Responses to Referee #3

Many studies have been performed whereby otherwise freely-running chemistry-climate models have had the day-to-day evolution of the dynamical fields constrained to follow the historical evolution as represented by reanalysis datasets. Here, Davis et al. present an analysis of different nudging schemes, using different combinations of variables or only nudging to zonal anomalies that are calculated in different ways, to assess the impact on the residual circulation of the lower stratosphere with a particular emphasis on how the nudged simulations differ with freely-running simulations and with the reanalysis dataset used for nudging. The study is very nicely performed and includes a convincing mechanistic diagnosis of the ways in which nudging of different variables affects the trends in tropical upwelling.

I really have no major concerns on the methodology or analysis presented here and my comments are mostly minor. One concern I do have, however, is the presentation of the effects of nudging zonal mean temperature on reproducing trends. In the abstract, at lines 21 – 23, the authors state that nudging to anomalies better reproduces trends in stratospheric upwelling, period. Taking a broader view, it would seem that nudging anomalies produces trends in upwelling that are more similar to the trends produced by the free-running (AMIP) simulation. This is clearly shown in Figure 3a, where the schemes that involve nudging to anomalies are much closer to the free-running simulation both in the TTL and in the lower stratosphere. As a consequence of the trends produced by the free-running AMIP simulation, the simulations nudged to zonal anomalies agree better with MERRA2 in the TTL but agree more poorly through the lower stratosphere. The degree of differences to the AMIP simulation across different nudging schemes also extends to the analysis of the EP-flux trends where it is stated (lines 392 - 394) that 'the response of the "no zonal-mean temperature nudging"simulations can be understood as the superposition of the "zonal-mean temperature nudging" simulation response - a slightly-incorrect MERRA2 response – and the AMIP response.' From both the analysis of trends and the analysis of the mechanism it appears that the response of schemes that do not affect the zonal-mean temperature produce trends that are more like the AMIP free-running simulation. I would suggest the authors should not overstate the conclusions of the effects of nudging temperature on the ability of the nudged model to reproduce the trends in the reanalysis as it would seem to depend significantly on the underlying behaviour of the free-running simulation.

Thanks - this is a good point. It was not consistent to discuss the EP flux results as AMIP-like or MERRA2-like but not discuss upwelling trends in this way.

In the abstract, we now state “None of the schemes substantially alter the structure of upwelling trends - instead, they make the trends more or less AMIP-like.”

The revised discussion of Figures 3 (lines 303-304) makes clear that the zonal anomaly nudging is only superior in the TTL.

Not at all a criticism, but more of a puzzled commentary. Figure 11 shows that nudging the zonal mean of temperature from MERRA produces temperature trends that disagree with the trends in MERRA. I can accept that the cause of the differences in the stratophere are not fully understood and may be related to unintended secondary circulations, but the anomalous trends found only in simulations that nudged the zonal mean temperature extend deep into the troposphere. In fact, the trends in the upper tropical troposphere appear to be three or four times larger than the trends in the same region found in MERRA2. Do you have any explanation for the discrepancy in trends in the troposphere and could there be links to the trends in the lower stratosphere?

The temperature difference between AMIP and MERRA2 maximizes around the level of net zero radiative heating in the TTL, where longwave cooling is close to zero and shortwave heating is at a minimum, whereas above and below the TTL the longwave cooling is substantially stronger [Fueglistaler et al. 2009]. So because the radiative terms are so small, it may be that this region is particularly sensitive to temperature perturbations and can more rapidly convert temperature nudging to perturbed heating. How this drives trends is quite unclear to us.

The extension into the troposphere may have something to do with convective parameterizations. MERRA2’s temperatures (and any meteorological input data set’s temperatures) will have convective effects baked in, so that nudging WACCM to those temperatures will result in a kind of double-counting as WACCM also has convection. Again, how this could contribute to the “wrong” temperature trends is unclear, but it could certainly present an inconsistency. It’s also possible that the incorrect trends in the TTL are just the decaying signal of this problem in the troposphere.

We think future work using a single model, like Smith et al. [2017], would be a more self-consistent system and might be the best avenue for understanding this problem.

