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.