We thank the three anonymous reviewers for their constructive comments and discussion. Below follows our point-by-point response to all referee comments, while the tracked-changes manuscript and updated supplemental figures are posted as separate comments.

-The authors

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 aformentioned 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."

Response to Referee #2

The authors of this study probe the impacts of nudging WACCM towards MERRA-2 meteorology in a number of Specified-Dynamics (SD) simulations. The novelty of this study lies in exploring discrete nudging "flavours" in order to determine the degree in which they can accurately simulate the mean state, the interannual variability and the upwelling trends of the residual circulation seen in MERRA-2 reanalysis product. This is a commendable attempt to understand how the implementation of various nudging frameworks affects the self-consistency of a chemistry-climate simulation highlighting the possible implications regarding the accuracy of transport processes associated mainly with ozone trends in the lower stratosphere as seen in recent studies. Apart from the question of which nudging scheme appears to better reproduce MERRA-2 (and to a second degree the free-running WACCM) spatial and temporal upwelling characteristics, the study comprehensively investigates the attribution of these discrepancies by shedding some light on the physical drivers of the upwelling trends. Nudging acts as an additional non-physical tendency in the model equations and it is quite important to evaluate their potential artificial effects on the model dynamics. Therefore, this study greatly improves the understanding regarding the degree of impact arising from the choice (or not) of a particular nudging scheme, albeit in a single model framework. For the reasons above, I recommend this study to be accepted and published with minor revisions. There are a few points that I think they should be addressed by the authors, more to do with enhancing the introduction of the paper by adding a substantial amount of discussion on the nudging studies literature.

Specific comments

1. I find the introduction to be relatively short and lacking in terms of literature related to nudging studies. In order to set the scene better and highlight that nudging studies are not just used for lower stratospheric ozone trends, more references would be extremely valuable to the reader. One of the first attempts to obtain a comparison between a GCM relaxed towards analyses and the analyses themselves is detailed in Jeuken et al. (1996). There are multiple studies looking at specific meteorological events, such as van Aalst et al. (2004) looking into the Arctic winter transport processes at the end of the 20th century or the SSW during 2009-2010 winter in Akiyoshi et al. (2016). Similarly, nudged simulations were used to focus on the effects of volcanic eruptions on stratospheric tracers such as water vapour in Loffler et al. (2016), to infer the global-mean volcanic effective radiative forcing over the satellite era in Schmidt et al. (2018), as well as to estimate the chemical effects of monsoon circulations on volcanic sulphur *particles seen in Solomon et al. (2016). Additionally, Solomon et al. (2015), studied the polar ozone depletion in 2011, using a nudged version of WACCM. In fact, the latter three studies, used a previous version of CESM1-WACCM which is nudged towards an older version of the reanalysis (MERRA), with and without nudging the temperature respectively. A bit of discussion regarding the differences between the nudging schemes used in the aforementioned studies and in the current study certainly wouldn't hurt. I would also recommend some discussion with respect to the differences in between a CTM and an CCM-SD model run as nudging, especially when trying to interpret the differences between Ball et al. (2018) and Chipperfield et al. (2018). CTMs are forced directly with the full 3-D circulation from reanalyses and after many years of optimizations they have been proven quite successful at simulating stratospheric tracers on various timescales (Chipperfield, 1999; Mahieu et al. 2014). On the other hand, CCMs are much more recent tools and exhibit deep-rooted differences when compared to a CTM when looking at their tracer advection, and some discussion regarding their differences would be also a good idea.*

Thank you very much for this reference list! We have included these references, in addition to a few others, in the revised introduction. We tried to keep our introduction short and to the point, but realize it may have left out some details.

We have expanded the introduction on lines 48-57 (a discussion of different "flavors" of nudging), lines 66-67 (errors in nudging schemes), and lines 71-76 (the sensitivity of residual circulation trends and CTM's). We think this will give a better background on the differences between CTMs and nudged CCMs and the difficulties in constraining the residual circulation.

2. There is a hint in the study that the overarching aim of the study is to capture the free-running WACCM climatology rather than the climatology of MERRA2 - some justification is required if that is the case. Although the research question (reproducing the residual circulation variability and trends of MERRA2) of the study is clearly stated in the introduction (lines 56-57), there are various places across the text where the message appears to be the reproduction of free-running WACCM. As an example, it is stated that the nudging will create conflicts due to the differences between WACCM and MERRA2 underlying climatologies (explained in section 2 - lines 88-92). I would suggest rephrasing the relevant parts where the point seems to be lost. The point of nudging is exactly to reproduce the reanalyses themselves albeit exhibiting spurious features in *the stratospheric residual circulation. It should be noted that between reanalysis products and among different estimates of wbar* there lies significant uncertainty with respect to upwelling trends as seen in Abalos et al. (2015) as well as Kobayashi and Iwasaki (2016).*

As we discussed in our response to Referee #1, we placed the goal before the motivation in our discussion of (not) nudging the mean. Rather than suggesting that preserving the climatology is the goal, we know explicitly state our hypothesis on lines 78-81 that any differences in the climatology between the input meteorology and the model may lead to spurious circulations and trends.

Thank you for suggesting we discuss the variability of the residual circulation across reanalysis products; lines 74-76 connect the variability in the residual circulation trends to some recent work suggesting the tropospheric meridional circulation in reanalyses is in an unphysical balance.

3. There's no mention of the MERRA2 output the authors have used throughout this study. How did you calculate the TEM diagnostics? Did you perform the calculation on the native MERRA2 levels? What output have you used? How about its temporal frequency? This information needs to be included by describing all the above in either section 2 or section 3.

Thanks, this was an oversight. We now state on line 97 that we use the ASM product, and on lines 198-199 state that eddy fluxes are calculated from 3-hourly instantaneous output on native levels.

Minor comments

Line 22 : See discussion above (point 2)

Here and elsewhere, we've cleaned up the manuscript to better reflect our hypothesis.

Line 55 : Some more references are needed here such as Abalos et al. (2015) and Kobayashi and Iwasaki, (2016) with findings related to the discrepancies in the trends of the residual circulation between reanalysis.

Thanks - these references better connect the Chemke and Polvani paper to the residual circulation and its trends.

Line 62 : In the context of this study is quite clear that nudging zonal-mean temperatures alters the meridional eddy momentum and heat fluxes in the TTL without being successful in simulating the underlying MERRA2 trends. Applying a thermal nudging (temperature) could potentially lead to a sustained spurious heat source in the model, which leads to a stronger BDC in the lower stratosphere as seen in Miyazaki et al. (2005) with a different model. However, the last sentence is quite strong as a statement and generalizes a result which is model specific. Therefore, I would recommend rephrasing this bit so it doesn't strike as misleading.

We agree this was too strong of a statement and now say, "We find that not nudging zonal mean temperatures results in the best reproduction of upwelling trends, while nudging zonal mean temperatures tends to degrade these trends, consistent with our hypothesis."

We have added Miyazaki et al. [2005] to our discussion of temperature trends and spurious heating on line 519.

Line 69 : When nudging, the choice of the relaxation timescale can play an important role (Merryfield et al. 2013), although there is no consensus that a specific timescale necessarily leads to an improvement (Hardiman et al. 2017). I'm aware of a standard full WACCM-SD CCMI simulation using a nudging timescale of 5 hours - towards MERRA though (Orbe et al. 2018; Orbe et al. 2019 under review in ACPD), have you performed any additional runs with this timescale (or have plans) to compare with?

We did attempt to run with a more aggressive nudging timescale of 5 hours (10% per timestep). However, this required a substantial increase in the physics timestep subcycling to prevent instabilities in the convection scheme. The resulting simulations would not be numerically comparable to our existing simulations, so we did not investigate the issue of nudging timescale. This is now discussed on lines 92-96.

Lines 84-87 : It would be very helpful (at least to me) if you could include a figure (in the supplement) showing the vertical profiles of the pressure levels in WACCM-L66/L88 and MERRA2 to better highlight their spacing differences throughout the depth of the atmosphere.

That's a great idea - we've added two versions of this figure, one with a linear pressure scale and one with a logarithmic pressure scale to emphasize the middle and upper atmosphere.

Lines 98-99 : Here you mention that 6-hourly MERRA2 anomalies are used for nudging WACCM, I assume you are interpolating in time to nudge every 3 hours (see line 71)?

This was an error, we used 3-hourly output in all cases. The model then interpolates the 3-hourly output to the current model time to determine the nudging terms.

Line 105 : Please clarify that this improvement refers specifically to aerosol-climate interactions in Zhang et al. (2014).

We have rewritten all references to preserving the mean climate of WACCM, so that it is clear that we see it as a potential solution to the problem of reproducing TTL trends - not something that is an end in and of itself. Here, we are simply citing Zhang et al. [2014] to give credit to the idea of climatological anomaly nudging - hopefully there is no sense of an implied "improvement" that needs clarification.

Lines 155 & 159 : Please correct the reference - it is : Hardiman et al. (2010)

Thanks, this was in error.

Line 164 : SF6 can be considered linear only by approximation, characterised by a fast growth rate and there needs to be a correction for this. See Garcia et al. (2011).

Thanks, we've added this reference and a clarification to line 217.

Line 224 : This line refers to Figure 3b, where you calculate the difference in the trends compared to MERRA2. For clarity, it would be better to rephrase and not use the word negative but either smaller or bigger to reflect their differences. E.g. in the TTL all WACCM runs + MERRA 2 have positive trends as seen in Figure 3a, and the term negative might be misleading as the trends are just smaller (but still positive).

That's a good point, it was somewhat ambiguous the way it was rewritten. It now states "The upwelling trends in all WACCM runs tend to be smaller in the TTL and larger aloft" on line 285.

Line 228 : " The standard UVT…" - Clarify that this holds true for both versions (L66 + L88).

Yes, good idea, it now states "The UVT L88, UVT, and UVT climatological anomaly…".

Lines 241 - 252 : Excellent discussion (and figure)!

Thank you.

Line 309 : By gravity wave (GW) momentum forcing are you referring to all the parameterizations? Meaning orographic (OGW) + non-orographic (NOGW) gravity wave drag put together? Please clarify.

Yes, correct, it's the sum of all gravity wave sources in the model. We have clarified this by stating, "However, gravity wave momentum forcing from orographic and non-orographic waves…"

Line 311 : McLandress and Shepherd (2009), using CMAM, show the total contribution of both resolved and parameterized wave drag occuring at the edge of the pipe in the lower stratosphere in boreal winter in their figure 18. However, this lumps together resolved (major contributor at the edge of the pipe) and all GW parameterizations while the orographic gravity wave drag contributes more in the NH mid-latitudes instead. I would suggest caution drawing parallels to this result which remains model specific. Different parameterizations lead to various magnitudes of contributions to the upwelling throughout the stratosphere and specifically for the versions of CMAM over the past decade, it has been shown that the NOGW contributes negatively in the upwelling in the lower stratosphere in SPARC, 2010 and more recently in Chrysanthou et al. (2019).

This is a fair point - it does appear this is a case where different models and different generations of models provide different answers. We aren't focused on the physical plausibility of any of the trends here, merely their correspondence with those in MERRA2, so rather than expand the discussion to note this is not a robust feature of models, we have simply eliminated it.

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.

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.

A Comprehensive Assessment of Tropical Stratospheric Upwelling in Specified Dynamics CESM1 (WACCM)

Nicholas A. Davis[1,a], Sean M. Davis[2], Robert W. Portmann[2], Eric Ray[1], Karen H. Rosenlof[2], and Pengfei Yu[1,b]

[1]University of Colorado Cooperative Institute for Research in Environmental Sciences (CIRES) at the NOAA Earth System Research Laboratory (ESRL) Chemical Sciences Division, Boulder, CO, USA

[2]NOAA Earth System Research Laboratory (ESRL) Chemical Sciences Division, Boulder, CO, USA

[a]*Current affiliation*: Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

[b]*Current affiliation*: Institute for Environmental and Climate Research, Jinan University, Guangzhou, GD, China

Correspondence to: Nicholas A. Davis (nadavis@ncar.edu)

Abstract. Specified dynamics (SD) schemes relax the circulation in climate models toward a reference meteorology to simulate historical variability. These simulations are widely used to isolate the dynamical contributions to variability and trends in trace gas species. However, it is not clear if trends in the stratospheric overturning circulation are properly reproduced by SD schemes. This study assesses numerous SD schemes and modeling choices in the Community Earth System Model (CESM) Whole Atmosphere Community Climate Model (WACCM) to determine a set of best practices for reproducing interannual variability and trends in tropical stratospheric upwelling estimated by reanalyses. Nudging toward the reanalysis meteorology as is typically done in SD simulations does not accurately reproduce lower stratospheric upwelling trends present in the underlying reanalysis. In contrast, nudging to anomalies from the climatological winds or anomalies from the zonal mean winds and temperatures better reproduces trends in lower stratospheric upwelling, possibly because these schemes do not disrupt WACCM's climatology. None of the schemes substantially alter the structure of upwelling trends - instead, they make the trends more or less AMIP-like. An SD scheme's performance in simulating the acceleration of the shallow branch of the mean meridional circulation from 1980-2017 hinges on its ability to simulate the downward shift of subtropical lower stratospheric wave momentum forcing. Key to this is *not* nudging the zonal-mean temperature field. Gravity wave momentum forcing, which drives a substantial fraction of the upwelling in WACCM, cannot be constrained by nudging and presents an upper-limit on the performance of these schemes.

