Dr. Clara Orbe
NASA Goddard Institute for Space Studies Code 611 New York · NY 10025
clara.orbe@nasa.gov

Dr. Paul Young
Editor,
Atmospheric Chemistry and Physics

December 6, 2019
re: manuscript number: acp-2019-625 Title: "Description and Evaluation of the Specified-Dynamics Experiment in the Chemistry-Climate Model Initiative (CCMI)"

Dear Dr. Young,

We thank the referees for their reviews of our manuscript. The generally positive tone of the reviews is encouraging. After careful consideration of all reviewers' suggestions we feel that the changes that we have made to the text have improved the manuscript. We have considered all comments carefully and modified our manuscript accordingly. We have provided two versions of the revised manuscript, one of which includes the corrections highlighted in red. The point-by-point responses to the referee's comments are also included. We hope that the manuscript is now acceptable for publication in ACP. I confirm that my coauthors, David Plummer, Darryn W. Waugh, Huang Yang, Patrick Jöckel, Douglas E. Kinnison, Beatrice Josse, Virginie Marecal, Makoto Deushi, Nathan Abraham, Alexander Archibald, Martyn Chipperfied, Sandip Dhomse, Wuhu Feng and Slimane Bekki concur with the submission of our manuscript in its revised form. The revised version of the manuscript has been resubmitted electronically.

Yours sincerely,

Dr. Clara Orbe

Response to Reviewer 1

We thank the referee for his/her careful and thoughtful review of the manuscript. Incorporation of this feedback into the new revised manuscript has helped improve the clarity and readability of the study.

Our response to the general comments raised are as follows:

1. *"Trends: There is no mentioning of long-term trends in the paper. It is found that interannual variability in the SD simulations is rather consistent, and thus it is concluded correctly that the SD simulations are most justified to be used for studies of interannual variability (page 17, lines 6ff). I strongly recommend to add a word of caution here that the agreement of interannual variability very likely will not transfer to long-term trends. Even more so, I suggest to the authors to add a small section that specifically analyses the trends; while this might seem like opening another "can of worms", in my opinion this could be done in a rather short way: If you just calculate the trends of the time-series presented in 7 and 8 and present them in one additional Figure, you clarify how well the trend is constrained (or not). Of course comparison to the FR simulations (and underlying reanalysis were appropriate) would be very valuable here as well. On a similar note, as stated by the authors the time-series were not de-trended before the interannual corrections were calculated, which might obscure the rating of the interannual variability. I recommend to detrend the timeseries for the correlations of interannual variability, and regard the trends separately as suggested above. As long-term trends are the subject of many studies (and you even mention "past trends" in the 2nd sentence of the abstract), and often the naive idea that nudged simulations should get the observed trend right is still around, I think this would be a very important addition to the paper."*

We agree that the incorporation of a discussion about trends renders the study more complete. To this end, and consistent with the additions requested from the reviewer, we have added a new section 4.2.3 that quantifies the trends in all dynamical and transport measures considered in the other sections, as well as a new concluding paragraph in Section 2.3 that explains how we performed these calculations. In addition, we added a new paragraph at the end of Section 5.2 which presents a trend analysis for the SD* and FR ensembles as well as new discussion of the trend analysis in the conclusions.

It is important to note to the reviewer that the trend analysis of the TEM circulation, specifically that of w*, may be considered redundant considering that such analysis was already performed in *Chrysanthou et al. (2019)*. However, we recognize that the reviewer feels strongly that this analysis be included and, therefore, we have incorporated it in our study. Given the previous analysis, though, we found it necessary that we put our trend analysis of the TEM circulation in the context of that study, as the reviewer will note in the second paragraph of Section 4.2.3 and in the new paragraph in Section 5.2 (i.e. with explicit references to their Figures 10 and 11).

In addition to the newly added text we have incorporated two new figures (Figures 9 and 13 in the revised manuscript). While the reviewer suggested that only one additional figure would be needed, we found that this would be impossible given the number of variables that had to be analyzed and in order to be consistent with the other sections in the manuscript. In particular, it was necessary that we present first the trends for the full SD ensemble (Figure 9) and then separately present the trends for the SD* and FR ensembles (Figure 10). As such, these additional figures increase the total number of figures to 14 which we feel is large, but still appropriate, given the broad scope of the study.

