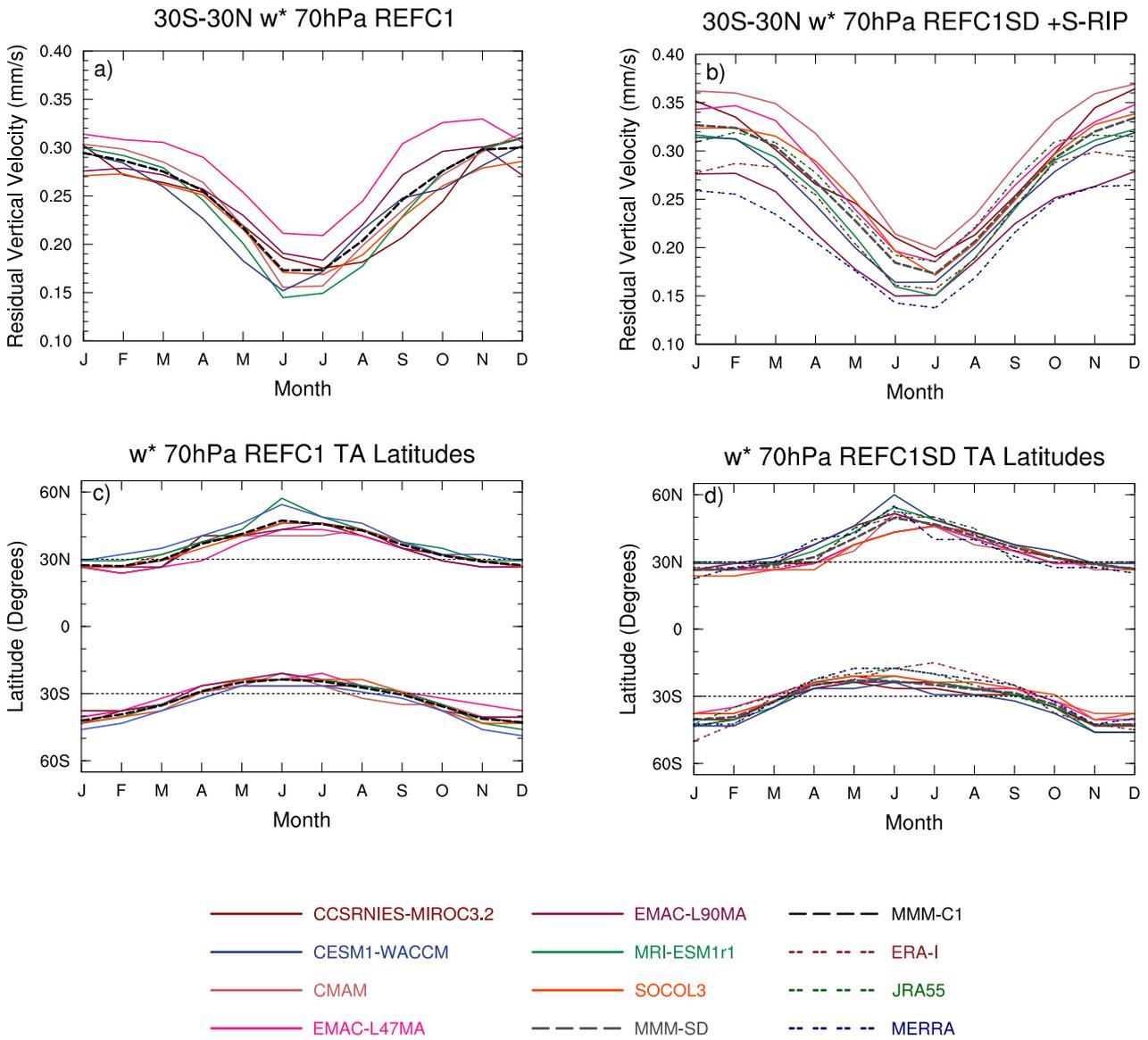