1 Introduction

Stratospheric ozone loss has been halted by a concerted international effort to eliminate emissions of ozone-depleting substances under the Montreal Protocol [WMO, 2018]. While ozone is recovering in the upper stratosphere, there is some

Deleted: Chemistry

Deleted: expectedly changes the model's mean upwelling compared to its free-running state, but

Deleted: preserves WACCM's climatology and indication of a decline in the tropical lower stratosphere since the late 1990's when ozone depleting substance emissions peaked [Ball et al. 2018]. Ozone in this region is strongly mediated by the vertical advection of ozone-scarce tropospheric air by the residual circulation - the wave-driven, thermally-indirect overturning circulation of the stratosphere [Butchart 2014, and references therein] - as well as eddy mixing of ozone-rich air from the extratropical lower stratosphere [Abalos et al. 2013].

Ball et al. [2018] evaluated two specified dynamics (SD) simulations to assess this unexpected decline. In SD simulations, a climate model's circulation is nudged toward the meteorology of an atmospheric reanalysis. The goal of such simulations is to constrain the known variability of the atmospheric circulation to better isolate the role of chemical processes, insofar as the reanalysis meteorology is reliable. Ball et al. [2018] found that their SD simulations were unable to reproduce an observed decline in lower stratospheric ozone, and offered several explanations, including emissions of short-lived ozone depleting substances, the quality of the meteorological data, and the quality of model-simulated tracer transport in the lower stratosphere. However, other studies using a chemical transport model [Chipperfield et al. 2018] and an SD-like model simulation [Wargan et al. 2018] found that observed meteorological variability could explain the recent changes in stratospheric ozone, with some disagreement on the sign of the change in the tropical lower stratosphere.

By construction nudging schemes target each model's prognostic variables, which are generally either the horizontal winds or vorticity and divergence, often in addition to temperature, surface pressure, and occasionally specific humidity. Both Van Aalst et al. [2004] and Loffler et al. [2016] nudged surface pressure and vorticity, which characterize the balanced flow, stronger than they nudged divergence and temperature. The motivation for such a scheme is that the essence of dynamical variability is the evolution of the large-scale balanced flow, and that model physics like convection should govern shorter time-scale variability in the thermodynamics and the unbalanced flow. Schmidt et al. [2018] nudged only the horizontal winds in CESM to assess the freely-evolving temperature response to volcanic aerosols, while Solomon et al. [2015] and [2016] nudged horizontal winds and temperatures to assess polar ozone and heterogeneous chlorine chemistry in the TTL, both of which require an accurate reproduction of absolute temperature. But from any sampling of past studies, it is difficult to understand the impact of the peculiarities of each nudging scheme on simulated dynamics, transport, and chemistry.

Lower stratospheric ozone is sensitive to modes of natural variability including the El Niño-Southern Oscillation [Randel et al. 2009, Diallo et al. 2018], which can drive variations in tropical stratospheric upwelling via vertical shifts in gravity wave momentum forcing [Calvo et al. 2010]. Regardless of the interpretation of recent ozone variability, the question remains, "why is there a discrepancy between the Ball et al. [2018] simulations and the simulations in Chipperfield et al. [2018] and Wargan et al. [2018]?" One possibility is that one or more of these schemes may not be reproducing the mean meridional circulation trends in the input meteorological data, which would present a critical flaw to their purported ability to constrain circulation variability. There is emerging evidence that nudging tends to increase the inter-model spread in measures of the residual circulation and chemical transport [Orbe et al. 2017, Orbe et al. 2018, Chrysanthou et al. 2019]. This is because the schemes do not perfectly reproduce dynamical variability, with substantial errors in the residual circulation [Akiyoshi et al. 2016]. Further, nudging - which is often implemented as a relaxation term - is substantially less

Deleted:

sophisticated than 3-dimensional data assimilation, which itself tends to degrade model performance in simulating the stratospheric age of air [Meijer et al. 2004].

Chemical transport models, which do not prognostically model the atmosphere but instead directly ingest the meteorology, are not as susceptible to these errors [Chipperfield et al. 1999, Mahieu et al. 2014]. However, they must still assess vertical motion either through mass continuity or diabatic heating, in addition to estimating convective transport [Stockwell and Chipperfield 1999, Chipperfield 2005]. It is also possible that reanalyses contain spurious circulation trends that cannot (and perhaps should not) be reproduced by climate models [Chemke and Polvani 2019], given the range of trends in the residual circulation [Abalos et al. 2015; Kobayashi and Iwasaki 2016].

Here we pose a simple question: do nudging schemes reproduce the variability and trends in the mean meridional circulation of the input reanalysis meteorology? If not, what can be done to improve that representation? Given that multi-decadal 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. In this study, we examine one of the models and nudging schemes employed in Ball et al. [2018], the Community Earth System Model-Whole Atmosphere Community Climate Model (CESM (WACCM)), analyzing a series of nudging experiments to assess the impact of various modeling choices and nudging scheme variations on tropical stratospheric upwelling trends. We find that *not* nudging zonal mean temperatures results in the best reproduction of upwelling trends, while nudging zonal mean temperatures tends to degrade these trends, consistent with our hypothesis.

Deleted: .

Deleted: ¶

2 Model

We use version 1.2.2 of CESM (WACCM) [Marsh et al. 2013] using the Community Atmosphere Model Version 4 (CAM4) [Neale et al. 2013] finite volume dynamical core [Lin 2004], covering an altitude range of 0 to 140 km on a 1.9x2.5 degree grid. All of our simulations are atmosphere-only experiments, with prescribed sea surface temperatures (SST's), Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al. 2015) historical and RCP4.5 fixed lower boundary conditions, time-varying solar forcings, and volcanic aerosols. The default WACCM-SD scheme nudges horizontal winds and temperatures by 1% per timestep, which, for a timestep of 30 minutes, corresponds to a nudging timescale of 50 hours. 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]. WACCM is nudged toward the MERRA2 reanalysis instantaneous assimilation ("ASM") product [Gelaro et al. 2011], with meteorological input supplied to the model every 3 hours on 72 hybrid levels. The default nudging is a simple relaxation term computed by linearly interpolating the meteorological input between meteorological times. Surface pressure is prescribed based on MERRA2 meteorology, but allowed to vary sufficiently to ensure mass conservation [Lamarque et al. 2012, and references therein]. MERRA2 surface geopotential, surface wind stress, surface latent and sensible heat fluxes, ice and ocean grid fractions, and surface skin temperature are also incorporated into the scheme to ensure consistency.

CAM4 WACCM has 66 native levels. However, standard practice is to run the SD simulations on reanalysis levels, the implicit motivation being that the reanalysis meteorology does not need to be interpolated before running the simulations [Marsh 2011, Chandran et al. 2013, Verronen et al. 2016]. 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 (Fig. S1 and S2). This substantially impacts cloud heating and gravity wave momentum forcing, in part because the MERRA2 grid triples the number of model levels in the boundary layer and lower troposphere (Fig. S3). In addition to the standard SD simulations with 88 levels, we test whether running the model on these non-native levels impacts the resolved circulation and its trends by nudging WACCM on its 66 native levels with interpolated MERRA2 meteorology (which interpolates MERRA2 meteorology from 72 to 46 levels, with the remaining 20 levels above MERRA2's lid). It is worth noting that while the number of levels in MERRA2 in the troposphere is much greater than in WACCM at its native resolution, the levels and their spacing correspond quite closely in the upper troposphere to the mid-stratosphere, such that any differences in the circulation in that region are likely to be driven from below.

Standard practice is to nudge toward the actual reanalysis meteorology, but this could create conflicts as WACCM and MERRA2 have different climatologies. As an example, the tropical lapse-rate tropopause is nearly 1 km higher in WACCM than in MERRA2, so standard nudging might disrupt WACCM's tropical tropopause height and cold-point temperature, which will go on to impact stratospheric water vapor, ozone chemistry, and transport. Such artificial disruptions to the climate will engage the feedbacks that stabilize WACCM around its free-running climate, potentially interfering with upwelling trends. Further, even modern reanalyses exhibit spurious global discontinuities in zonal-mean temperatures and winds due to satellite transitions, so it may be a valid choice to avoid nudging the zonal mean all together [Long et al. 2017].

Deleted: activate

Deleted: .

As a solution, we test a scheme which nudges WACCM toward climatological anomalies, rather than the full meteorological reanalysis fields [Zhang et al. 2014]. Monthly-mean seasonal cycles of U, V, and T are calculated in WACCM and MERRA2 and interpolated to daily values using cubic splines. This ensures a smooth annual cycle, avoids sampling issues, and only requires monthly-mean output. The annual cycle is calculated from 1980-2017 inclusive, from MERRA2 and from WACCM in free-running mode with 66 native levels and prescribed historical sea surface temperatures, hereafter "WACCM AMIP" (AMIP referring to the Atmospheric Model Intercomparison Project [Gates 1992]). To generate the nudging input, 3-hourly MERRA2 U, V, and T anomalies are calculated by subtracting the MERRA2 climatology from the MERRA2 meteorology, and then adding this anomaly to the CESM climatology. Explicitly, for a given field X at some point in level-latitude-longitude (σ, ϕ, λ), the nudging tendency is given as

Deleted: anomalies from the climatology

Deleted: 6

$$\frac{d}{dt}X_{nudge}(\sigma,\phi,\lambda) = \frac{1}{\tau}\left(X_{WACCM}(\sigma,\phi,\lambda) - \left(X_{MERRA2}(\sigma,\phi,\lambda) - \overline{X_{MERRA2}(\sigma,\phi,\lambda)} + \overline{X_{AMIP}(\sigma,\phi,\lambda)}\right)\right) \quad \text{(Eq. 1)}$$

where the overbar indicates the climatological value of X and τ is the nudging timescale.

While this climatological anomaly nudging scheme in theory better preserves the climatology than nudging toward the actual reanalysis meteorology, it still nudges the zonal mean. As the stratospheric circulation is wave-driven, the most important aspect of the circulation that must be reproduced in order to reproduce trends in stratospheric upwelling are the resolved wave momentum forcings (gravity wave momentum forcing is parameterized and cannot be nudged). We therefore test an additional scheme wherein only the zonal anomalies are nudged in WACCM. This allows WACCM to freely model the zonal-mean circulation and climate, bypassing differences in the climatologies of the input meteorology and WACCM and limiting the influence of spurious reanalysis trends and features. As this scheme does not nudge the zonal-mean winds, it could be combined with a separate QBO nudging scheme, and would also not disrupt a spontaneous QBO [Garcia and Richter 2019]. Explicitly, for a given field X at some point in level-latitude-longitude (σ, ϕ, λ), the nudging scheme is given as

$$\frac{d}{dt}X_{nudge}(\sigma,\phi,\lambda) = \frac{1}{\tau}\left(\left(X_{WACCM}(\sigma,\phi,\lambda) - [X_{WACCM}](\sigma,\phi)\right) - \left(X_{MERRA2}(\sigma,\phi,\lambda) - [X_{MERRA2}](\sigma,\phi)\right)\right) \quad \text{(Eq. 2)}$$

Deleted: s in theory an improvement over

Deleted: -

Deleted: climate

Deleted: Modern reanalyses can still exhibit spurious global discontinuities in zonal-mean temperatures and winds due to satellite transitions [Long et al. 2017].

Deleted: in WACCM

Deleted: potentially

Deleted: spurious

Deleted: it

where the vertical brackets indicate the zonal mean value of X. An advantage of this method over climatological anomaly nudging is that it does not require substantial preprocessing and a reference AMIP simulation, but it does require source code modification as the zonal mean of a given field in the model and input reanalysis meteorology must be calculated on-line at every time step.

Deleted: Thus, the zonal mean of a given field in the model and input reanalysis meteorology must be calculated at every time step.

Deleted: .

The final dimension along which we test SD simulations is the subset of nudged variables. While standard practice is to nudge U, V, and T, it is unclear which, if any, are most important for reproducing past variability and trends. To gain further insight, we tested several other combinations of nudging variables – UV, UT, and VT. We found that UT nudging does not constrain the meridional circulation (as might be expected), and found that VT nudging is too similar to UVT nudging to warrant further investigation (probably because the zonal-mean zonal winds are strongly constrained by temperature through geostrophy). For simplicity, we ignore these other combinations of variables and focus only on UVT and UV.

We also consider a simulation identical to the AMIP run, but with Quasi-biennial Oscillation (QBO) nudging in the tropics, designated "AMIPQBO". The free-running version of CESM 1.2.2 (WACCM) does not spontaneously generate a QBO, instead requiring a scheme to nudge the tropical zonal winds to the observed QBO.

Our full set of WACCM simulations is documented in Table 1. All simulations are run from January 1, 1980 to December 31, 2017.