Finally, as recommended by the referee we have first detrended all timeseries (summarized above) and then recalculated all correlations of interannual variability. As reflected in the new entries in Table 4 this additional step has little effect on the resulting correlations, which change between 0.1-0.5, depending on the variable. We hope that we have addressed all concerns raised by the referee in this specific comment.

2) *"Reanalysis data of TEM circulation: Is there a particular reason why the analysis of v* and w* is not contrasted to the respective data from the reanalysis? While I understand that calculating those diagnostics is a large effort, they are available from the "SRIP" dataset (see https://www.earth-syst-sci-data.net/10/1925/2018/ and https://catalogue.ceda.ac.uk/uuid/b241a7f536a244749662360bd7839312). If you choose not to include the data, please give a short explanations for the reasons in the paper."*

In response to this comment we now incorporate the v* and w* fields from MERRA, JRA-55 and ERA-I directly from the S-RIP files, as recommended by the reviewer. It is important to note in our response to the reviewer that there was an inconsistency in the units of w* between the fields provided by S-RIP and those provided in the CCMI output. In particular, for all of the S-RIP reanalysis products we needed to apply the following in order to convert the available ω^* fields (units of Pa/s) to w* (units of m/s): $w^*=-(H/p)\,\omega^*$, where $H = RT_s/g$ is a mean scale height of the atmosphere, here taken to be 7 km, corresponding to T_s~240K, a constant reference air temperature. As the reviewer provided no recommendation as to how to perform this conversion we hope that she/he agrees this step was appropriate.

The new v* and w* fields for the reanalyses are reflected in the updated (currently numbered) Figures 2, 5, and Figure 8 in terms of meridional profiles, seasonal cycle and timeseries, respectively. A new mention of this data source, as well as a description of the conversion from ω^* to w* has also been added to the last paragraph of Section 2.2 in which we now have created an entire stand-alone paragraph devoted to explaining the reanalyses products that are used in this study.

In addition to our responses above to the general comments we now respond here to the referee's specific comments:

1) *"Abstract: The abstract emphasizes the differences in the SD simulations strongly, but does not mention that the interannual variability is indeed constrained in the simulations. While I agree*

that the text should clearly state the "warning" to users of SD simulations, it would be fair to also mention the positive outcomes of this study. On the other hand, I think it would be good to mention the results on the dynamical inconsistency in the Abstract, as I think this is a major result to keep in mind when working with SD simulations: the dynamics are not only not well constrained, but they are actually internally consistent (which makes sense as an non-physical term is added to the budgets). This important finding should also be mentioned in the conclusions."

We agree with the referee about both points. However, just to clarify: Does the author mean "internally inconsistent", as opposed to "internally consistent"? Given that this makes the most sense in the context of the reviewer's comment we assume that this is a typo. To this end, the new abstract (see revised manuscript) now explicitly states that the interannual variability of the SD simulations is well constrained, especially compared to the overall magnitude, seasonal cycle amplitude and trends. (Note the latter comment about trends is a new addition to the abstract as well). We have also added a brief clause to the abstract mentioning the lack of dynamical consistency in the SD simulations, a point that we have expanded into a new conclusion "bullet point" (please see new conclusion #5). This new point in the conclusions reads as follows:

"Interestingly, the relationship between the meridional and zonal components of the flow is fundamentally different between the SD simulations and the FR simulations. Unlike the free-running simulations, the specified-dynamics simulations do not exhibit a strong correlation between indices of the Hadley Cell derived separately from the zonal versus meridional winds. This reveals that different components of the flow are not dynamically consistent in all of the SD simulations."

2) *"page 3, line 26: I find the way you introduce the possible reasons for the SD differences a little confusing. Why not list all three points first, and then mention that in the following with "implementation differences" you refer to the named two points (and give a reason why you group them - presumably because it is hard to quantify their relative roles?)"*

OK. We have made this paragraph more easily readable using the suggestions from the referee.

