Response to Referee #1

Overall Comments: In this study the authors explore a broad range of specified-dynamics (SD) simulations in which WACCM is nudged to MERRA-2 meteorological fields in an attempt to quantify and understand the extent to which such SD simulations can reproduce upwelling trends in the underlying reanalysis. Given its implications for, among others, recent investigations into lower stratospheric ozone trends this study is very relevant. It is also an admirable attempt to understand in detail the mechanics of nudging, delving nicely into the momentum balance of upwelling and discrepancies that may arise in these balances among different nudged runs. In this sense the study does stand out as an attempt to address with more rigor than is standard the ways in which nudging can produce non-intuitive trends/variability/etc. For these reasons I recommend that this paper be accepted with minor revisions. However, there are still several key points that must be addressed. As these are more directed at the delivery and presentation of the results, and not related to fundamental problems that I have with the paper, I have not recommended "major revisions." Nonetheless, they need to be addressed.

Main Point 1: Throughout the authors argue that it is most desirable to preserve WACCM's free-running climatology (e.g. see discussion at the top of page 8, and various other places). Since this is a not a standard goal of nudging this needs to be better justified. In particular, I find the justification in lines 88-92 unsatisfying. Why is it bad that WACCM-SD reproduce the tropopause (or, more generally, temperature) structure of MERRA-2? Even if that impacts transport isn't that the point? What I would understand is if the authors argued that doing so creates dynamical inconsistencies in the circulation (assuming either the nudging tendency is large enough that it implies spurious vertical velocity analogous to the situation presented in Weaver et al. (1997)).Is this what the authors mean? Given that the use of nudging to climatological means is a central component of this work (and not conventional) I think this needs to be much better explained.

Weaver, Clark J., Anne R. Douglass, and Richard B. Rood. "Thermodynamic balance of three-dimensional stratospheric winds derived from a data assimilation procedure."Journal of the atmospheric sciences 50, no. 17 (1993): 2987-2993.

We apologize for being unclear and stating our goal before its motivation. We suspected that a key factor driving the inability of WACCM to produce the proper upwelling trends were disagreements in the climatologies of the meteorological input and WACCM. One can imagine it's easy for a model to have its variability nudged - but to nudge the mean could incur substantial undesired responses in the model, exactly analogous to the idea presented in Weaver et al. [1993] and some other references suggested by other reviewers.

We now explicitly state in the introduction on lines 78-81 that "Given that multidecadal trends in the earth system tend to be the residual of a balance of much larger terms, we hypothesize that disagreements between the climatologies of the input meteorology and the nudged model may lead to spurious circulations that interfere with upwelling trends."

We have also cleaned up the manuscript by revising any mention of "preserving WACCM's climatology" to more accurately reflect that this is not an end goal in itself, but rather a hypothesis that doing so may improve the upwelling trends in the TTL and lower stratosphere.

Weaver et al. [1993] is now cited in our discussion of spurious heating and temperature trends on line 519.

Main Point 2: I think the authors need to be much more cautious in generalizing the result that zonal mean temperature nudging should not be applied. As the conclusions read (especially point 1 spanning lines 400-404) the authors seem to suggest that this is a general result. However, given that the zonal mean temperature nudging trends are failing to reproduce MERRA-2 through their effects on eddies (via discrepancies in the meridional heat fluxes) I'm highly suspect that other nudging frameworks using different models (with different balances of resolved vs. parameterized momentum forcing of upwelling) will automatically corroborate these findings. In short, I think the authors need to state clearly how this conclusion depends very specifically on the particular way in which the momentum forcing in WACCM is driving w and how that may depend on horizontal/vertical resolution and other factors. Of course I notice that line 408-409 seems to direct these questions to future work but this is a bit unsatisfying.If the authors do not wish to do any test simulations (at higher horizontal resolution, for example) they should at minimum be very clear that these results are not likely to be generalizable to other nudging frameworks.*

We agree that the summary of results gave the impression we were prescribing a set of best practices generally, and not just for WACCM. We have added "In WACCM" to conclusion #1 on line 486.