Short name	Case	Nudging variables	Vertical levels
AMIP	AMIP	None	66
AMIPQBO	AMIPQBO	U (tropics only)	66
UVT L88	UVT L88	U, V, T	88 (MERRA2 levels)
UV L88	UV L88	U, V	88 (MERRA2 levels)

UVT	UVT	U, V, T	66
UV	UV	U, V	66
$U_{ca}V_{ca}T_{ca}$	UVT climatological anomaly	U, V, T climatological anomalies	66
$U_{ca}V_{ca}$	UV climatological anomaly	U, V climatological anomalies	66
$U_{za}V_{za}T_{za}$	UVT zonal anomaly	U, V, T zonal anomalies	66
$U_{za}V_{za}$	UV zonal anomaly	U, V zonal anomalies	66

Table 1: Full list of WACCM simulations examined in this study, the MERRA2 variables with which they are nudged, the number of vertical levels, and the short name by which the figures refer to the simulation. See text for further information.

3 Methods

The Transformed Eulerian Mean (TEM) residual circulation, an estimate of the mean meridional mass circulation, is calculated by solving the zonal momentum and mass balance equations so that different physical contributions to trends in tropical upwelling can be parsed directly. The "downward control" principle (Haynes et al., 1991) relates the steady-state zonally averaged vertical mass transport at any level to the vertically-integrated wave forcing in the column above. If one relaxes the steady-state constraint, the residual circulation streamfunction can be diagnosed similarly from the sum of the time tendency of the zonal wind and the vertically-integrated wave forcing in the column above [Randel et al. 2002, Abalos et al. 2013], and is given by

$$\Psi^*(p,\phi) = \frac{1}{g}\int_p^0 \left\{\frac{a^2 cos^2(\phi)((acos(\phi))^{-1}\nabla\cdot\vec{F}+[D_{gw}]-[u]_t)}{[m]_\phi}\right\}_{\phi=\phi(p')} dp' \qquad \text{(Eq. 3)}$$

where Ψ^* is the TEM residual circulation streamfunction, a is the radius of the earth, p is the pressure, ϕ is latitude, g is the acceleration due to gravity, $\vec{F}$ is the Eliassen-Palm (EP) flux vector, D_{gw} is the subgrid-scale (gravity) wave momentum forcing, u is the zonal wind, m is the angular momentum per unit mass, and the subscripts outside of the brackets indicate derivatives. The vertical integral is computed along lines of constant angular momentum by interpolating all fields from latitude to angular momentum space, with p' denoting the pressure element along isolines of angular momentum. All calculations are performed in pressure space to ensure consistency with WACCM's and MERRA2's vertical discretization using 12 month low-pass filtered monthly-mean fields, primarily to stabilize the integration along angular momentum contours. Eddy fluxes are calculated every 3 hours in MERRA2 on native levels, while eddy fluxes are output as a monthly-mean value in WACCM. We use averaged output for zonal-means and gravity wave tendencies. For illustrative purposes, we display the

Deleted: ,

Deleted:

Deleted: values

approximate altitude, in addition to the pressure, by showing the average of the geopotential height from 30S to 30N in the AMIP simulation. Angular momentum per unit mass in the shallow atmosphere approximation is given by

Deleted: taking

$$m = a\cos(\phi)\,(u + \Omega\, a\cos(\phi)) \qquad \text{(Eq. 4)}$$

where Ω is the rotation rate of the Earth. The meridional and vertical components of the Eliassen-Palm flux are given by the isobaric coordinate version of Equations 25-26 in Hardiman et al. [2010],

Deleted: 09

$$F^{(\phi)} = a\cos(\phi)\left([u]_p \frac{[v'\theta']}{[\theta]_p} - [u'v']\right) \qquad \text{(Eq. 5)}$$

$$F^{(z)} = a\cos(\phi)\left(\left\{f - \frac{1}{a\cos(\phi)}\frac{\partial}{\partial\phi}([u]\cos(\phi))\right\}\frac{[v'\theta']}{[\theta]_p} - [u'\omega']\right) \qquad \text{(Eq. 6)}$$

where f is the Coriolis parameter, v and ω are the meridional wind and pressure velocity, and θ is the potential temperature. See Hardiman et al. [2010] for a description of the divergence operator in Eq. 3.

Deleted: 09

The total upwelling mass flux through the tropical lower stratosphere is a useful metric for the strength of the residual circulation as it diagnoses the total tropospheric air mass entering the stratosphere from below. This air is low in ozone, which is efficiently produced in the stratosphere [Fueglistaler et al. 2009], higher in chlorofluorocarbons and other chlorinated species, which can destroy polar ozone [Tegtmeier et al. 2016], and retains a record of the time the air entered the stratosphere based on the concentrations of gases with linearly-increasing tropospheric concentrations, such as sulfur-hexafluoride [Linz et al. 2017] (a correction for non-linearity is given by Garcia et al. [2011]), or based on gases with a strong seasonal cycle, such as water vapor [Mote et al. 1996]. The total upwelling mass flux, hereafter simply "upwelling", convolves the speed of the upward circulation with its spatial extent, and is expressed as in Rosenlof [1995] as

Deleted: [

$$M^*(p) = 2\pi a(max[\Psi^*(p,\phi)] - min[\Psi^*(p,\phi)]) \qquad \text{(Eq. 7)}$$

All trends are computed with linear least-squares fits. Statistical significance of trends is assessed with two-sided Student's t-tests using each sample's effective degrees of freedom estimated through its lag-1 autocorrelation, while a test for the difference of means is used to assess differences in the climatology.

4 Results

Upwelling monotonically decreases with altitude from a peak of 20×10^9 kg/s at 120 hPa (15.5 km) in the tropical tropopause layer (TTL) to 0.5×10^9 kg/s at 1 hPa (47.7 km) in the upper stratosphere, indicating divergent poleward flow at all levels (Fig. 1). The difference in upwelling between MERRA2 and AMIP is at a maximum in the TTL and approaches zero at 20 hPa, indicating that MERRA2 has less total mass outflow in the TTL and lower stratosphere than AMIP – a weaker residual circulation shallow branch (the outflow is proportional to the derivative of the mass flux with respect to pressure).

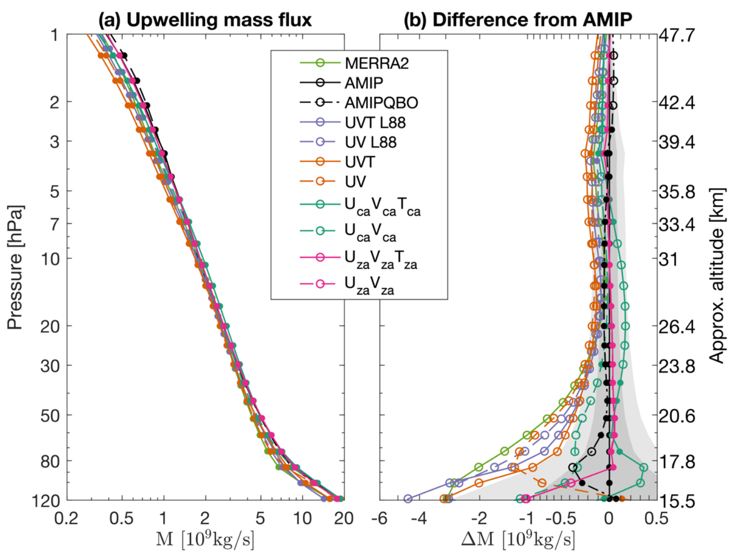

Figure 1: Tropical upwelling mass flux (a) climatology and (b) difference from the AMIP simulation. Dark and light shading indicate 1- and 1.96-times the standard error. Open circles in (b) indicate the average is statistically significantly different from the AMIP simulation at the 95% confidence level.

In the lower stratosphere, the upwelling in the standard nudging simulations (UV/T L88 and UV/T) falls between the weak upwelling in MERRA2 and the strong upwelling in the AMIP simulation, with the upwelling generally falling outside of the standard error of the AMIP simulation. By contrast, the upwelling in the climatological anomaly and zonal anomaly simulations is well within the standard error bounds, with the UV and UVT zonal anomaly nudging simulations most similar to AMIP over most of the stratosphere (the upwelling in the two zonal anomaly simulations is nearly identical). There are minor differences between the upwelling in AMIP and AMIPQBO, except in the TTL where AMIPQBO has weaker upwelling near the cold point. Across all nudging varieties, UVT nudging tends to have faster upwelling in the upper TTL and lower stratosphere than UV nudging. Overall, the runs which do not nudge the climatology generally preserve WACCM's free running climatological tropical upwelling mass flux.

Deleted: it seems reasonable to conclude that

Deleted: hypothesized benefit of both climatological anomaly and zonal anomaly nudging - less disruption to the climatology

Deleted: - is generally true for the

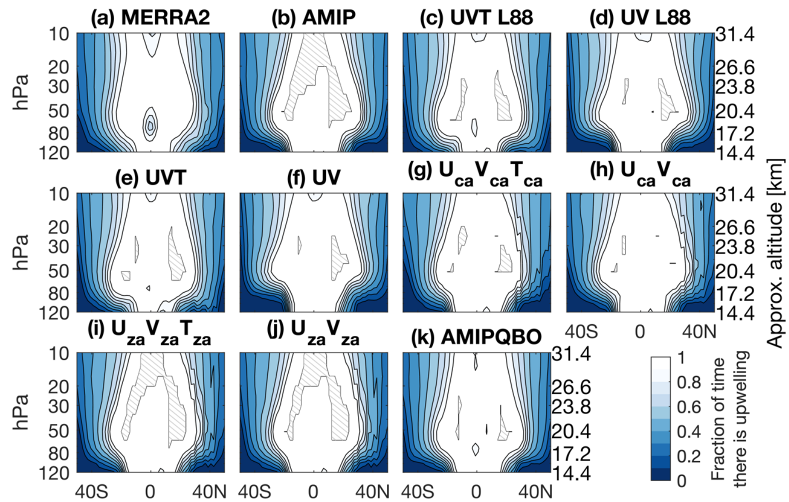

Figure 2: Frequency of upwelling based on the streamfunction definition of the residual circulation for each model simulation. Hatching indicates there is upwelling 100% of the time.

A statistical measure of tropical upwelling in the latitude-pressure plane is the frequency with which upwelling motion is observed (Fig. 2). Here, the TEM residual vertical velocity is calculated using the definition

$$[\omega]^* = [\omega] + \frac{1}{a\cos(\phi)}\left(\frac{[v'\theta']}{[\theta]_p}\cos(\phi)\right)$$

Major differences exist between WACCM and MERRA2. The upwelling frequency in MERRA2 tends to be split in the middle stratosphere, whereas there is consistent upwelling in both AMIP and AMIPQBO, indicating that the deep branch of the circulation in MERRA2 tends to be more consolidated in latitude in the winter seasons (see Fig.'s S4 and S5). While AMIP has constant upwelling (100% of the time) throughout the tropics, AMIPQBO only has scattered regions of constant upwelling in the subtropics. The secondary meridional circulations associated with the QBO are apparently strong enough to drive downwelling 1-5% of the time on the equator. In the standard nudging simulations, the morphology of upwelling appears as the average of the AMIP simulations and MERRA2, with slightly-split upwelling in the middle stratosphere and pockets of constant subtropical upwelling. The middle stratospheric upwelling in both of the anomaly nudging varieties appears more similar to AMIP than to MERRA2, further evidence that these nudging schemes preserve WACCM's climate. Upwelling in the zonal anomaly simulations is constant in the tropics as it is in AMIP, because neither have a QBO. While the local minimum in upwelling in MERRA2 on the equator in the lower stratosphere may be related, in part, to the anomalous circulations associated with the QBO (compare AMIP and AMIPQBO), it is also possibly related to the minimum in ozone leading to a local minimum in diabatic heating [Ming et al. 2016].

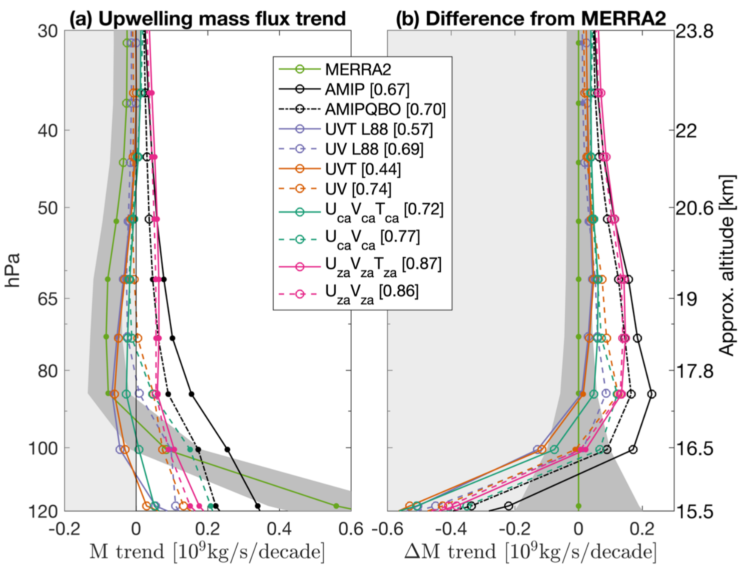

Figure 3: Upwelling mass flux (a) trends and (b) trend differences from MERRA2. Shading indicates the 95% confidence interval on the MERRA2 trends. Numbers in the legend indicate the correlation coefficient squared between each simulation's trends and the trends in MERRA2. Open circles in (a) indicate trends not statistically significant at the 95% confidence level, while open circles in (b) indicate trends statistically significantly different from MERRA2 at the 95% confidence level.