3) *"page 5, line 3-4: Could the transformation of w to omega contribute to model spread in this variable, as probably the density is approximately calculated using the given zonal mean monthly mean temperatures? What is the reasoning behind not using w?"*

We suppose this could be the case, but in practice the differences in w are as large as those between ω. The motivation for recasting in terms of pressure velocities was more one of consistency with our approach in a prior (related) study:

Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., ... & Josse, B. (2018). Large-scale tropospheric transport in the Chemistry–Climate Model Initiative (CCMI) simulations. *Atmospheric Chemistry and Physics*, *18*(10), 7217-7235.

4) *“page 5, line 22: why is the interpolation for "tropospheric variables" performed? Because the CCMI model levels are too coarse in the troposphere?”*

This interpolation was performed so that we could compare all models at the same pressure levels. Even for models with relatively high vertical resolution the surfaces that we used for evaluating all fields (30 mb, 80 mb, 300 mb, 800 mb) did not always correspond to levels on which all models’ output was available. A similar approach for facilitating our comparisons across the simulations was adopted in *Orbe et al. (2018)* (see reference included in previous response).

5) *“page 6, line 12: before mentioning ensemble member, define that the ensemble here is the multi-model suite of SD experiments (earlier, it was mentioned that there are multiple ensembles of the REF-C1 simulations in the conventional meaning of "ensemble", so this is somewhat confusing).”*

Sorry for the confusion. We now explicitly state that we are considering the SD ensemble. Please see the revised text.

6) *page 6, line 24, and entire section 3: In general, I find your way of using "e.g." when listing models a little confusing - this implies that you only list examples, when indeed you do list all the models with certain properties. So please just remove the "e.g." (see also page 7, line 23 and 26). Similarly, when mentioning that "few" models have a certain property, please specify the number. For example in Section 3.3 you say "most" models use HadISST, while "some" use other forcing. While I agree you do not need to list all details here, please be a bit more specific, otherwise this information is more confusing that helpful.*

We agree with the reviewer. In response, throughout section 3 we now have removed all “e.g.” notations in the places recommended by the reviewer (as well as other places). By including all simulations that apply to these statements we have in turn resolved the other issue raised in this comment, which is the ambiguity about “some” versus “most”. Please see the revised manuscript.

7) *“page 9, line 21: the listed differences occur quite close to the equator, where the variable changes it sign, does it? As those result from small shifts in the distribution, it might not be entirely fair to list those as range or difference - rather refer for example to difference in maximum values?”*

The referee is correct that the differences quoted here occur somewhat close to the equator. However, we do not agree that these differences simply reflect small shifts in the underlying meridional profile of v^*. That certainly is plausible but that is not what is happening here. We refer the referee to Figure 2a, from which it is clear that the shape of the meridional profile is relatively consistent among all of the SD simulations. By comparison, several models seriously underestimate the amplitude of the subtropical maxima. These differences do not simply reflect latitudinal shifts in the profile, but rather structural differences over the subtropics. Therefore, we refrain from revising this statement.

8) *"page 9, line 30ff: Good to see you applied the recalculation to check on the possible impact of this issue. According to the Figure in the supplement, the magnitude of differences appears to be larger to me in the recalculated w* values? Therefore, I would recommend to reformulate that statement, in that the differences are at least as high as the ones shown in Fig. 2 (or in other words, that the differences cannot be explained by differences in the calculation method of w*, as the recalculation even emphasizes the differences.)"*

We thank for the reviewer for the suggestion. We have revised this statement accordingly by adding the clause "and even larger" in reference to the differences in w*. Please see the revised manuscript.

9) *"page 10, line 1-2: Because they are clustered by reanalysis product, correct? As you do not present the values from the reanalysis this is indirectly inferred."*

Correct – the reader can see in Fig. 2b that for some variables (as in the quoted example here, w* at 30 hPa) the SD simulations are clustered by reanalysis product. We thought this inference was obvious but if the reviewer is indicating that we should explicitly state this then, yes, we agree. We have modified this sentence accordingly. Please see the revised text.