Minor comments

Lines 98 – 99, for the case where WACCM is nudged towards anomalies it is stated 'To generate the nudging input, 6-hourly MERRA2 U, V, and T anomalies are calculated...'but a bit earlier, at lines 68 – 71, when the default nudging scheme is described it is stated that the MERRA2 reanalysis is supplied to the model every 3 hours. Is this difference real or just a typo? And if it is real, have the authors considered the differences in model behaviour that may be caused by reducing the frequency by a factor of two? Part of the motivation behind pointing this out is an open question about the effect of linearly interpolating in time between the available reanalysis.

This was indeed a typo and has been fixed - we use the 3-hourly output in all cases. However, the model's nudging scheme does interpolate the meteorological input to the current model time (lines 98-99).

Line 180 – Figure 1, I might suggest reducing the vertical extend to maybe 5 hPa so that the horizontal scale can be expanded. None of the other graphs extend beyond 30hPa.

While it's true the vertical extent is substantially higher than in the other plots, our intent was to begin with a macroscopic view of the whole stratosphere and display the large-scale structure of upwelling (the mass flux monotonically decreasing with height throughout the stratosphere), the remarkably rapid decrease with height of upwelling through the TTL compared to all heights above (indicating the strong poleward flow in the shallow branch), and the consistency of WACCM vs. MERRA2 upwelling (e.g., AMIP essentially always has more upwelling, throughout the entire stratosphere). We feel that including the log-scale difference plot alleviates the need to expand the axis, as it emphasizes the differences lower in the stratosphere.

Line 216 – missing 'A' in 'MERR2'

Thanks, this has been fixed.

Lines 239-240: Here it is stated that 'This all demonstrates that (incidentally) nudging zonal-mean MERRA2 temperatures has a negative impact on the upwelling trend morphology and magnitude.' I see how the findings of the correlation coefficient of trends with MERRA2 being largest for simulations that do not nudge the zonal-mean temperature supports the statement on morphology. But the magnitude of the trend

over large regions of the vertical profile shown in Figure 3 is closest to MERRA2 for the simulations that do nudge zonal temperature. The magnitude of the trends in UVT is closer to MERRA2 than UV, and UVT(ca) is closer than UV(ca) between 90 hPa and40 hPa. While the magnitude of the trends in UVT(za) are the furthest from MERRA2 everywhere above 90 hPa. The experiments where zonal average temperature is not nudged are closer to the AMIP simulation and this is an advantage in the TTL as theAMIP simulation has the largest positive trends and is thus closest to MERRA2. But producing trends closer to the freely-running AMIP simulation becomes a disadvantage higher up where the freely-running AMIP simulation produces more positive trends than MERRA2. Having read a bit further, I see how you eventually address this (and I particularly like Figure 4) but the statement at Lines 239-240 about the effect of nudging zonal-mean temperature on the magnitude of trends seems unsupported.

Thanks, another reviewer made this point as well. We have edited the discussion to be more specific to the TTL, and to make the point that the recent ozone trends in Ball et al. [2018], part of the motivation of this work, depend on the dynamics in this region being accurately resolved.

“This all demonstrates that (incidentally) nudging zonal-mean MERRA2 temperatures - the UVT, UVT L66, and UcaVcaTca simulations - has a negative impact on upwelling trend morphology and magnitude in the TTL. While it is true that the trends in the zonal anomaly nudging simulations are too positive above the TTL, key for constituent transport into the stratosphere and for recent ozone trends is the upwelling trend at and above the tropopause.”

Lines 284 - 285: It is stated here that the poor performance of the zonal anomaly nudging in reproducing variability in upwelling below 85 hPa suggests ‘a strong role for the zonal-mean circulation in transforming wave dynamics into zonal-mean momentum forcing and therefore upwelling (Fig. 6).’ Are you suggesting that the QBO has a role to play in upwelling in the TTL? Is there anything to be seen correlating the MERRA2 variability with that of the AMIPQBO run? [Okay, way down at Line 460 I see where you address the role of the QBO on variability in the TTL using UVT(za) nudging.]

Right, sorry that we leave this idea until the end of the paper. We tried to be linear in our discussion of the results, but obviously there are many cases like this where we don’t revisit an idea until later.

Line 363 – 366 – The caption for Figure 9 does not mention what is indicated by the thick black line. Is it the lapse rate tropopause?

Thanks - that is correct, and we have added this to the figure caption.

Line 414 – minor typo on '\hypothesize'

Thanks, fixed.