In similar fashion to the upwelling frequency, there are categorical differences between the upwelling trends in all WACCM simulations and MERRA2 over the full 1980-2017 analysis period (Fig. 3). MERRA2 shows a substantial upwelling acceleration in the TTL and upwelling deceleration in the lower and middle stratosphere. Physically, this pattern indicates enhanced tropospheric air mass transport into the TTL and a shift of upward mass transport from the deep to the shallow branch of the residual circulation. The upwelling trends in all WACCM runs tend to be smaller in the TTL and larger aloft

Deleted: be more negative

Deleted: more positive

compared to MERRA2, and there are no simulations that appear structurally different from each other. In fact, the behavior of the upwelling trends in the WACCM simulations can be well-described by a simple linear shift in the trend pattern.

The UVT L88, UVT, and UVT climatological anomaly nudging simulations tend to reproduce the slightly negative-to-zero MERRA2 upwelling trend in the lower and middle stratosphere, while they completely fail to reproduce the acceleration in the TTL. Standard UVT nudging on both MERRA2 and native levels even suggests a decrease in the upwelling mass flux in the TTL, while the UVT climatological anomaly nudging suggests no significant trend. Upwelling trends in the standard UV and UV climatological anomaly nudging simulations is shifted to more positive values, with upwelling trends matching MERRA2 at 100 hPa, but underestimating the trend below and overestimating the trend aloft. There is no distinction between the upwelling trends in the UVT and UV zonal anomaly simulations, with both indicating acceleration at all levels. One objective measure of the disagreement in the morphology (and not sign) of the upwelling trends is the correlation coefficient squared (R^2) between the trends in the nudging simulations and the trends in MERRA2 (Fig. 3). In general, the UV nudging schemes show better agreement, especially for the standard nudging simulations, while the zonal anomaly nudging simulations show the best agreement of all (probably due to the kinked structure in the trend, with a minimum at 85 hPa similar to MERRA2). This all demonstrates that (incidentally) nudging zonal-mean MERRA2 temperatures - the UVT, UVT L66, and $U_{ca}V_{ca}T_{ca}$ 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.

How can we understand the tendency for WACCM to produce such a different upwelling trend structure than MERRA2? In its free-running mode, with or without QBO nudging, WACCM produces an acceleration in upwelling throughout the TTL and the stratosphere (consistent with another AMIP simulation with different initial conditions; not shown), with a slightly stronger acceleration in the TTL. As nudging is introduced, first nudging only the resolved waves ($U_{za}V_{za}T_{za}$ and $U_{za}V_{za}$) and then only the zonal-mean and resolved wave horizontal winds (UV L88, UV, and $U_{ca}V_{ca}$), the trends become more negative (Fig. 4). The more comprehensive nudging, nudging both zonal-mean and resolved wave temperatures and horizontal winds (UVT L88, UVT, and $U_{ca}V_{ca}T_{ca}$), leads to even more negative trends. The only simulations with TTL and lower stratospheric trends unexplained by a regression of all WACCM simulations are the zonal anomaly nudging simulations, which fall outside of the prediction interval. In general, then, nudging schemes are unable to change the structure of the upwelling trends that are inherent to WACCM over this analysis period, only their average value. MERRA2's trend structure exists in a completely different phase space, well outside of the range of behavior predicted by the WACCM simulations. The upwelling trends in the UV nudging schemes tend to be "closest" to the trends in MERRA2, but they cannot be said to reproduce these trends.

Deleted: standard

Deleted:

Deleted:

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Deleted: the

Deleted: with

Deleted: waves

Deleted: a greater

Deleted: the

Deleted: N

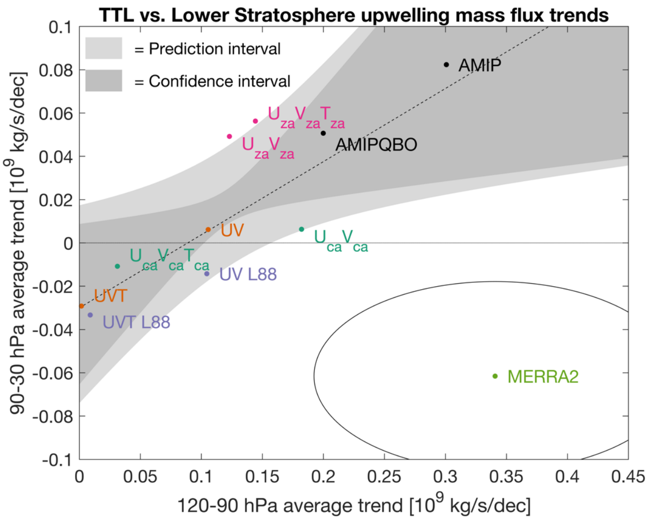

Figure 4: Tropical upwelling trends averaged over the lower stratosphere versus upwelling trends averaged over the TTL. The regression line, 95% confidence interval, and prediction interval are indicated by the dashed line, dark shading, and light shading, respectively. The circle around MERRA2 indicates the 95% confidence interval.

4.1 Variability

To gain some further insight, it is worth examining the time series of upwelling mass flux anomalies at 85 hPa at the top of the TTL. At this level, MERRA2 and the standard UVT nudging simulations indicate deceleration and the other simulations indicate acceleration. Upwelling variability in the standard nudging simulations, which comprehensively nudge the meteorology in both the time (climate and anomaly) and spatial (zonal mean and resolved wave) dimensions, agrees well with the variability in MERRA2 (Fig. 5). Anomalies in the UV nudging simulations tend to be more negative than those in the UVT nudging simulations and MERRA2 early in the record, and more positive later in the record, leading to a more positive trend in the UV nudging simulations at this altitude (Fig. 3).

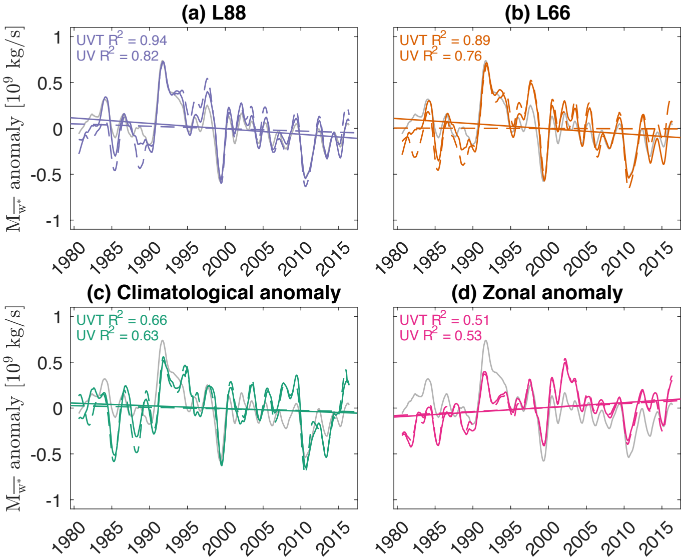

Figure 5: 12-month low-pass filtered time series of tropical upwelling mass flux and its linear trend at 85 hPa in the (a) standard L88 nudging, (b) standard nudging, (c) climatological *anomaly nudging, and (d) zonal anomaly nudging simulations, with UVT and UV nudging indicated by the solid and dashed lines and MERRA2 indicated by the solid grey line. The correlation coefficient squared between each time series and MERRA2 is shown in each panel.*

Climatological anomaly nudging constrains the temporal anomalies in both the zonal mean and the resolved waves. At first glance, then, it is surprising that the anomalies in the upwelling mass flux in the climatological anomaly nudging simulations disagree with those in MERRA2. In particular, while the magnitude of the anomalies seems consistent between the two, the peaks in the climatological anomaly simulations tend to be shifted relative to those in MERRA2. Similarly, while the magnitude of the anomalies in the zonal anomaly nudging simulations is much weaker than in MERRA2, the peaks are more aligned. This suggests that the zonal mean and its climatology, in particular, can substantially dictate the projection of zonal mean and resolved wave variability onto upwelling. This inference is obvious from the momentum balance equation - the vertical and meridional shears in the zonal-mean zonal wind and the static stability kinematically determine the projection of the eddy heating, and dynamically determine wave propagation by modulating potential vorticity gradients. In other words, the same resolved wave variability produces different upwelling for different climates. This raises the question of whether it is even possible to construct a nudging scheme that reproduces both variability and trends from a model with a different climatology.

Deleted: preserves a model’s climate and

Deleted: different

Below 85 hPa, the zonal anomaly nudging exhibits extremely poor correlations with the variability in MERRA2 and the other simulations, suggesting a strong role for the zonal-mean circulation in transforming wave dynamics into zonal-mean momentum forcing and therefore upwelling (Fig. 6). Prescribing either the zonal-mean circulation or its climatological anomaly seems to guarantee better performance, suggesting that the TTL upwelling is governed by interannual variability in the zonal-mean circulation. Above 85 hPa, there is the opposite behavior - zonal anomaly nudging performs slightly worse than the conventional nudging and substantially better than the climatological anomaly nudging. How could it be that prescribing both the zonal-mean climate and its variability, or neither, is better than prescribing only the variability? A plausible hypothesis is that there may be physical incoherencies in the climatological anomaly nudging simulations, arising from the combination of the zonal-mean anomalous circulation in MERRA2 with the zonal-mean climatological circulation in WACCM.

Deleted: primarily

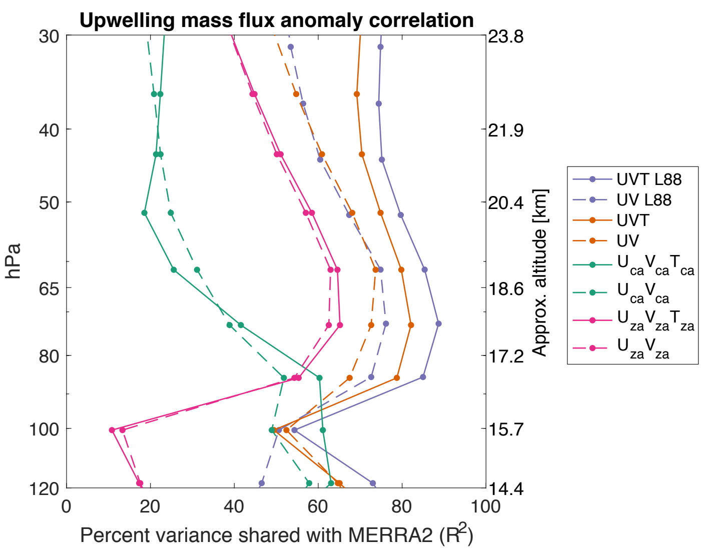

Figure 6: Correlation of tropical upwelling mass flux anomalies between each simulation and MERRA2.

4.2 Physical drivers of upwelling trends

Examining the contribution of specific terms in the momentum balance to the upwelling trends may provide some cursory answers as to why these nudging schemes produce disparate trends and variability. The upwelling trend due to a particular term of interest is given by the linearization about the total trend,

Deleted: in either equation

$$Trend_{term}(p) = Trend_{total}(p) - Trend_{detrendedterm}(p)$$

where $Trend_{total}(p)$ is the total upwelling trend, $Trend_{detrendedterm}(p)$ is the upwelling trend where the term of interest has been detrended, and $Trend_{term}(p)$ is the upwelling trend due to the term of interest. To detrend, we remove the linear trend in the term, including its intersect, and then add the mean of the trended data. This ensures that the term has the same average value before and after detrending.

First, we examine the contributions in the momentum balance equation (Fig. 7). Here, the contribution from $[u]$ is with respect to its time tendency and the meridional angular momentum gradient - it does not consider the role of the shear terms in the Eliassen-Palm flux divergence. The total upwelling trend and its variation among the simulations is dominated by the contribution from resolved wave momentum forcing. However, gravity wave momentum forcing from orographic and non-orographic waves drives a surprisingly large fraction of the upwelling trend in the WACCM simulations - up to 100% of the trend above 70 hPa in the zonal anomaly simulations. In MERRA2, there is no detectable gravity wave contribution to the upwelling trend at any altitude, despite having an overhauled gravity wave scheme that minimizes analysis tendencies in the tropics [Molod et al. 2015, Coy et al. 2016]. This partly explains why the WACCM simulations tend to have a more positive upwelling trend than MERRA2 in the lower stratosphere and a more negative trend in the TTL.

Deleted: the

Deleted: term

Deleted: (consistent with, for example, Fig. 18 of McLandress and Shepherd [2009]; gravity wave momentum forcings dominate the contribution at the edge of the pipe at 70 hPa during boreal winter)

Next, we examine the contributions to the upwelling trend from specific terms in the Eliassen-Palm flux divergence (Fig. 8). There is no contribution from the zonal-mean zonal wind, and while the vertical eddy momentum flux term drives deceleration in the TTL, its spread among the WACCM simulations and MERRA2 is small. There is some spread among the WACCM simulations due to the zonal-mean potential temperature term in the TTL, which curiously implicates different zonal-mean temperature trends. In general, though, it is the meridional eddy heat and momentum fluxes that drive the bulk of the variation among the WACCM simulations and between the WACCM simulations and MERRA2. The meridional eddy momentum flux drives an acceleration in upwelling in all of the WACCM simulations except UVT L88, but at a substantially weaker rate than in MERRA2 in the TTL. However, most of the spread among the WACCM simulations is due to the meridional eddy heat flux term, which drives deceleration in MERRA2 and the UVT nudging simulations and drives acceleration in all of the UV and the UVT zonal anomaly nudging simulations. This suggests that the zonal-mean temperature determines whether the eddy heat flux will drive acceleration or deceleration. It is not a kinematic impact (or else it would manifest in the upwelling trend due to zonal-mean potential temperature), but rather an impact on the projection of the wave physics within the simulations.