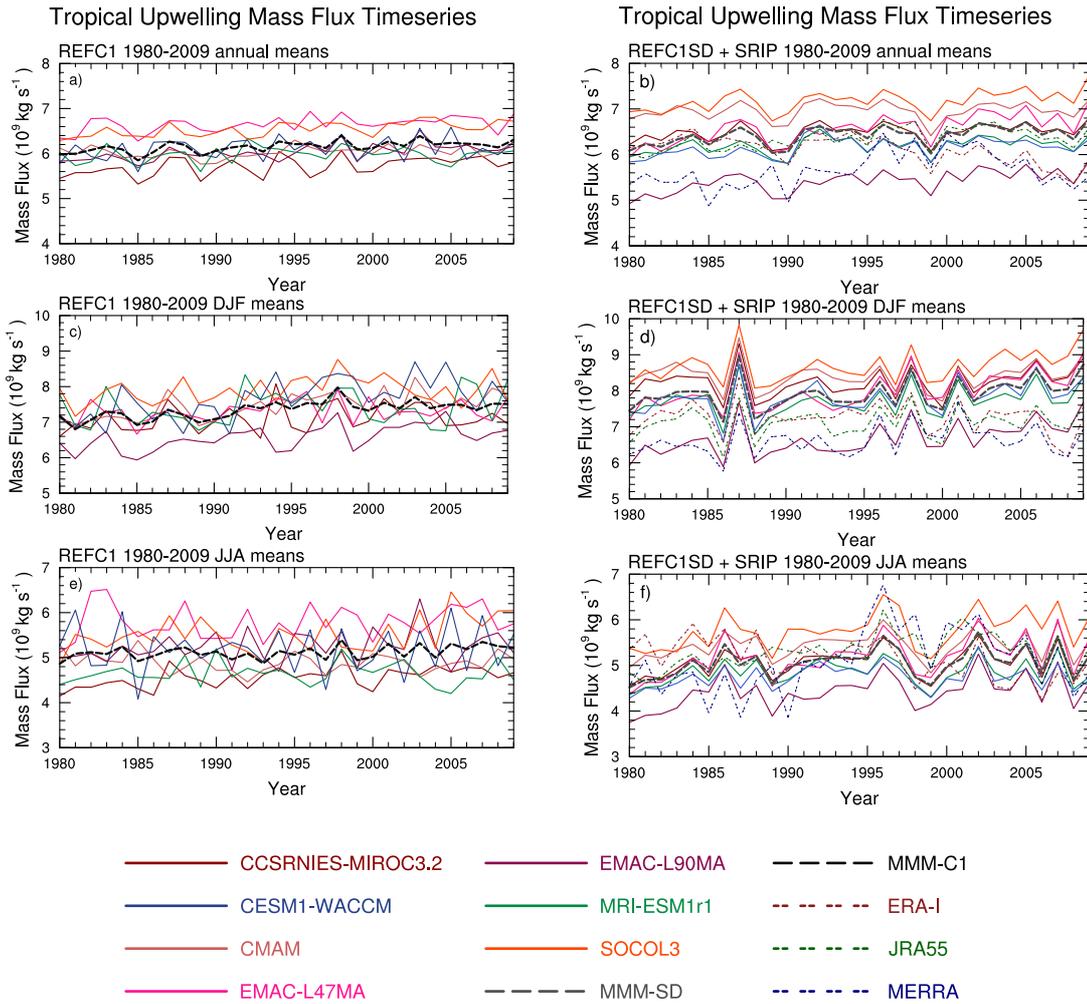