10) *"page 10, line 5: I find the references to Table 3 mentioned here (and elsewhere) not very helpful. In particular in this paragraph, there is no explanation of why the respective potential reasons are thought to explain the differences. In general, I think Table C4 ACPD Interactive comment Printer-friendly version Discussion paper 3 could be removed, and rather the potential reasons should be mentioned/discussed where appropriate."*

We disagree with the reviewer. While we acknowledge that the analysis could be improved if the reasons for the SD differences were more easily understandable (which would require more simulations, especially targeted simulations suitable for single-model studies) we still think it is useful to remind the reader about the three fundamental sources of differences between the simulations. While one can accomplish this in the text we feel that the use of a table makes this conceptual understanding more clear and evident to the reader. Therefore, we wish to keep the table in the manuscript.

11) *"page 11, line 1: Isn't the spread in U850 almost as large?"*

We are not sure how this comment applies directly to page 11, line 1. The referenced line discusses how we chose to perform averages (required for performing our seasonal cycle analysis) for w* and ω* over 30S-30N (versus 60S-60N for all other variables). As indicated in that paragraph this was informed by the meridional profile of those variables (which tend to change sign around/close to these subtropical latitudes). We are not sure how the reviewer's comment relates to the referenced material.

12) *“page 11, line 3ff: I think here you mix up the phase and the amplitude: the spread in V300 tau_max is present for all three RA (largest for MERRA, but still 3-4 months for the other RA). If you refer to the amplitude rather than phase here, this would make more sense (then the text needs to be rephrased).”*

We are not entirely in agreement with the referee about this point. The seasonal cycle phase plots in Fig. 3b show large uncertainty especially in tau_max, larger than is typically exhibited for the other variables (with the noted exception of omega850, as well as omega300). The referee is absolutely right that discrepancies in the amplitude of the seasonal cycle for this variable (and location) are large, but that does not obscure the fact the phase is also poorly constrained. Indeed, the two are interrelated for this case, as it is the small values of SCA for the MERRA ensemble that contribute to the uncertainty in phase.

13) “*Fig. 4: Please specify more clearly what the individual "dots" represent for the phase: are those the individual models? (and only few are seen in the phase plot because they lay on top of each other?). In the caption it says "show the spread...", which does not clarify what is shown.”*

Correct – each individual dot shows an individual simulation. The fact that only a few dots are sometimes evident is because they all lie on top of each other. We have added a sentence to that figure caption clarifying this point.

14) *“page 11, line 17: Could the effect of small annual means amplify the differences? If one model has zero mean, its annual cycle would be infinity...?”*

Perhaps – the reviewer seems to be more generally asking the question “How well defined is the seasonal cycle?”. We acknowledge that in several cases the SCA is quite small but we also emphasize that we were quite judicious in our choice of latitudinal averaging so as to ensure that the various measures explored in this paper were well defined. Unfortunately, this is difficult to do when considering several (here ~10) variables at a given time. This is why we show the reader a more regionally resolved distribution of the seasonal cycle amplitude in Supplementary Figures 6 and 7. As it is not clear whether or not the reviewer is simply making an (appropriate) comment or, rather, requesting a change we will leave the text as it currently reads.

15) *“page 12, line 16: Do you really mean "positive correlations" here, or "high"? Negative correlations would be very surprising.”*

OK – we have replaced “positive” with “strong.” Please see the revised text.

16) *“page 13, line 9: could you mention the value for N here to give the reader a feeling of the ensemble size (without having to go to Table 1 and count)”*

Good point – indeed, it would be nice to have this defined in the text. However, N depends on which variable is shown since some models output only a subset of variables. The reviewer can note this by looking at the values of N denoted in the titles to (old numbered) Figures 9-11. Note how N varies among the panels--we think that this is a reasonable way to present the varying ensemble sizes for the different variables.

17) *“Fig. 9: I really like this Figure and think it provides valuable information!”*

Thank you. We also like this figure – it is interesting to see how the relative skill of the free-running vs. specified-dynamics simulations changes as a function of altitude.