Deleted: imposed trend in the

Deleted: onto the zonal-mean climate

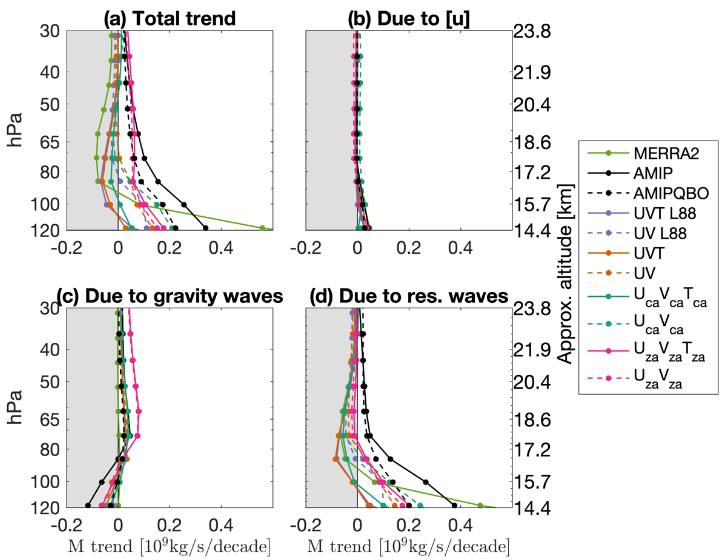

Figure 7: Tropical upwelling mass flux trends: (a) total trend, (b) trend due to zonal-mean zonal wind, (c) trend due to gravity wave momentum forcing, and (d) trend due to resolved wave momentum forcing. See text for details.

To gain a more comprehensive understanding, we examine the trends in the Eliassen-Palm flux and its divergence, as well as changes in the index of refraction (Fig.'s 9-10). The Eliassen-Palm flux is parallel to the group velocity for linear Rossby waves, such that Rossby wave generation, dissipation, and propagation can be assessed from the combined flux and divergence pattern [Edmon et. al 1980, and references therein]. These quantities directly diagnose the wave dynamics, while the index of refraction diagnoses whether the Rossby wave solution to the potential vorticity equation is oscillatory or evanescent. Waves will tend to refract toward positive values of the index, while negative values indicate that waves will evanesce exponentially away from their source region. Here, we take the difference of the quasi-geostrophic form of the index between 1994-2017 and 1980-1993 [Matsuno 1970]. Because of differencing the wavenumber term vanishes and the resulting changes in the index are applicable to all wavenumbers, though we note the upwelling in the TTL is primarily driven by planetary-scale waves [Ortland and Alexander 2014, Kim et al. 2016]. We also plot the latitudes of the streamfunction maxima and minima, indicative of where the mean meridional circulation vertical velocity switches from upward to downward, commonly called the "turn-around latitudes" [Rosenlof 1995], as the total upwelling mass flux within the pipe can be diagnosed solely by the net momentum tendency along this contour.

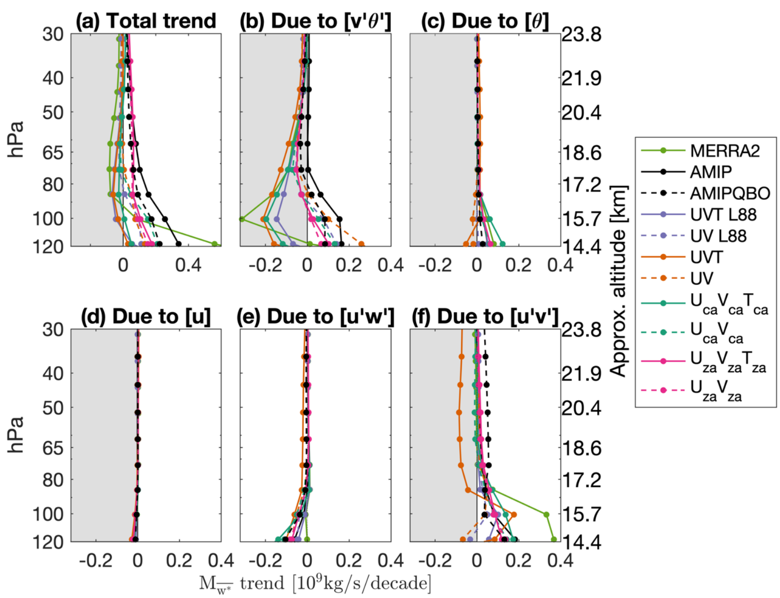

Figure 8: Tropical upwelling mass flux trend: (a) total trend, (b) trend due to the meridional eddy heat flux, (c) trend due to the zonal-mean potential temperature, (d) trend due to the zonal-mean zonal wind, (e) trend due to the vertical eddy momentum flux, and (f) trend due to the meridional eddy momentum flux. See text for details.

The climatological Eliassen-Palm flux divergence in the tropics is characterized by an equatorial minimum in negative wave forcing that increases poleward and, in the TTL, downward. In MERRA2 there is a downward consolidation and equatorward shift of the wave forcing at the edge of the pipe in the subtropics from 1980 to 2017, which drives the strong increase in upwelling in the TTL and the minor decrease in upwelling aloft (Fig. 3). Here, the increasing wave drag in the TTL originates from the subtropical lower stratosphere and from within the upwelling region. Contrast this pattern with the pattern observed in the AMIP simulation: enhanced wave propagation from the extratropical troposphere into the subtropical TTL and lower stratosphere, producing acceleration over a deep layer. This pattern resembles the canonical greenhouse gas response observed in CMIP-type simulations [Garcia and Randel 2008, Shepherd and McLandress 2011], and is an enhancement and

Deleted: two

Deleted: s

upward shift of the climatological wave forcing. The AMIPQBO simulation exhibits this pattern only in the Southern Hemisphere.

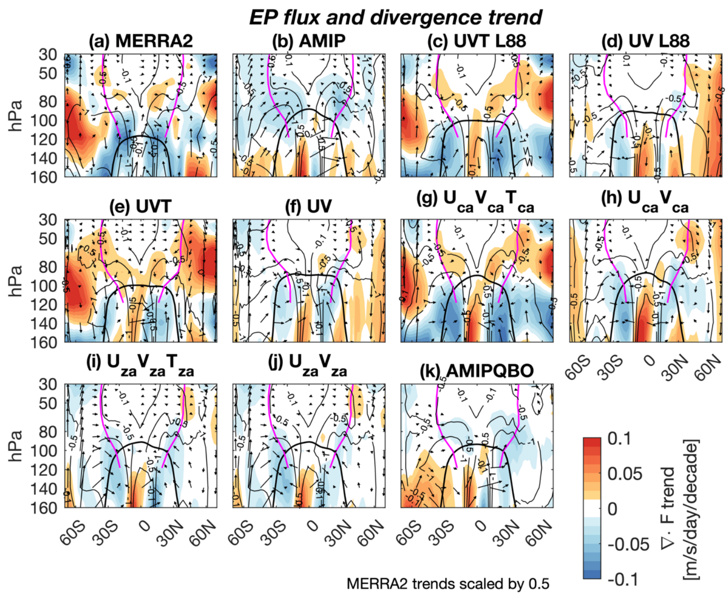

Figure 9: Eliassen-Palm flux (vectors) and divergence (shading) trends, climatological Eliassen-Palm flux divergence (black contours, -0.1, -0.5, and -1.0 m/s/day), and climatological turn-around latitudes for all simulations (magenta). MERRA2 Eliassen-Palm flux and divergence scaled by 0.5. Eliassen-Palm flux vectors have been scaled to appear consistent with the divergence field, to an arbitrary maximum scale consistent across panels [Edmon et al. 1980]. Tropopause indicated by the thick black line.

The specified dynamics simulations have patterns that are so similar that they can simply be collapsed into two composites - the UVT L88, UVT, and UVT climatological anomaly simulations, which we will refer to as the "zonal-mean temperature nudging" composite, and the UV L88, UV, UV climatological anomaly, and both zonal anomaly simulations, which we will refer to as the "no zonal-mean temperature nudging" composite (Fig. 10). We emphasize these composites are not merely based on the commonality of this pattern, but also on the binary distinction of whether they nudge zonal-mean temperatures.

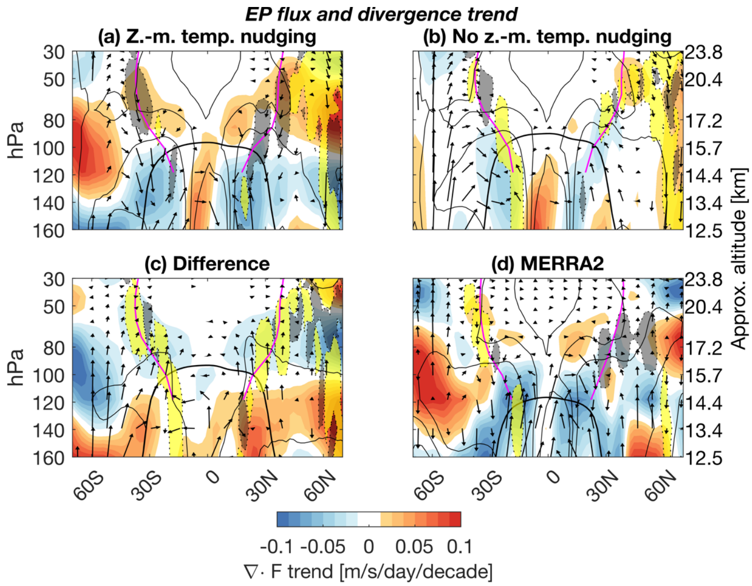

Figure 10: As in Figure 9, but for the composites of simulations with (a) and without (b) zonal-mean temperature nudging, (c) the difference between the composites, and (d) MERRA2. Also shown are positive (yellow translucence) and negative (black translucence) changes in the index of refraction greater than $1x10^{-11}$ m^{-2}.

In the "zonal-mean temperature nudging" simulations, the change in wave momentum forcing resembles the pattern in MERRA2, but the consolidation of wave forcing from within the pipe to the subtropical lower stratosphere amplifies the climatological drag, rather than shifting it equatorward as in MERRA2. In the Southern Hemisphere in MERRA2, the downward consolidation and equator shift of drag is associated with increases in the index of refraction along the pipe edge between 160 and 100 hPa, while in the "zonal-mean temperature nudging" simulations, there is a decrease in the index in the same location. This suggests a possible mechanism whereby the erroneous decrease in the index of refraction in the "zonal-mean temperature nudging" simulations may be refracting the waves away from the pipe and preventing them from accelerating the upwelling in the TTL. Contrast this with the "no zonal-mean temperature nudging" simulations, in which the positive wave momentum forcing trend in the TTL is due to enhanced propagation from the extratropical troposphere into the lower stratosphere and TTL, similar to the AMIP simulations. However, the waves arc over the subtropics and refract toward the increasing index of refraction, unlike the AMIP simulations and similar to the downward consolidation of wave momentum forcing seen in MERRA2 and the "zonal-mean temperature nudging" simulations. The differenced flux and divergence trends between the "no zonal-mean temperature nudging" and "zonal-mean temperature nudging" simulations resemble the changes seen in the AMIP simulations. Therefore, 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. There is more agreement in the wave momentum forcing changes among the composites and MERRA2 in the Northern Hemisphere, at least at the turn around latitude where it has consequences for tropical upwelling.

5 Conclusions and Discussion

Our primary conclusions can be summarized as follows:

1. In WACCM, the particular specified dynamics methodology can have a substantial impact on the climate, trends, and variability of stratospheric upwelling
    a. While specified dynamics schemes that (incidentally) nudge zonal-mean temperatures reproduce variability in tropical stratospheric upwelling, they do not reproduce stratospheric upwelling trends
    b. Specified dynamics schemes that do not nudge zonal-mean temperatures tend to better (but not entirely) reproduce stratospheric upwelling trends, at the expense of variability
2. Nudging the zonally-anomalous circulation tends to most consistently preserve WACCM’s climate and reproduce stratospheric upwelling trends in the TTL
3. Gravity wave parameterizations can interfere with a specified dynamics scheme’s ability to reproduce upwelling trends

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.

From these conclusions, we can infer that prescribing zonal-mean MERRA2 temperatures, and in particular the climatological anomalies in zonal-mean temperatures, restricts the temperature response that otherwise spontaneously occurs in the AMIP simulations. However, prescribing the zonal-mean temperatures does not lead to a better reproduction of tropical upwelling trends or wave momentum forcing trends along the pipe edges present in the input meteorology. While we cannot begin to fully answer why this is the case, we hypothesize that it may be due to a mismatch in the climatologies of WACCM and MERRA2.