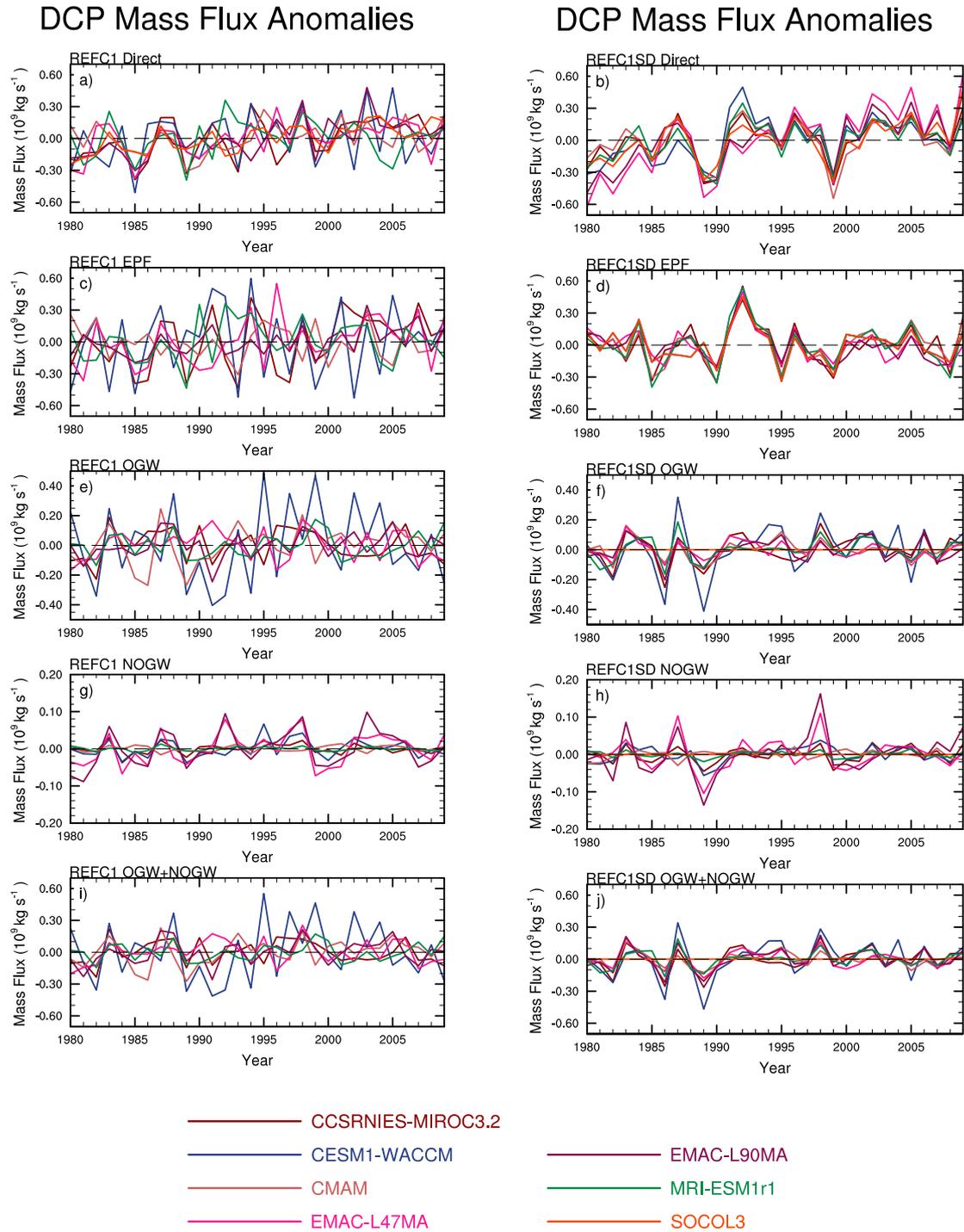