18). *“page 14, top: I also like the Figures 4 and 5 of the supplement (but agree it is too much to show them in the paper). Some additional interesting features are to be identified, e.g. for omega, it turns out that the RMS at higher latitudes is stronger for FR than for SD. Also for w*, this Fig highlights that the RMS appears to be generally higher in the SD simulations, a result that I’m not surprised by. Mentioning of those interesting features could be worthwhile”*

We thank for the reviewer for her/his thoughtful review of the manuscript – clearly, this is evident in comments like these that require careful scrutiny of the supplementary figures. For that we are exceptionally grateful. At the same time, however, we would appreciate if the reviewer understand that we are struggling to represent a readable summary of the CCMI SD simulations that will not be overly ridden with (what may be) interesting details. While we agree that several of the features noted are interesting, we are afraid that we are already at our limit and that further discussion would detract from the main messages of the manuscript. Therefore, in this spirit we will refrain from including those details in the current draft.

19) *“page 14, line 12: I’d summarize the findings in that SCA is (much) better constrained for T and U for the SD runs, and similar for V, but the spread in the SCA is even larger in the SD runs for omega.”*

We feel that the paragraph in which these statements are embedded already presents this summary point. No changes made to the manuscript.

20) *“page 14, line 16: "poorly (equally)" - confusing formulation, please clarify”*

We agree with the referee – this is poor/unnecessarily ambiguous language on our part. Our apologies for the confusing clause. We have replaced with “…is, by comparison, less well constrained…”.

21) *“page 15, top: Again, one interesting result that could be mentioned here is that the correlations for w* are lower in the SD simulations in the tropics compared to FR, a result that is not as clear from Fig. 11 (where the spread is similar or lower in SD).”*

The reviewer makes a good point – that is interesting. Thank you for noting this point.

22) *“page 17, line 35: Do you need to make the assumption that advection scheme biases are small? I think you can rate it under "underlying free-running model" biases. So even if the circulation would be constrained perfectly (which it is not, see TEM diagnostics), one would expect the advection schemes to induced differences, but those should be smaller than in the FR simulation - but they are about as high (see Fig. 9), i.e. spread in the circulation itself must contribute”*

We do not necessarily assume that the advection scheme biases are small – rather, we assume that any biases due to underlying advection scheme errors should be similar in magnitude for simulations run in free-running vs. nudged modes (i.e. the advection error is fundamental to the numerical scheme and not to the nudging itself). Thus, as the reviewer indicates, this type of error would fall under “free-running model biases.”

23) *“page 18, top: I recommend to include somewhere in the paper the important result of dynamical inconsistency. Also, you can refer to Chrysanthou et al. for the dynamical inconsistency of the TEM budget (the "downward control" calculation works less well in the SD runs).”*

We completely agree with the reviewer (please see our response to Specific Comment 1 above). As noted there we have now modified the abstract and conclusions to better address this results about dynamical inconsistency. We have also indicated that the *Chrystanthou et al. (2019)* study noted similar behavior (please see revised conclusions).

24) *“Table 1: A few questions to be clarified: Are the models with non-constant nudging timescales those that give a range, and the others have constant nudging timescales? Why include information on models that are not used?”*

Correct – ranges in nudging timescales are only presented for simulations in which the nudging timescale is not constant. As regards models that are mentioned in the table, but for which no output is presented (e.g. GFDL) our main commitment in this manuscript was to present the results from those simulations that performed the CCMI REF-C1SD experiment. Unfortunately, some of those simulations did not contain the output required for the analysis addressed herein (e.g. GFDL) but this does not mean that the study does not deserve mention. Thus, given the rather encompassing/broad nature of this manuscrupt’s title we feel strongly that we mention all simulations produced, even those that may not have all necessary output.

In addition to our response to specific comments we thank the reviewer for her/his technical comments:

25) *“Abstract, line 2, and later: the word "online" GCM is not well defined. I would advise not to use it in the Abstract, and if used later on it needs to be defined.”*

We agree with the reviewer that the term “online” GCM has not been defined in the text. Following the suggestion from the reviewer we now have removed that reference in the abstract and included a definition at the end of the third paragraph in the Introduction. Please see the updated manuscript.