Deleted: The

Deleted: points

Deleted: in

Deleted: also

Deleted: \

The "no zonal-mean temperature nudging" simulations, the AMIP simulations, and MERRA2 exhibit similar temperature trends, but the trends in the "zonal-mean temperature nudging" simulations are wildly different, with warming everywhere in the lower stratosphere (Fig. 11). Consider the difference in the climatological temperature between MERRA2 and the AMIP simulation. AMIP is 3 Kelvin colder in the tropical TTL and lower stratosphere, likely due to stronger adiabatic cooling from the greater upwelling frequency and stronger upwelling mass flux (Fig.'s 1, 2). If the zonal mean temperature is nudged, this will induce a warming tendency in the TTL and lower stratosphere (and a cooling tendency in the upper troposphere; see Weaver et al. [1993] and also Miyazaki et al. [2005] - temperature nudging essentially induces spurious heating). Such a warming tendency would induce a response akin to the westerly-shear QBO phase, with anomalous subsidence on the equator and rising motion in the subtropics (see for example Plumb and Bell [1982], Fig. 1). Indeed, the models in which the zonal-mean temperature is nudged tend to have a reduced upwelling frequency in the equatorial lower stratosphere and TTL (Fig. 2), though the distinctions are less obvious when one considers the upwelling mass flux (Fig. 1).

Deleted:)

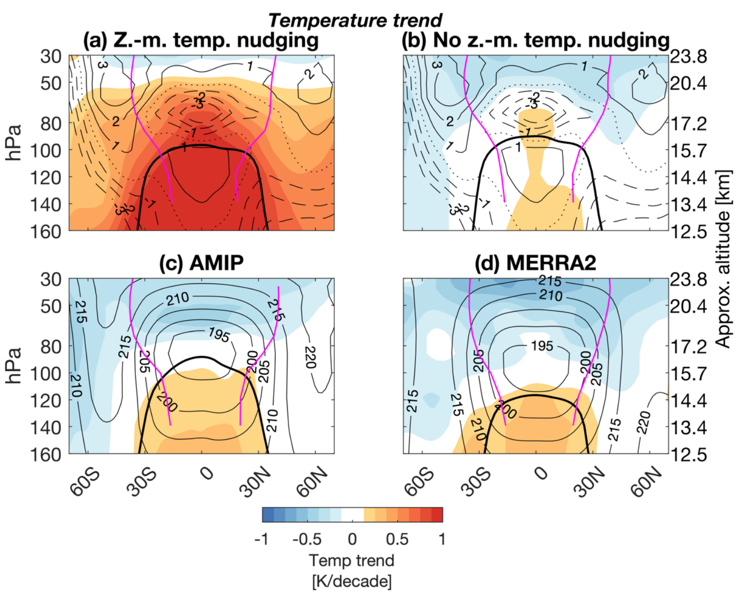

Figure 11: Zonal-mean temperature trends (shading) in the simulations with (a) and without (b) zonal-mean temperature nudging, (c) the AMIP simulation, and (d) MERRA2. Climatological temperatures shown for the AMIP simulation and MERRA2 (black contours, every 5 K), while the climatological temperature difference between the AMIP simulation and

MERRA2 is shown on the upper panels (black contours, every 1 K). Also shown are the climatological turn-around latitudes (magenta).

These meridional circulation responses to the nudging tendencies could, in principle, influence trends in potential vorticity which impact wave propagation and ultimately the tropical upwelling mass flux trend (Fig. 10), but a precise mechanism is not obvious. Another possible complication is that differences in the native horizontal resolution between the nudged and meteorological models could lead to non-negligible differences in the wave physics [Boville 1991, Held and Phillipps 1993, Bèguin et al. 2013, Davis and Birner 2016]. While the schemes nudge the eddy terms, the model still has the freedom to determine the final wave generation and breaking processes that lead to zonal-mean wave momentum forcings. Both of these possibilities present potentially serious challenges for specified dynamics schemes. By construction the schemes produce zonal-mean heat and momentum forcings that will engage feedbacks within the model that act to preserve its climatology. Ostensibly, it is not possible in WACCM to nudge zonal-mean temperatures without inducing compensating changes in the model that cause even its zonal-mean temperature trends to diverge rapidly from the trends in the input data (Fig. 11). We should not be surprised, then, that such simulations fail to reproduce tropical upwelling trends, which are driven by nuanced changes in wave propagation and dissipation (Fig.'s 9 and 10).

Parameterized gravity waves also lead to errors between the nudged and meteorological reference simulations [Smith et al. 2017]. Here, they drive deceleration in the TTL and acceleration in the lower stratosphere in WACCM, in part explaining why WACCM tends to have a more positive (negative) upwelling trend in the lower stratosphere (TTL) compared to MERRA2. As long as the parameterization is active, it presents an upper limit on performance in simulating the MERRA2 trends.

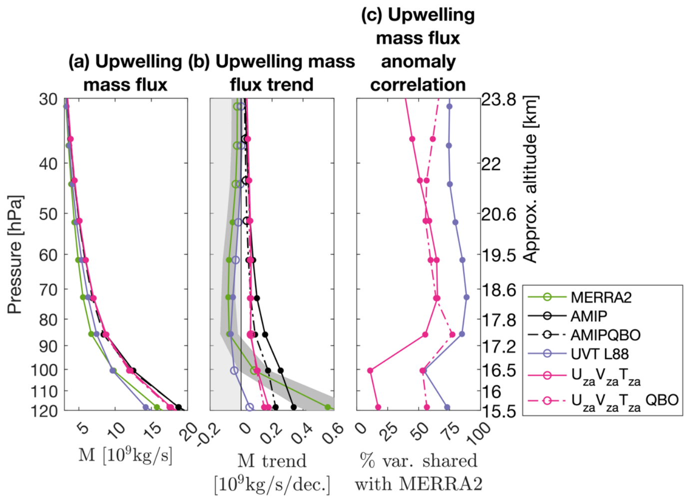

Figure 12: Climatological (a) upwelling mass flux, (b) upwelling mass flux trend, and (c) anomaly correlation with MERRA2. Shading and symbols in (a) and (b) as in Figures 1 and 3, respectively. $U_{za}V_{za}T_{za}$ QBO is a $U_{za}V_{za}T_{za}$ simulation with zonal-mean zonal wind (QBO) nudging in the tropical stratosphere.

We have shown that by *not* nudging the zonal-mean temperatures or, even better, *not* nudging zonal-mean variables at all, the simulations are able to respectably reproduce tropical upwelling trends without damaging WACCM's climatology. As the conversion of resolved waves to zonal momentum forcings depends on the zonal mean circulation, the zonal anomaly nudging scheme may be further improved by the inclusion of a separate QBO nudging scheme. Here the combination of QBO nudging (of the zonal-mean zonal winds in the tropics) with zonal anomaly nudging does not impact the climatology or the trends (Fig. 12), but drastically improves the anomaly correlation between WACCM and MERRA2 in the TTL and middle stratosphere to the degree that there is little gain in skill in these regions using the standard UVT L88 scheme. This additional QBO nudging ameliorates the primary drawback of the pure zonal anomaly nudging scheme – its lack of QBO-driven circulation variability. Nevertheless, over this sufficiently-long period, the upwelling trend in the UVT zonal anomaly QBO simulation is almost identical to that in the AMIPQBO simulation, suggesting no gain in skill by constraining historical variability.

In certain situations, nudging to zonal-mean temperatures might be necessary - for example, when there is interest in temperature-dependent chemistry [Froidevaux et al. 2019]. In these cases, the disadvantages of temperature nudging should be investigated and weighed against its advantages. Regardless of the application, model, or input meteorology, care should be taken in interpreting specified dynamics simulations, especially in regard to their modeled trends. Trends are often the small residuals of the balance of large terms; they may be overshadowed by the nudging acting on the climatological differences between the models.

Data availability

MERRA2 is provided by NASA's Global Modeling and Assimilation Office at https://gmao.gsfc.nasa.gov/reanalysis/. Source code modifications are archived at doi:10.5281/zenodo.3376232 and can also be provided on request by nadavis@ucar.edu – please note they are compatible with CESM 1.2.2, but may not be compatible with all versions of CESM. Raw and post-processed model output is also available upon request.

Author contributions

N. A. Davis, S. M. Davis, R. W. Portmann, and P. Yu designed the model experiments, and N. A. Davis and P. Yu performed the experiments. N. A. Davis produced source code modifications, analyzed model output, and wrote the manuscript with editing and support from all authors.

Acknowledgements

Three anonymous reviewers are thanked for their efforts to provide comments and improve this manuscript. We thank the Global Modeling and Assimilation Office (GMAO) at the NASA Goddard Space Flight Center for producing and distributing the MERRA2 reanalysis, and thank Peter Hitchcock for constructive comments and suggestions.

References

Abalos, M., Randel, W. J., Kinnison, D. E., and Serrano, E.: Quantifying tracer transport in the tropical lower stratosphere using WACCM, *Atmos. Chem. Phys.*, 13, 10591-10607, doi:10.5194/acp-13-10591-2013, 2013.

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, 7534-7554, doi:10.1002/2015JD023182, 2015.

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, 1361-1380, doi:10.1002/2015JD023334, 2016.

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, 1379-1394, https://doi.org/10.5194/acp-18-1379-2018, 2018.

Bèguin, A., Martius, O., Sprenger, M., Spichtinger, P., Folini, D., and Wernli, H.: 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, 2013.

Boville, B. A.: 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, 1991.

Butchart, N.: The Brewer-Dobson circulation, *Rev. Geophys.*, 52, doi:10.1002/2013RG000448, 2014.

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, 2331–2340, https://doi.org/10.1175/2010JAS3433.1, 2010.

Chandran, A., Garcia, R. R., Collins, R. L., and Chang, L. C.: Secondary planetary waves in the middle and upper atmosphere following the stratospheric sudden warming event of January 2012, *Geophys. Res. Lett.*, 40, 1861– 1867, doi:10.1002/grl.50373, 2013.

Chemke, R. and Polvani, L. M.: Opposite circulation trends in climate models and in reanalyses, *Nature Geosci.*, 12, 528-532, https://doi.org/10.1038/s41561-019-0383-x, 2019.

Chipperfield, M. P.: Multiannual simulations with a three-dimensional chemical transport model, *J. Geophys. Res.: Atmos.*, 104, 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, 1179-1203, doi:10.1256/qj.05.51, 2006.

Chipperfield, M. P., Dhomse, S., Hossaini, R., Feng, W., Santee, M. L., Weber, M., J. P. Burrows, J. D. Wild, D. Loyola, and M. Coldewey-Egbers: On the cause of recent variations in lower stratospheric ozone, *Geophys. Res. Lett.*, 45, 5718–5726, https://doi.org/10.1029/2018GL078071, 2018.

Chrysanthou, A., Maycock, A. C., Chipperfield, M. P., Dhomse, S., Garny, H., Kinnison, D., Akiyoshi, H., Deushi, M., Garcia, R. R., Jöckel, P., Kirner, O., Morgenstern, O., Pitari, G., Plummer, D. A., Revell, L., Rozanov, E., Stenke, A., Tanaka, T. Y., Visioni, D., Yamashita, Y., and Zeng, G.: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry-climate models, *Atmos. Chem. Phys. Discuss.*, https://doi.org/10.5194/acp-2019-260, in review, 2019.

Coy, L., Wargan, K., Molod, A. M., McCarty, W. R., and Pawson, S.: Structure and dynamics of the Quasi-Biennial Oscillation in MERRA-2, *J. Climate,* 29, 5339–5354, https://doi.org/10.1175/JCLI-D-15-0809.1, 2016.

Davis, N. and Birner, T.: 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, 2016.

Diallo, M., Riese, M., Birner, T., Konopka, P., Müller, R., Hegglin, M. I., Santee, M. L., Baldwin, M., Legras, B., and Ploeger, F.: Response of stratospheric water vapor and ozone to the unusual timing of El Niño and the QBO disruption in 2015–2016, *Atmos. Chem. Phys.*, 18, 13055-13073, https://doi.org/10.5194/acp-18-13055-2018, 2018.

Edmon, H. J., Hoskins, B. J., and McIntyre, M. E.: Eliassen-Palm Cross Sections for the Troposphere, *J. Atmos. Sci.*, 37, 2600–2616, https://doi.org/10.1175/1520-0469(1980)037<2600:EPCSFT>2.0.CO;2, 1980.

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

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W. (2009), Tropical tropopause layer, *Rev. Geophys.*, 47, RG1004, doi:10.1029/2008RG000267.

Garcia, R.R. and Randel, W.J.: Acceleration of the Brewer–Dobson Circulation due to Increases in Greenhouse Gases. *J. Atmos. Sci.*, 65, 2731–2739, https://doi.org/10.1175/2008JAS2712.1, 2008.

Garcia, R. R. and Richter, J. H.: On the Momentum Budget of the Quasi-Biennial Oscillation in the Whole Atmosphere Community Climate Model. *J. Atmos. Sci.,* 76, 69–87, https://doi.org/10.1175/JAS-D-18-0088.1, 2019.

Gates, W. L.: An AMS Continuing Series: Global Change - AMIP: The Atmospheric Model Intercomparison Project, *Bull. Amer. Meteor. Soc.*, 73, 1962–1970, https://doi.org/10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2, 1992.

Gelaro, R., McCarty, W., Suárez M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, R., Nielsen, J. E., Partyka, G. Pawson, S, Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *J. Climate*, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2011.