However, we disagree that this result isn't likely to be generalizable to any other nudging simulation at a different resolution or in a different model. There is certainly some dependence of the resolved wave field on resolution, but it does not really change the dynamical regime or dominant balance of terms [Boville 1991; Held and Phillipps 1993; Béguin et al. 2013; Davis and Birner 2016]. The likelihood that different resolutions/models produce substantially different dynamical balances when averaged over almost half of the earth seems unlikely.

We do not have a robust mechanism that explains why the full nudging cannot produce the correct upwelling trends in the TTL/lower stratosphere. However, we have shown that when we craft the nudging scheme so that it does not nudge the climatological mean, or the climatological nor zonal mean, and temperature in particular, the upwelling trends are in better agreement with the input meteorology.

We have expanded our discussion on lines 496-501 to note our hypothesis moving forward. It is left as future work, for either us or other authors, to assess whether it is true because it is well outside the scope of this paper.

"We emphasize we have only assessed these conclusions using WACCM, and have not explicitly examined the impact of the nudging timescale, model resolution, or parameterizations. However, there is no obvious reason why this mechanism should be WACCM-specific. We offer the hypothesis, confirmed here in WACCM, that if there are differences in the climatologies of any nudged model and its input meteorology, upwelling trends will be more poorly reproduced when nudging zonal-mean temperatures than when not nudging to zonal mean temperatures, with the magnitude of error scaling with the difference in the zonal-mean climate."

Minor Comments:

Abstract, Ln. 22: I'm a bit confused why the goal is to "preserve WACCM's (free-running) climatology". The whole point of nudging is to draw the free-running model towards the reanalysis in as dynamically consistent a way as possible so I'm not sure why one would want to preserve a (biased) free-running climatological state. I'm sure there's a clear motivation for this but I couldn't identify one in the text (neither here nor in the sections later). See Major Comment 1.

See our response to your main comment.

Abstract Ln. 22: "climatological winds" -> Is that also just zonal or meridional too?

We are not sure what you mean by “also just zonal”, by not specifying “zonal” or “meridional” we thought “climatological winds” would imply the full horizontal wind field. We are happy to address this if you can clarify.

Ln.40: What does "the quality of the meteorological data" mean? Please specify.

This is a suggestion from Ball et al. [2018] as a possible contributor to errors in nudging schemes. Reanalyses are known to have difficulty conserving mass, momentum, and heat, so nudging can introduce inconsistencies that the model must find a way to balance. See our further discussion on lines 74-76.

Ln. 71: Does nudging occur everywhere?

We note on lines 107-109: “In this configuration, the model instead runs on 88 levels - 72 levels from the surface to the lower mesosphere, on MERRA2 hybrid levels, with a further 16 free-running levels in the upper atmosphere.” We have also added two plots to the supplemental information detailing the MERRA2 levels. The log-scale plot illustrates the lid of MERRA2 and the 16 free-running levels above.

Line 88: This wouldn’t happen, though, if one were to nudge "hard" to T (using, for example, a relaxation timescale of a few days, not 50 days). I’m not sure I really understand the point here. Sure, it would change WACCM’s tropopause (and other fields) but why is that necessarily a bad thing? Clearly, this would not be good if it were done in such a way that violated dynamical balance but that is more likely to be an artifact of the nudging machinery. What is fundamentally wrong about nudging to the full time-varying reanalysis field?

Any nudging term is an unphysical quantity; by construction it violates any momentum, heat, or mass balance in the model. Given that, hard nudging introduces even larger unphysical terms than weak nudging.

Nudging variability about the mean is probably not too unphysical, given that it is a temporary departure from the mean, but nudging the mean itself will engage the processes that set the modeled climate in the first place.

One way to limit the unphysical tendencies is to lengthen the nudging timescale, but eventually the model will no longer actually reproduce variability. We instead took a different approach to try to see which physical field was most responsible for the inability of WACCM to reproduce upwelling trends in the TTL and lower stratosphere.