26) *"page 3, line 22: I suggest to remove "it is important to note" here, but rather just state that in addition to your paper, there is a study that focuses on the stratospheric circulation (also, the next sentence also starts with "it is important to note")"*

Thank you for the comment. We have replaced the current phrasing with simply "Note the study by *Chrysanthou et al. (2019)*…".

27) *"page 3, line 30: two times "overall" in one sentence"*

Thanks for the comment. We have removed one of the "overall" references as part of our other (broader) revisions of this paragraph. Please see our responses to earlier comments.

28) *"page 3, bottom: sections should be labels 1,2,3,... as in the text"*

OK – we have replaced the numerals I/II/III/IV/V with 1-5 as requested by the reviewer.

29) *"page 4, line 17: rather introduce the model names than institutions? It is not really scientifically relevant whether the contributions are from the same institution, but whether it is the same model that contributes with different set-ups (and as I understand it, the NASA contributions are two different models). Furthermore, also other models contribute more than one set-up (e.g. EMAC L47 and L90)."*

30) *"page 6, line 30: add "information" after "for more""*

Ok – we thought the "information" was implied. Clearly, though, the reviewer does not agree so, per her/his recommendation, we have added "information" after "for more."

31*) "page 9, line 11 and general: mbar or hPa? There is a mixture, and as this an European journal you might just want to use hPa."*

OK – all references to mb replaced to hPa throughout. Please see the revised manuscript.

32) *"page 11, line 3: please put for example V300 in brackets after "the latter", otherwise hard to follow for the reader."*

OK – this has been added. Please see the revised text.

33) *"page 11, line 9: again, please put e.g. meridional winds in brackets after "the former" "*

Done – thanks for catching this oversight.

34) *“page 15, line 5: remove "e.g."”*

This has been removed. Please see the revised text

Response to Reviewer 2

We thank the referee for her/his review of the manuscript. Our response to the reviewer's comments are as follows:

1) *"Specific Comment 1: Language: In general the manuscript is clear and relatively easy to follow, and written in a concise style. In a few places however, some redundant language seems to have crept in to the text. For example on page 9, line 11 "Moving next to the vertical winds" is absolutely not needed in this sentence at all. Similarly, at the beginning of the following paragraph, the sentence could be rewritten without "Next we compare". There are some other examples of this meta-level discussion of what you do first, then what you do next, which for me are simply not necessary. See also page 13, line 12, and page 15, line 3."*

We thank the reviewer for this comment and we agree that our phrasing is, at places, too "meta". We have revised our sentencing at all of the places in the text recommended by the referee as well as in other locations where our phrasing was not appropriate/needed.

2) *"Specific Comment 2: Interannual variability: It would be good to be more specific about what you mean by "interannual variability" in the context of this paper. Clearly you are referring to the correlation of annual average quantities from the model experiments for specific years with the corresponding annual averages from the reanalysis products. This is mostly clear by also referring to this as "covariability". But perhaps some readers might have other ideas about what it means for a model to be representing "interannual variability" well, for example how well periodic variations (eg. ENSO) are represented in a statistical sense (eg. by reproducing probability distributions of their frequency and amplitude), rather than their exact timing. You could avoid this potential confusion by being clearer about what you mean by this term."*

We thank the reviewer for the constructive feedback. Indeed, we could be more precise in our description of "interannual variability". Although we considered our equation of "interannual variability" with "covariability" sufficient it is clear that this is not the case, especially given that previous studies (e.g. Chrysanthou et al. (2019)) have taken other approaches (in the latter case by explicitly regressing timeseries on different modes of variability, including ENSO, QBO, etc.)). We have added new text to Section 2.3 clarifying this. Specifically, the beginning of that modified paragraph now reads as follows:

"In addition to the seasonal cycle we also assess how well the simulations covary with each other on interannual timescales over years 1980-2009. As such, our assessment of interannual variability, which evaluates only the degree of correlation between timeseries, differs from previous studies (Chrysanthou et al. (2019)), in which timeseries were further decomposed in terms of different modes of interannual variability (i.e. the El-Nino Southern Oscillation, the Quasi-Biennial Oscillation, etc). More precisely, for each member within the SD ensemble we identify a given variable χ for which we first remove the linear trend and then calculate the correlation coefficient between the annual mean time series corresponding to that ensemble member i and the annual mean time series of its corresponding ensemble mean…

Our responses to the other comments are as follows:

1) *“Page 2, line 11: Perhaps be explicit that by "boundary conditions" you mean the SSTs and SICs. The flow fields could also be considered a kind of boundary condition.”*

OK – we have changed that line accordingly by making explicit reference to SSTs and SICs. Please see the modified text.