Figure 7. (Top) Climatological annual cycle in \bar{w}^* [mm s^{-1}] at 70 hPa between 30°S - 30°N in (a) REF-C1 and (b) REF-C1SD. (Bottom) Climatological annual cycle in turnaround latitudes at 70 hPa for each model in (c) REF-C1 and (d) REF-C1SD.

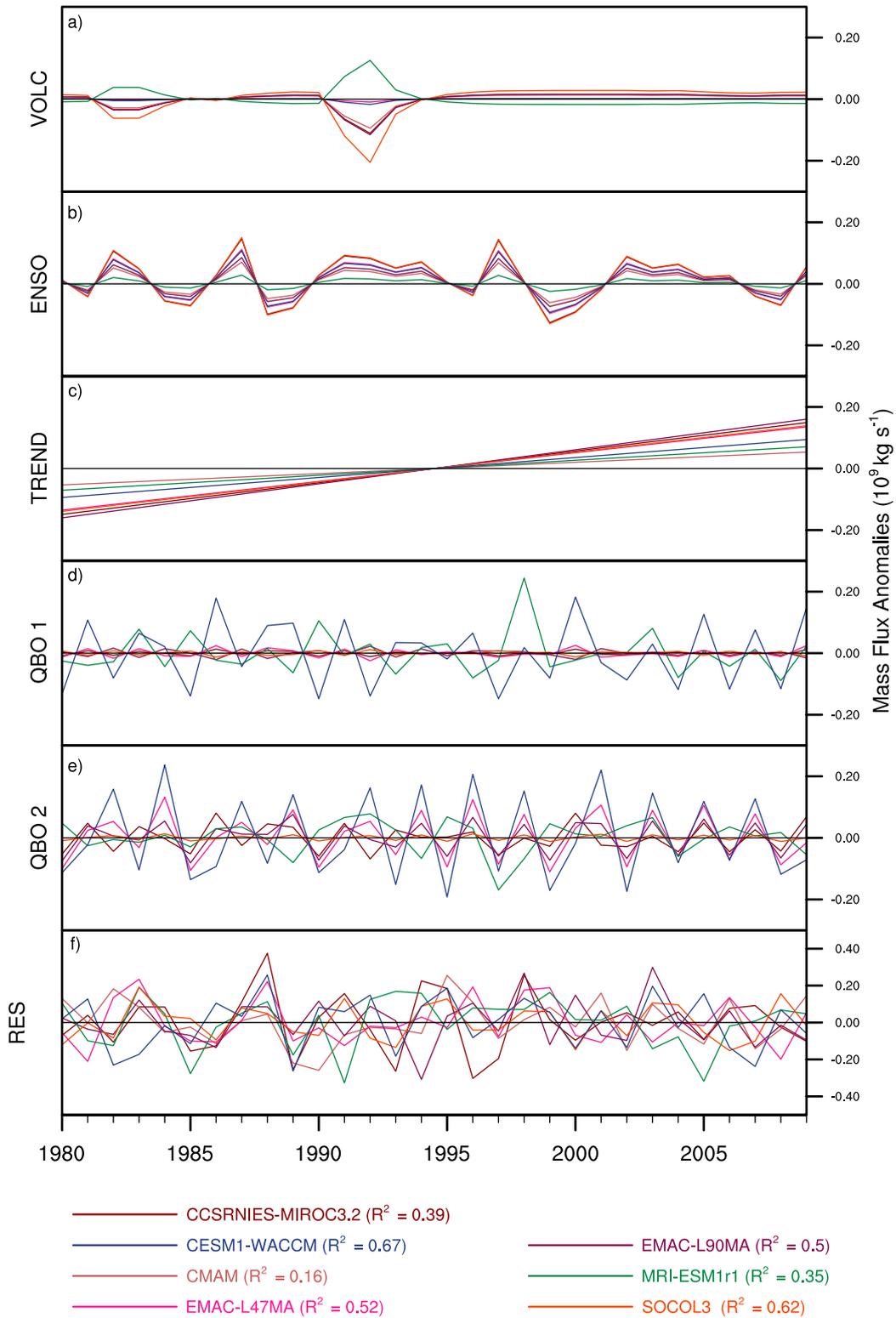


1635 **Figure 8.** (Top-to-bottom) Timeseries of Annual annual (ANN), DJF and JJA means of tropical upward mass flux [$\times 10^9 \text{ kg s}^{-1}$] at 70 hPa for (left panels) REF-C1 simulations and (right panels) REF-C1SD simulations.