Line 99: You write that three-hourly MERRA-2 input is used in Line 70 but six-hourly here. Which is it? If six-hourly why was the decision made to coarsen the resolution temporally?

This was a typo, it is 3-hourly in all cases.

Line 104: Again, can you please justify what you mean by "climatological anomaly nudging scheme is in theory..."? If the nudging was perfect (i.e. converged to assimilation) then it's not obvious to me that there's any fundamental problem with nudging to the full time-varying field.

See our response to your main comment #1, and also the response to the specific comment on line 88. We have cleaned up the manuscript so it is clear that our hypothesized way to improve the trends is to preserve the climate of WACCM.

Line 121: I am assuming other more standard tests have been done (i.e. vertical profile of nudging? changes in nudging timescale?). If so, it should be clarified that these have been done and they have not produced any satisfying simulation in which w reproduces w* in the underlying reanalysis (here MERRA-2).*

We think it's fair to argue that a study performing standard tests might not add as much to the literature as one performing more novel tests. Our goal was to determine which physical fields are most impactful, not the strength of the nudging.

We did investigate the timescale issue, but to get the timescale short enough to sample the phase space (for example, 12 hours) we had to decrease the physics timestep. This presented a conundrum, because shortening the physics timestep can change how the convective and gravity wave schemes impact the circulation. We now note on lines 92-96:

"We attempted to run WACCM at up to 10% per timestep (or, 5 hour timescale), but this required increasing the physics parameterization sub-cycling due to convective scheme errors - the "nsplit" parameter. Such simulations are not numerically comparable so we have chosen to avoid assessing the impact of nudging timescale, though it is known to have varied impacts [Merryfield et al. 2013, Hardiman et al. 2017, Orbe et al. 2017]."

The biggest constraint on the scope of this study was the amount of computing resources available to us. Repeating experiments at different timescales or resolutions was just not feasible.

Line 168: How did you calculate this from MERRA-2 (as shown in future figures?)? Where did you get all of the components (specifically the subgrid-scale wave momentum forcing)? And which product did you use? You indicated the third hourly fields initially but were six-hourly used here?

Thanks for this, we should have noted which product we used. We now state on lines 96-97 "WACCM is nudged toward the MERRA2 reanalysis instantaneous assimilation ("ASM") product", and state on lines 198-199, "Eddy fluxes are calculated every 3-hourly output interval in MERRA2 on native levels, while eddy fluxes are output as a monthly-mean value in WACCM".

Regarding the subgrid-scale forcings, we now note on line 199, "We use averaged output for zonal-means and gravity wave tendencies", to clarify that instantaneous fields are only used to calculate the eddy fluxes. As gravity wave tendencies can be highly variable in time, an average over the instantaneous values would not be accurate like a averaged output.

Lines 184-193: Again, I am confused. Don't you want to reproduce MERRA-2? See earlier comments.

Not necessarily, no, see our response to your main comment #1.

Figure 2 caption: The hatching definition is strange. Per the colorbar definition white contours in all panels should indicate regions where there is upwelling 100% of the time (i.e. fraction of 1). Why doesn't all hatching align with white?

The colorbar indicates any region with upwelling >= 90% of the time will be white; the hatching therefore is used to indicate the exceptional areas where there is exactly 100% upwelling.

Line 198: Is this frequency calculated daily/monthly/etc. Does the temporal sampling used to evaluate this measure matter?

As we note in the methods section, all values examined in this study are monthly-means. We are sure that the temporal sampling matters - in the annual mean there will clearly be upwelling everywhere in the tropics, but that washes out the variability we are interested in.

Line 200: Are you taking w directly from MERRA-2 or calculating offline in a consistent fashion as for the WACCM simulations? This relates to my earlier question about MERRA-2 mass flux estimates. How exactly are all measures derived from the MERRA-2 output?*

See our response to your comment on line 168.