Hardiman, S. C., Andrews, D. G., White, A. A., Butchart, N., and Edmond, I.: Using different formulations of the transformed Eulerian mean equations and Eliassen–Palm diagnostics in general circulation models, *J. Atmos. Sci.*, 67, 1983–1995, https://doi.org/10.1175/2010JAS3355.1, 2010.

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, 1209-1232, doi:10.5194/gmd-10-1209-2017, 2017.

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, 651–678, https://doi.org/10.1175/1520-0469(1991)048<0651:OTCOED>2.0.CO;2, 1991.

Held, I. M. and Phillipps, P. J.: 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, 1993.

Kobayashi, C. and Iwasaki, T.: Brewer-Dobson circulation diagnosed from JRA-55, *J. Geophys. Res*., 121, 1493-1510, doi:10.1002/2015JD023476, 2016.

Kim, J., Randel, W. J., Birner, T., and Abalos, M.: Spectrum of wave forcing associated with the annual cycle of upwelling at the tropical tropopause, *J. Atmos. Sci*., 73, 855–868, https://doi.org/10.1175/JAS-D-15-0096.1, 2016.

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R.,Sassi, F., Boville, B. A., Marsh, D., Harvey, L., Randall,C., Randel, W., Lamarque, J.-F., Emmons, L. K., Hess, P.,Orlando, J., Tyndall, J., and Pan, L.: Sensitivity of chem-ical tracers to meteorological parameters in the MOZART-3 chemical transport model, *J. Geophys. Res.*, 112, D20302, https://doi.org/10.1029/2006JD007879, 2007.

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 Model, *Geosci. Model Dev.*, 5, 369-411, https://doi.org/10.5194/gmd-5-369-2012, 2012.

Lin, S.: A “vertically Lagrangian” finite-volume dynamical core for global models, *Mon. Wea. Rev.*, 132, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2, 2004.

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, 663, https://doi.org/10.1038/ngeo3013, 2017.

Löffler, M., Brinkop, S., and Jöckel, P.: Impact of major volcanic eruptions on stratospheric water vapour, *Atmos. Chem. Phys.*, 16, 6547-6562, https://doi.org/10.5194/acp-16-6547-2016, 2016.

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., and Wright, C. J.: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), *Atmos. Chem. Phys.*, 17, 14593-14629, https://doi.org/10.5194/acp-17-14593-2017, 2017.

Mahieu, E., Chipperfield, M., Notholt, J., Reddmann, 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., Nakajimi, H., Palm, M., Paton-Walsh, c., Russell, J. M., III, Schneider, M., Servais, C., Smale, D., and Walker, K. A.: Recent Northern Hemisphere stratospheric HCl increase due to atmospheric circulation changes, *Nature*, 515, 104–107, doi:10.1038/nature13857, 2014.

Marsh, D. R.: Chemical-dynamical coupling in the Mesosphere and Lower Thermosphere, in Aeronomy of the Earth's Atmosphere and Ionosphere, IAGA Special Sopron Book Ser., vol. 2, edited by M. Abdu, D. Pancheva, A. Bhattacharyya, 1st ed., 370 pp., ISBN: 978-94-007-0325-4, Springer, Dordrecht, 2011.

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. Climate*, 26, 7372–7391, https://doi.org/10.1175/JCLI-D-12-00558.1, 2013.

Matsuno, T.: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere, *J. Atmos. Sci.*, 27, 871–883, https://doi.org/10.1175/1520-0469(1970)027<0871:VPOSPW>2.0.CO;2, 1970.

McLandress, C. and Shepherd, T. G.: Simulated anthropogenic changes in the Brewer–Dobson circulation, including its extension to high latitudes, *J. Climate,* 22,1516–1540, https://doi.org/10.1175/2008JCLI2679.1, 2009.

Meijer, E. W., Bregman, B., Segers, A., and van Velthoven P. F. J.: The influence of data assimilation on the age of air calculated with a global chemistry-transport model using ECMWF wind fields, *Geophys. Res. Lett.*, 31, L23114, doi:10.1029/2004GL021158, 2004.

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, 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, 1889-1901, https://doi.org/10.1175/JAS-D-15-0293.1, 2016.

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, 1339-1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

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, 909-918, doi:10.2151/jmsj.83.909, 2005.

Mote, P. W., Rosenlof, K. H., McIntyre, M. E., Carr, E. S., Gille, J. C., Holton, J. R., Kinnersley, J. S., Pumphrey, H. C., Russell, J. M., and Waters, J. W.: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, *J. Geophys. Res.*, 101, 3989– 4006, doi:10.1029/95JD03422, 1996.

Neale, R. B., Richter, J., Park, S., Lauritzen, P. H., Vavrus, S. J., Rasch, P. J., and Zhang, M.: The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments, *J. Climate*, 26, 5150–5168, https://doi.org/10.1175/JCLI-D-12-00236.1, 2013.

Orbe, C., Waugh, D. W., Yang, H., Lamarque, J.-F., Tilmes, S., and Kinnison, D. E.: Tropospheric transport differences between mod-els using the same large-scale meteorological fields, *Geophys. Res. Lett.*, 44, 1068–1078,doi:10.1002/2016GL071339, 2017.

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, 7217-7235, https://doi.org/10.5194/acp-18-7217-2018, 2018.

Ortland, D. A. and Alexander, M. J.: The residual-mean circulation in the tropical tropopause layer driven by tropical waves, *J. Atmos. Sci.,* 71, 1305–1322, https://doi.org/10.1175/JAS-D-13-0100.1, 2014.

Plumb, R. A. and Bell, R. C.: A model of the quasi-biennial oscillation on an equatorial beta-plane, *Q. J. R. Meteorol. Soc.*, 108, 335-352, doi:10.1002/qj.49710845604, 1982.

Randel, W. J., Garcia, R. R., and Wu, F.: Time-dependent upwelling in the tropical lower stratosphere estimated from the zonal-mean momentum budget, *J. Atmos. Sci.*, 59, 2141–2152, https://doi.org/10.1175/1520-0469(2002)059<2141:TDUITT>2.0.CO;2, 2002.

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, L15822, doi:10.1029/2009GL039343, 2009.

Rosenlof, K. H.: Seasonal cycle of the residual mean meridional circulation in the stratosphere, *J. Geophys. Res.*, 100, 5173–5191, doi:10.1029/94JD03122, 1995.

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, 12,491-12,508. https://doi-org.cuucar.idm.oclc.org/10.1029/2018JD028776, 2018.

Shepherd, T. G. and McLandress, C.: A Robust Mechanism for Strengthening of the Brewer–Dobson Circulation in Response to Climate Change: Critical-Layer Control of Subtropical Wave Breaking. *J. Atmos. Sci.,* 68, 784–797, https://doi.org/10.1175/2010JAS3608.1, 2011.

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, 933–947, https://doi.org/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. Atmos.*, 120, 7958-7974. doi: 10.1002/2015JD023365, 2015.

Solomon, S., Kinnison, D., Garcia, R. R., Bandoro, J., Mills, M., Wilka, C., Neely, R. R., III, 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, 12,624-12,633, doi:10.1002/2016GL071778, 2016.

Stockwell, D. Z. and Chipperfield, M. P.: A tropospheric chemical-transport model: Development and validation of the model transport schemes, *Q. J. R. Meteorol. Soc*., 125, 1747-1783, doi:10.1002/qj.49712555714, 1999.

Tegtmeier, S., Hegglin, M.I., Anderson, J., Funke, B., Gille, J., Jones, A., Smith, L., von Clarmann, T. and Walker, K.A.: The SPARC Data Initiative: comparisons of CFC-11, CFC-12, HF and SF6 climatologies from international satellite limb sounders, *Earth Sys. Sci. Data*, 8, 61-78, 10.5194/essd-8-61-2016, 2016.

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, 81-93, https://doi.org/10.5194/acp-4-81-2004, 2004.

Verronen, P. T., Andersson, M. E., Marsh, D. R., Kovács, T., and Plane, J. M. C.: WACCM-D - Whole Atmosphere Community Climate Model with D-region ion chemistry, *J. Adv. Model. Earth Syst.*, 8, 954-975, doi:10.1002/2015MS000592, 2016.

Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L., and Knowland, K. E.: Recent decline in extratropical lower stratospheric ozone attributed to circulation changes, *Geophys. Res. Lett.*, 45, 5166– 5176. https://doi.org/10.1029/2018GL077406, 2018.

World Meteorological Organization (WMO), Executive Summary: Scientific Assessment of Ozone Depletion: 2018, World Meteorological Organization, Global Ozone Research and Monitoring Project - Report No. 58, 67 pp., Geneva, Switzerland, 2018.

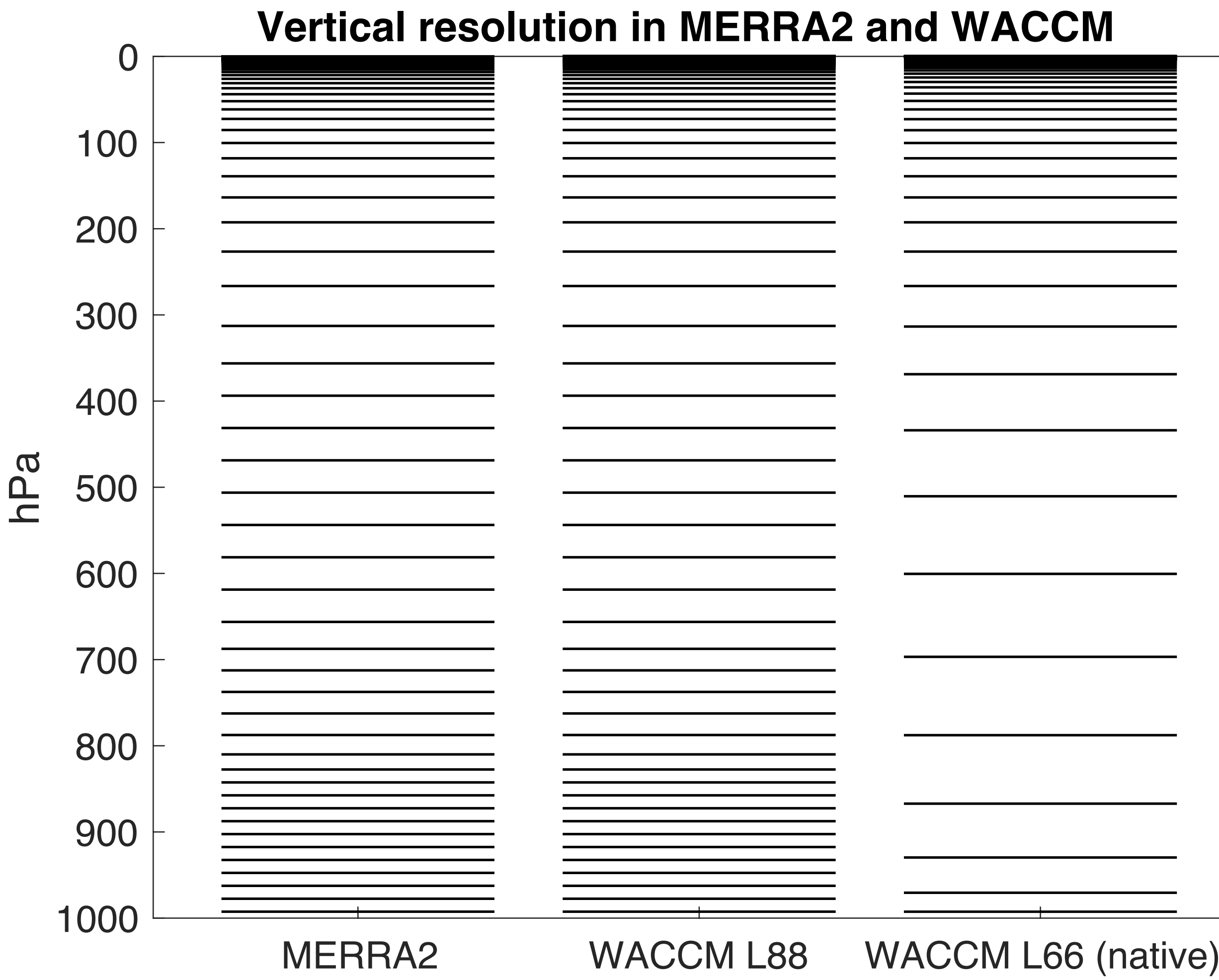

Figure S1: Vertical levels in MERRA2, WACCM on MERRA2 levels with 16 free-running levels above MERRA2's lid, and WACCM on native levels, assuming a 1000 hPa surface pressure.