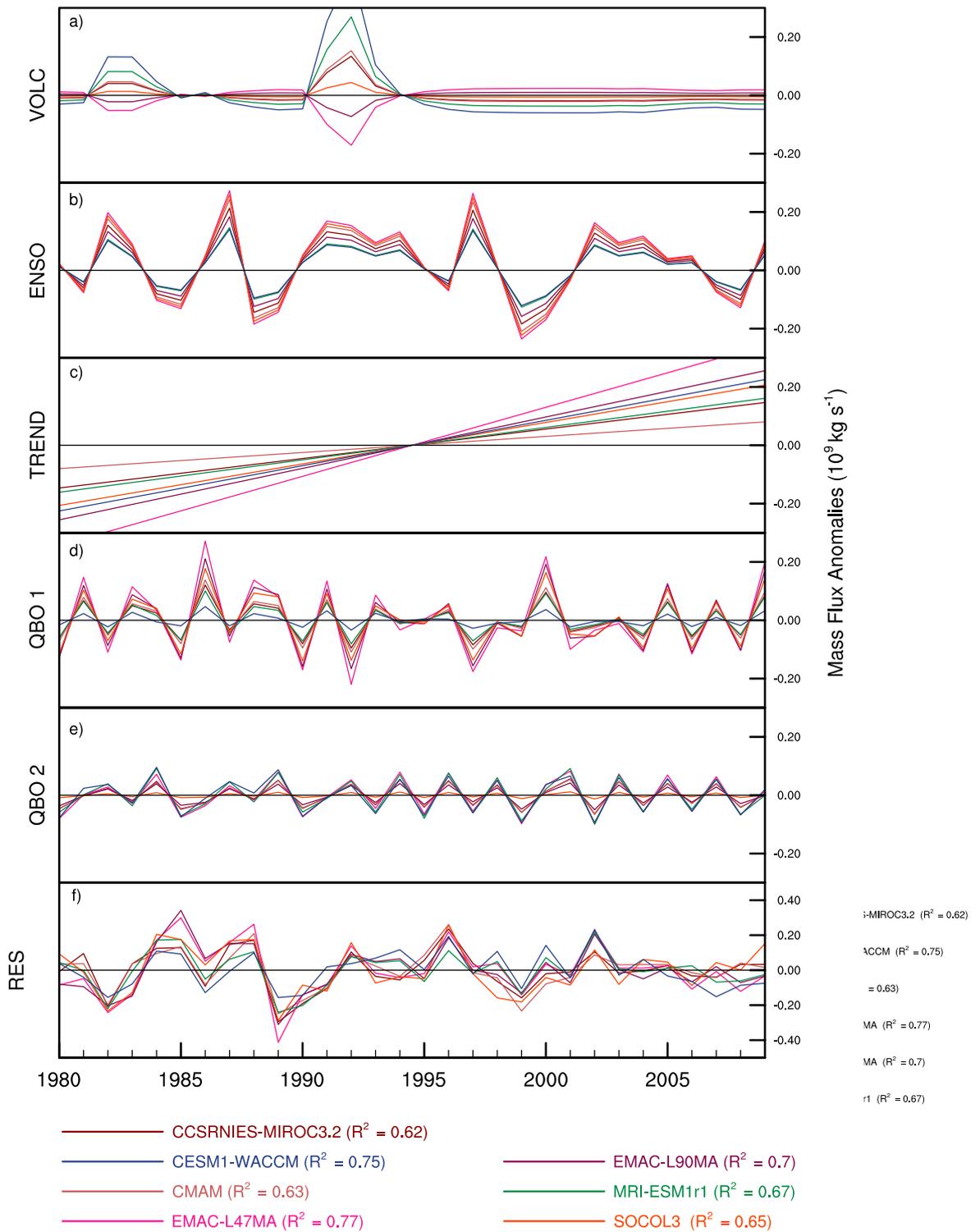
1640 **Figure 9.** Timeseries of the annual tropical upward mass flux anomalies [$\times 10^9 \text{ kg s}^{-1}$] calculated from (top-to-bottom) $-\bar{w}^* w^*$ (a, b), and the downward control principle DCP inferred contributions from resolved (EPF Divergence) wave driving (c, d), orographic gravity (OGWD) wave drag (e, f), non-orographic gravity (NOGWD) wave drag (g, h) and from the total parameterized (OGWD+NOGWD) gravity wave drag (i, j) for (left panels) REF-C1 simulations and (right panels) REF-C1SD simulations.