Line 201: What if you just compare climatological annual mean w between WACCM and MERRA-2? That's more standard – does that show the same sort of difference (i.e. w* smaller in MERRA-2)? I find this "split" in upwelling frequency in MERRA-2 curious only because it doesn't appear to manifest in the climatology of w* (see Figure 10-3 in Bosilovich et al. (2015)). Note that in MERRA this region of anomalous downwelling was present but it was corrected in MERRA-2. This seems to be at odds with what the current study is showing. Can the authors explain this discrepancy? The easiest thing to do would be just to plot the climatology and see if you can reproduce the aforementioned figure…*

Bosilovich, M. G. (2015). MERRA-2: Initial evaluation of the climate.NationalAeronautics and Space Administration, Goddard Space Flight Center. Available at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf

Upwelling frequency (Fig. 2) is a complementary measurement to the upwelling mass flux (Fig. 1), and provides useful information about the permanency of upwelling at any location - something the average vertical velocity cannot describe.

See the plot below. The annual-mean w* is similar between the two. This suggests that MERRA2 often has periods of slight downwelling on the equator in the lower stratosphere, but small enough that it still has net upwelling in the annual mean.

Lines 248:251: So this is a really important conclusion – the lack of any convergence of the trends to MERRA-2 in Figure 4 is striking (and frustrating!). This is a merely a comment that I like this figure.

We agree this is important; another reviewer has suggested we describe the behavior of the nudging schemes as making the trends more or less AMIP-like, which can be easily seen in this figure.

Line 280: Indeed. Hence, why is this the primary goal of the paper? Again, more justification needed. See earlier comments.

See our response to your main comment #1; primarily because nudging to the full meteorology produces the wrong sign of the upwelling trend in the TTL, which probably influenced the conclusions of the Ball et al. [2018] paper which claimed models could not explain the reduction in lower stratospheric ozone seen in observations.

Line 313: Given the larger role played by the (parameterized) GWD in contributing to upwelling trends in WACCM does this imply that your conclusions will depend largely on horizontal and vertical resolution? One would think that as more of the waves contributing to w are resolved then the disparities with MERRA-2 (in terms of the physical mechanisms forcing the trends) will get smaller. Have you looked at SD simulations at different horizontal resolutions?*

It is possible that the GWD may drive different trends at higher vertical resolution - for example, at the 110-level WACCM in CESM2 that generates a spontaneous QBO. But the intermodel spread in the upwelling trends is due to resolved wave drag. We would not expect that over the range of reasonable horizontal resolutions, say 2.8 vs. 1 degree, that we will resolve substantially more convective gravity waves, which are dominant in the tropics - their scales are on the order of the latent heating within deep convective clouds, which is orders of magnitude smaller in scale.

Line 357: You can see that enhanced wave propagation clearly in the AMIP run but not so clearly in the AMIPQBO run (no evidence in NH extratropics)...please check.

Thanks, this evaded us. On lines 444-445 we now state “The AMIPQBO simulation exhibits this pattern only in the Southern Hemisphere.”

References

Béguin, A., O. Martius, M. Sprenger, P. Spichtinger, D. Folini, and H. Wernli, 2013: Tropopause level Rossby wave breaking in the Northern Hemisphere: A feature-based validation of the ECHAM5-HAM climate model. *Int. J. Climatol.*, 33, 3073-3082, doi: 10.1002/joc.3631.

Boville, B. A., 1991: Sensitivity of simulated climate to model resolution. *J. Climate*, 4, 469-485, doi:10.1175/1520-0442(1991)004,0469:SOSCTM.2.0.CO;2.

Davis, N. and T. Birner, 2016: Climate model biases in the width of the tropical belt. *J. Climate,* 29, 1935-1954, https://doi.org/10.1175/JCLI-D-15-0336.1.

Held, I. M., and P. J. Phillipps, 1993: Sensitivity of the eddy momentum flux to meridional resolution in atmospheric GCMs. *J. Climate*, 6, 499-507, doi: 10.1175/1520-0442(1993)006,0499:SOTEMF.2.0.CO;2.