2) *“Page 2, line 13: "improves transport". What would constitute an improvement of transport exactly? It’s not clear what you really mean here.”*

By “transport” we mean, literally, the transport circulation – that is, the integrated advective-diffusive transport circulation that, in the stratosphere, is described by measures like the stratospheric mean age (Hall and Plumb (1994)). In the troposphere, the analogous measure of transport is the tropospheric mean age, as defined in Waugh et al. (2013). We have added a clause to this sentence that hopefully clarifies that we mean something quite precise.

3). *“Section 3.4: Can you go into a little bit more detail about why the different groups chose such different strategies for their choice of nudged fields, domains, and timescales? Can you synthesise relevant parts of earlier work by the respective modelling groups, or even draw on unrelated literature to provide more detail about how and why these choices are made? I think some more context could be useful here.”*

Unfortunately, the feedback we received from modeling centers suggests that a lot of the nudging specifics are very model-specific. Any attempt to generalize this would not be successful, besides perhaps the fact that all centers tend to agree that it is important to nudge both components of the horizontal wind and, in most cases, temperature as well. Besides that, the breadth in nudging timescales and domains is testimony itself that there is no overarching strategy guiding how nudged simulations should be implemented. Our fear about going too much into detail about each modeling center’s strategy is that this would lengthen the (already long) draft and would distract from the main results of the study in the sections that follow. More to the point, we do not think adding these details will help in the interpretation of the results (i.e. why some simulations cluster together in some cases and not in others).

4). *“Section 5.3: I think the lack of dynamical consistency in the SD simulations is as interesting a result of this study as any of the other results you discuss, and could probably be highlighted more in the abstract and conclusions. I would also like to see some discussion here about the implications of this for tracer transport, especially with respect to the use of SD for tropospheric chemistry simulations (especially involving long-range transport), and inverse modelling studies of greenhouse gases.”*

We agree with the reviewer that the lack of dynamical consistency in the SD simulations is an interesting result. Another reviewer raised a similar point and, in response to both comments, we have added a clause to the abstract as well as a new “bullet point” in the conclusions section which references this result. However, we do not agree with the reviewer that this result necessarily has direct implications for inverse modeling studies, which almost always use CTMs, not online nudged CCMs. Therefore, we refrain from referencing inversion studies in the current draft.

5) *“Page 17, lines 18 and 20: The use of the word "cases" is confusing here. It’s not immediately clear what a "case" is. Is it maybe a specific model run? With some effort the reader can see that you actually mean variables or fields (eg. T850hPa), so perhaps use "fields" here to be more consistent with the immediately proceeding sentence.”*

We agree – this is confusing. We have corrected this error – please see the revised text.

6) *“Conclusions: For the benefit of readers who like to read the conclusions section before reading the rest of the paper, it would be good to define all acronyms again in this section. Please consider the following acronyms: SD; TEM; CTM; and CCM. CCMI should be clear enough.”*

We agree – the suggested changes would help readers conducting a more cursory review of the manuscript. We have redefined all acronyms accordingly. Please see the revised text.

Response to Reviewer 3

We thank the referee for her/his review of the manuscript. Our response to the reviewer's comments are as follows:

1) *"page 1, line 2: "Here" could be removed"*

We agree – this sentence has been modified. Please see the revised abstract.

2) *"page 10, line 20: "Next" seems a strange way to start sentence here; how about "We therefore also compare ..."*

We thank the reviewer for the comment and suggested revision. Another reviewer also found this phrasing awkward. We have revised that sentence by beginning with "To explore this last point further we…".