REFC1 Annual means TUMF MLR Analysis



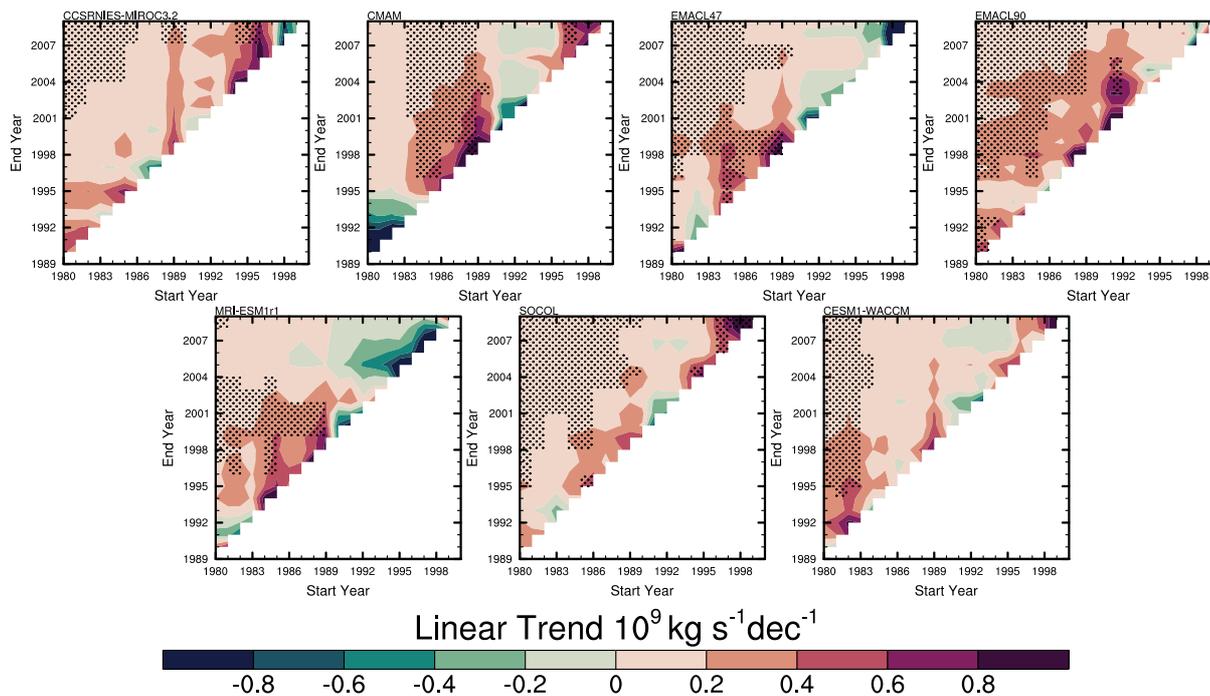
1645 **Figure 10.** Timeseries for REFC1 simulations of the components of the annual mean tropical upward mass flux [$\times 10^9 \text{ kg s}^{-1}$] attributed to (a) volcanic aerosol, (b) ENSO, (c) linear trend, (d, e) the QBO, and (f) the regression residuals from the mass flux timeseries and that reconstructed from the MLR.