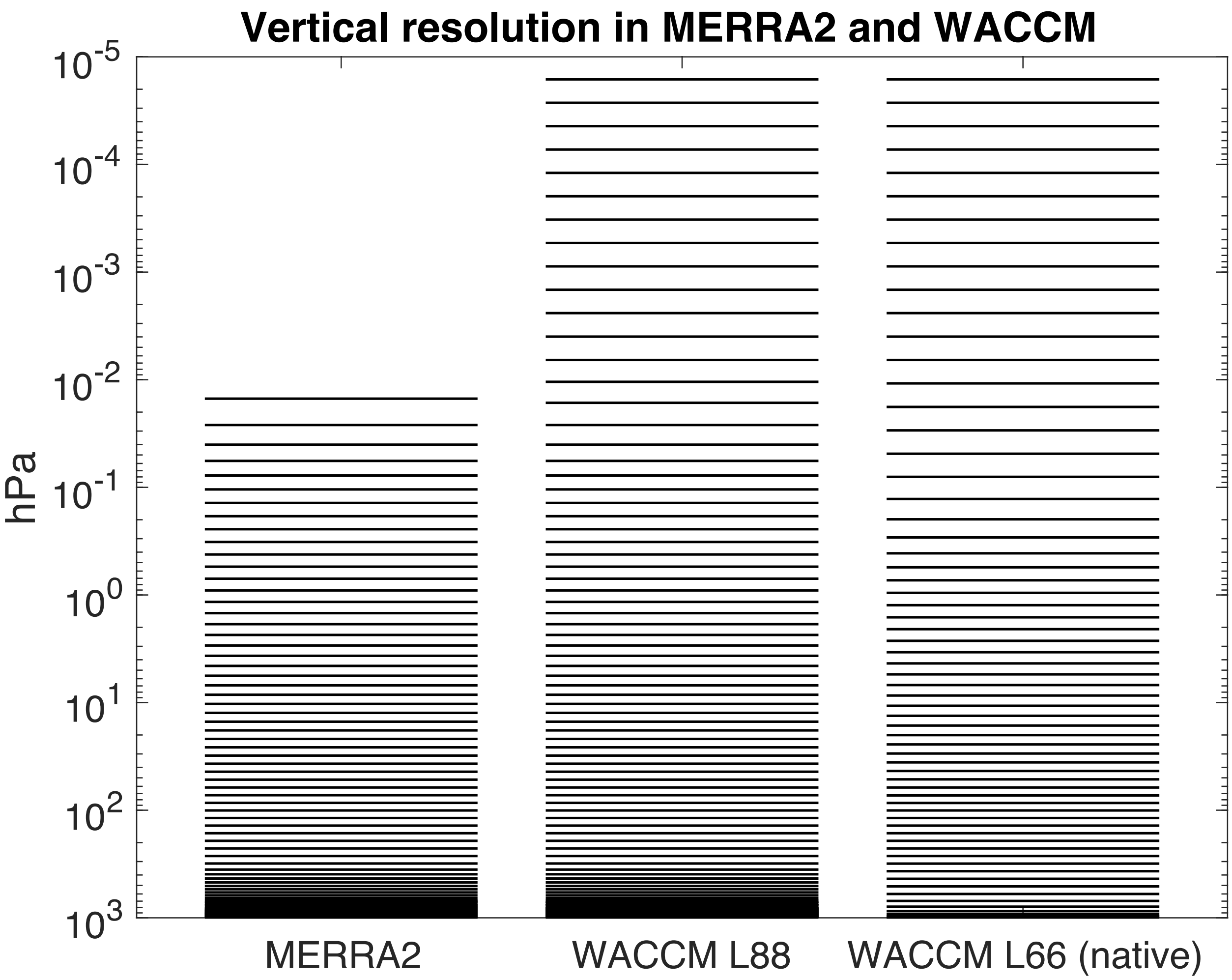

Figure S2: As in Fig. S1, but on a logarithmic pressure grid.

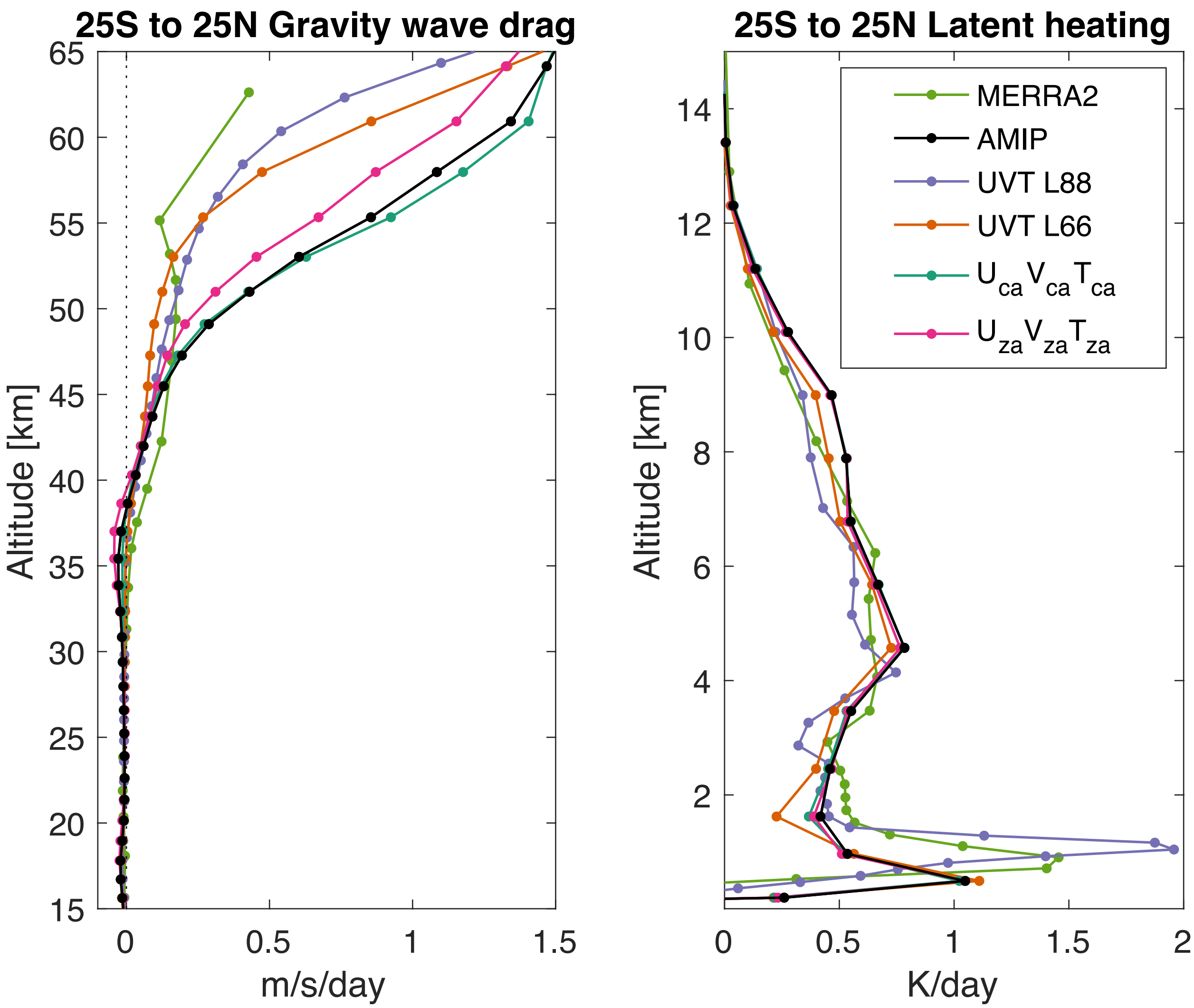

Figure S3: Climatological (left) gravity wave drag and (right) latent heating averaged between 25S and 25N for selected runs. Behavior of UV nudging simulations is quantitatively similar to their UVT counterparts. Gravity wave drag is the sum of the output variables UTGWSPEC, BUTGWSPEC, and UTGWORO.

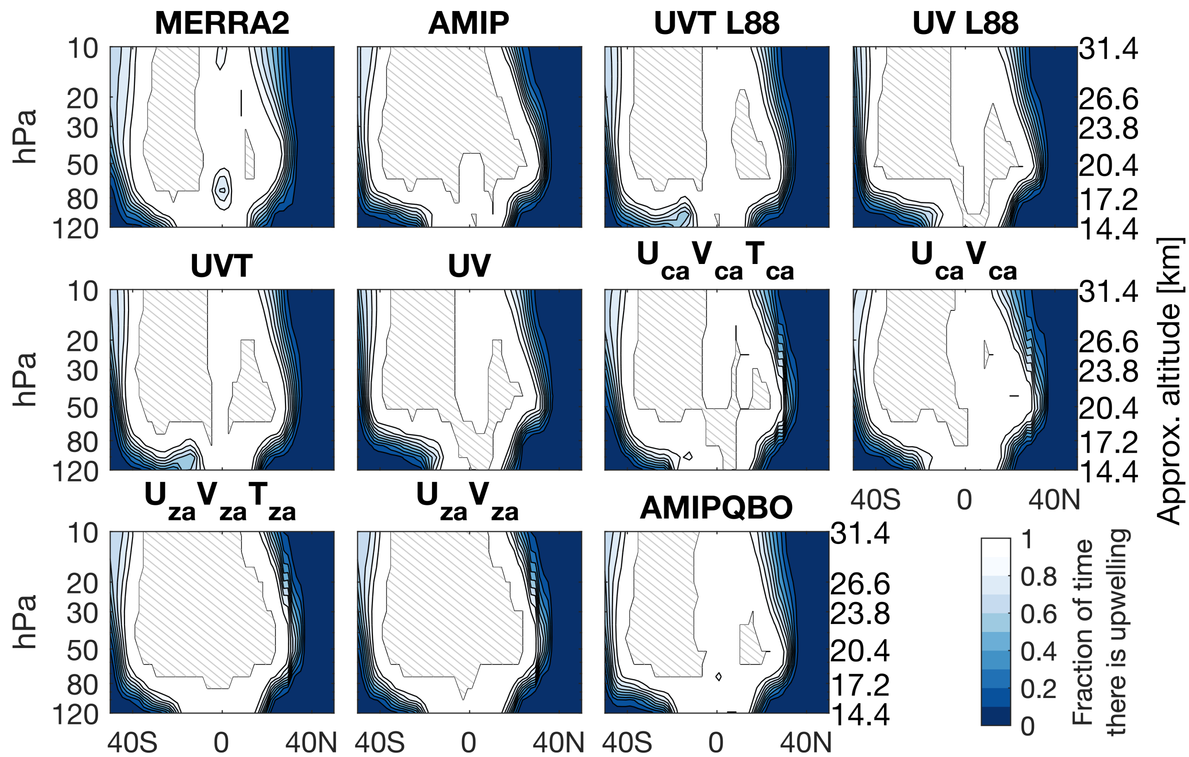

Figure S4: Frequency of upwelling based on the streamfunction definition of the residual circulation for each model simulation for December-January-February only. Hatching indicates there is upwelling 100% of the time.

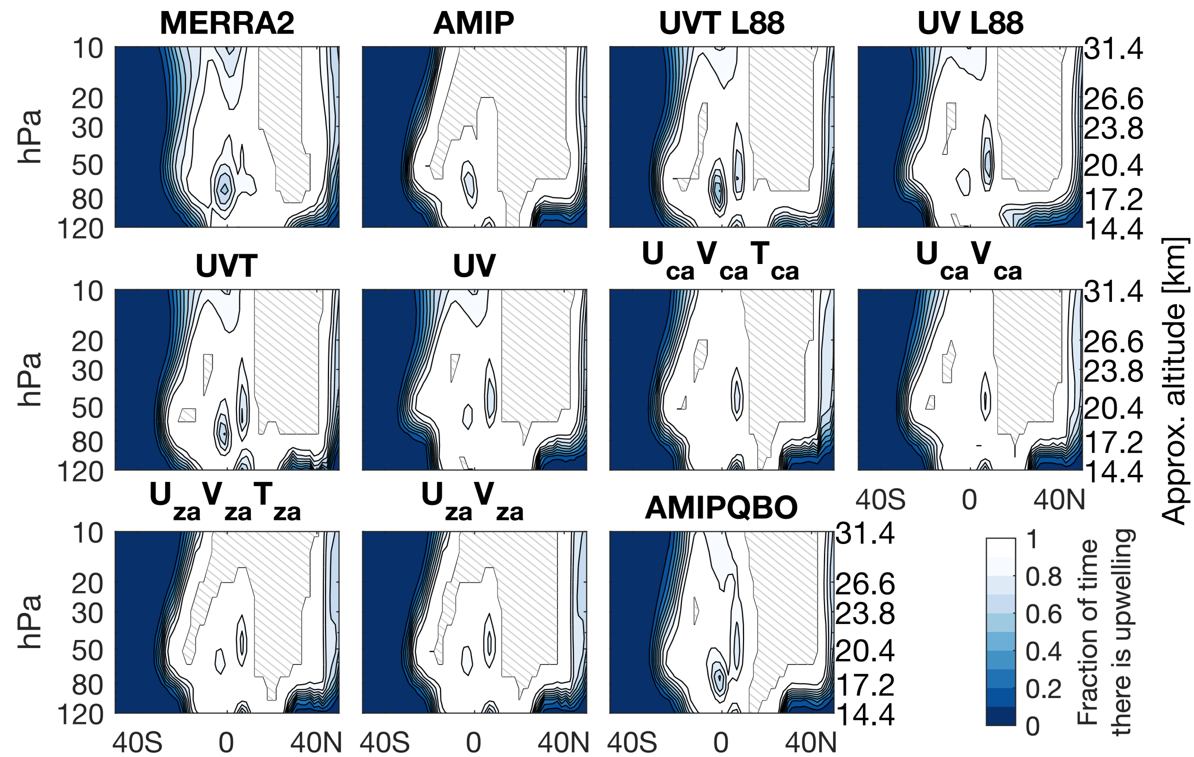

Figure S5: As in Fig. S4, but for June-July-August only.