3) *"page 10, line 21: I don't quite follow this argument for restricting the average to 60S60N. If you do the usual cos(lat) area weighting then grid points near the pole are naturally de-emphasised. If there really are issues with a few grid points then restricting it to ∼85S-85N would have been more reasonable. The promised discussion about the sensitivity of choice of latitude bounds at the end of this paragraph (section 5) is a bit cursory (in section 5 you basically state that you've looked at the latitudinal distribution and it looks okay – that's probably not the kind of sensitivity analysis that most readers would expect)."*

Our motivation for restricting our analysis to 60S-60N is primarily to focus on the tropics and midlatitudes. While the reviewer claims that restricting our access to 85S-85N would be sufficient, we do not necessarily agree, since several latitudes in Antarctica south of 70S span regions of high topography. Nudging over these regions may produce still more inconsistencies between models, depending on how different models interpolate the analysis fields to the local grid in regions of strong topographic gradients. As this kind of error, however, is more associated with artifacts related to topography, we did not want to consider that in our more general analysis (which is meant to be global). In order, therefore, to cut out those latitudes but also retain much of the midlatitudes we settled on 60S-60N. Furthermore, as the reviewer points out we tried several ranges of latitudinal bounds when analyzing the model results (e.g. 45S-45N, 30S-30N for all variables) and the main results did not change (see discussion in Section 5).

4) *"page 12: I would find it helpful if you included a bit of discussion on the sources of interannual variability (e.g., due to ENSO, QBO)"*

In response to this comment as well as that from another reviewer we now clarify what we mean more precisely when we refer to "interannual variability." In particular, we have added new text to Section 2.3 clarifying this. Specifically, the beginning of that modified paragraph now reads as follows:

“In addition to the seasonal cycle we also assess how well the simulations covary with each other on interannual timescales over years 1980-2009. As such, our assessment of interannual variability, which evaluates only the degree of correlation between timeseries, differs from previous studies (*Chrysanthou et al. (2019)*), in which timeseries were further decomposed in terms of different modes of interannual variability (i.e. the El-Nino Southern Oscillation, the Quasi-Biennial Oscillation, etc).”

An analysis of each individual mode of variability for all models and for all variables considered here is beyond the scope of this study and, rather, more appropriate in studies like that of *Chrysanthou et al. (2019)*, which only focused on one measure of the flow. We feel that this point is now clarified through the revised text described above.

5) *“page 13, line 22: "not intuitive and" could be removed”*

We agree – we have removed that part of the sentence.

6) *“page 13, line 25: "worse" -> "greater"”*

We agree with this edit – the new draft reflects this change.

7) *“page 13, line 28: "including" -> "included"”*

Thank you for the correction – this has been fixed.

8*) “page 15, line 6: "better interannual variability" – do you mean "more realistic interannual variability"?*

Thank you for catching this – we mean more “consistent” variability. This sentence has been changed – please see the revised text.

9) *“page 17, line 6: "... when inferring dynamics-tracer relationships" – it may be important to clarify that this refers to impact of dynamics on tracers, but not the other way around ("relationship" suggests it could go both ways) ...”*

Thank you for the suggestion. We agree with the reviewer and we have modified this sentence by replacing with “inferring the influence of dynamics on tracers.”

10) *“page 17, line 23: "including" -> "included"”*

Thank you for catching this error. It has been corrected.

Response to “Short Comment” posted by Beatriz Monge-Sanz

In response to the comment from Monge-Sanz we have now included the following two references in the updated manuscript:

Monge-Sanz BM, Chipperfield MP, Untch A, Morcrette J-J, Rap A, Simmons AJ. 2013. On the uses of a new linear scheme for stratospheric methane in global models: water source, transport tracer and radiative forcing. Atmospheric Chemistry and Physics. 13, pp. 9641-9660, https://doi.org/10.5194/acp-13-9641-2013.

Monge-Sanz BM, Chipperfield MP, Dee DP, Simmons AJ, Uppala SM. 2012. Improvements in the stratospheric transport achieved by a CTM with ECMWF (re)analyses: Identifying effects and remaining challenges. Q. J. R. Meteorol. Soc. 139(672), pp. 654-673, https://doi.org/10.1002/qj.1996.