REFC1SD Annual means DCP MLR Analysis



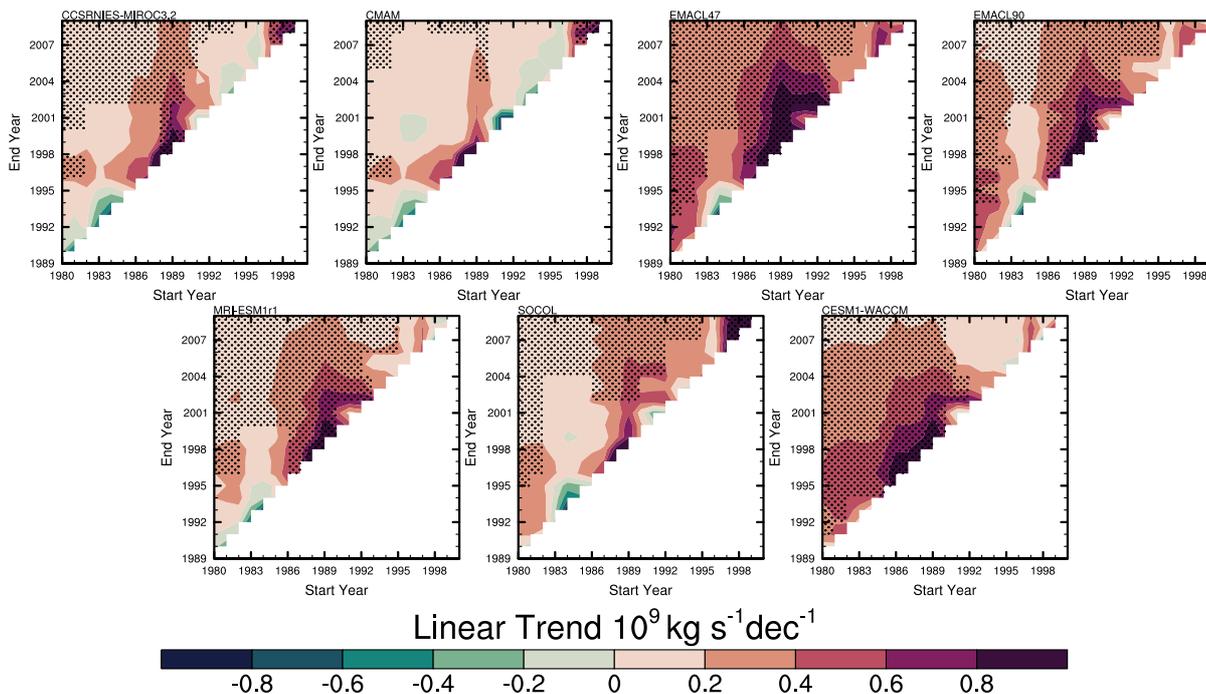
1650 **Figure 11.** Timeseries for REFC1SD simulations of the components of the annual mean tropical upward mass flux $[\times 10^9 \text{ kg s}^{-1}]$ attributed to (a) volcanic aerosol, (b) ENSO, (c) linear trend, (d, e) the QBO, and (f) regression residuals from the mass flux timeseries and that reconstructed from the MLR.

Start/End Date Trend Sensitivity REFC1



1655 **Figure 12.** 70 hPa tropical upward mass flux trends [$\times 10^9 \text{ kg s}^{-1} \text{ decade}^{-1}$] for different start (abscissa) and end (ordinate) dates ~~Mass flux linear trend partial regressor from the MLR sensitivity plots (values with statistical significance are stippled)~~ over the period 1980-2009 for the REF-C1 (r1i1p1) simulations. Trends are not shown for periods of less than 10 years. Values with statistical significance greater than the 95% level are stippled.

Start/End Date Trend Sensitivity REFC1SD



1660 **Figure 13.** As in Figure 12 but for the REF-C1SD simulations.

~~Mass flux linear trend partial regressor from the MLR sensitivity plots (values with statistical significance are stippled) over period 1980-2009 for REF-C1SD simulations.~~