Northern Hemisphere Stratospheric Temperature Response to External Forcing in Decadal Climate Simulations

Abdullah A. Fahad, Andrea Molod, Krzysztof Wargan, Dimitris Menemenlis, Patrick Heimbach, Atanas Trayanov, Ehud Strobach, and Lawrence Coy

Reviewers' comments are in black, responses are in blue color

I had two major concerns in my previous review. One was about the significance of the trends. Rather than computing trends over a few years which does not make sense in my opinion the authors now shows the spread of the ensemble of transient and perpetual-year simulations in Fig.1b. This gives a better insight of the involved uncertainties. I am now more convinced by the first figure and the fact that P2000 and P2020 simulations are significantly different from the P1992 simulation. Figure 8 has also been modified in a similar manner and I found it more satisfying.

My second major concern was about the lack of dynamical interpretation of the change in wave dynamics between P1992, P2000 and P2020 simulations. The authors argue that the paper provides already enough material by discarding the effect of low-frequency variability, volcanoes, radiation and the reasons for the change in wave dynamics is left for future studies. I understand the authors' view but it is a bit strange to finish a paper without providing any hypotheses or future directions of research to look at this problem.

The reader stays quite frustrated at the end and it would be nice to provide some sentences on how to tackle this problem in the future**.** At the same time, the paper has been reinforced by inserting a new figure (Figure 7) showing there is indeed no systematic low-frequency variability that could explain the differences between the different experiments.

I found this additional figure useful and helps convincing the reader. Despite several important improvements, there are several aspects of the paper that appear unclear to me and I would recommend publication once the authors carefully answered the following comments:

We appreciate the reviewer's positive feedback regarding the revisions. We have addressed the remaining concerns regarding the dynamical interpretation and the specific points of clarification individually below.

Major comments:

- Figure 8: I do not understand how it is possible that a high-top CMIP6 model performs less well than the 1°GEOS-MITgcm coupled model. Indeed, the GEOS-MITgcm model reproduces the timing of the warming between 1992 and 2000 very well (Figure 1b) where as the NASA GISS "high top" has a clear delay in the warming (Figure 8). The authors argue about the importance of resolving the stratosphere correctly at the end of the paper but what is the vertical resolution of the GEOS-MITgcm ? Is it low-top or high-top ? The model having 72 levels I would guess that its vertical resolution of the stratosphere is coarser than that of the NASA GISS hightop model.

The initial conditions for the ensemble simulations with GEOS-MITgcm are relatively recent in relation to the stratospheric warming period, and may experience some advantage in the timing of the stratospheric warming relative to CMIP6 historical simulations. The figure and the discussion in the text show the total lack of a simulated warming at any time in the low top simulation, due, in part, we suggest, to the lack of resolution in the model's stratosphere. We have added a line in the text describing Figure 8 to point out the difference between CMIP6 initial simulation dates and those of the GEOS/MITgcm simulations presented here.

Updated text Line numbers: 299-310.

- Figure 7: it is still not clear the reasons why the numerical protocol suppresses any effect of low-frequency modes. This should be better explained in the methodology section (see my minor comments). Is it related to the fact that initial conditions for the P1992, P2000 and P2020 simulations are the same ?

Figure 7 illustrates that the ensemble mean of 30 ensemble members effectively minimizes any low-frequency variability, despite the presence of low frequency variations in individual ensemble members. The first year of the P1992, P2000, and P2020 perpetual experiments were initialized from the same state and were run freely with the forcings specific to their respective years. The presence of the low frequency modes in different phases in the different ensemble members (phase spread in low frequency modes), as shown in the figure, acts to negate the impact of the low frequency modes in the ensemble mean, thus removing low frequency variability as an explanation for the stratospheric warming event as evidenced by the different behavior in the perpetual simulations (P1992, P2000 and P2020).

Texts are added for both the perpetual and transient experiments in the m,anuscript to explain this issue further in the methodology section: Lines: 131-140

- About aerosols effect. The abstract says "Each simulated year of these perpetual experiments is forced with the CO_2, Ozone, anthropogenic aerosol emissions" but at the end the conclusion is that the changes are said to be solely due CO_2 and Ozone. How did the authors conclude that anthropogenic aerosol emissions do not play a role in their simulations ? Maybe I missed a key explanation in the paper.

The budget term for the aerosol radiative impact was evaluated in relation to the other budget terms, as discussed at the end of Section 3.1 and was deemed too small an impact on the temperature to play a role in explaining the temperature trends which are the main subject of the study. Lines: 211-217.

Minor comments:

- Caption Figure 1: "The blue line in (a) shows a 7-year running mean" to be replaced by "The DJF mean is shown in thin black line and its 7-year running mean in blue."

Updated

- Line 78: Was the sharp warming period from 1992 to 2000 reported in previous studies ? Please cite references if appropriate.

Only a limited number of studies have specifically examined the pronounced increase in lower stratospheric DJF Northern Hemisphere temperatures during 1992–2000. Nonetheless, related literature does exist, particularly investigations focusing on the effects of volcanic aerosol loading and the dynamics associated with sudden stratospheric warmings. We have cited relevant references addressing this topic in lines: 49-76.

- Line 117: I do not understand the following sentence "The 30 years of simulation are regarded here as a 30-member ensemble of simulations of the 'perpetual' year, as the initial states for each perpetual year are random."

We have added the following text "The external boundary conditions of the `perpetual' experiments repeat annually while the internal atmospheric state varies, and so each year of the simulation functions as an independent ensemble member (or realization) of that specific year's climate". Lines: 119-121

- Line 118: The authors say "The 30-member ensemble mean of these experiment does not include a realistic simulation of the phase of low-frequency modes of internal variability". This statement should be carefully explained. Is it because the initial conditions of the P1992, P2000, P2020 simulations are the same ?

We ask the reviewer to see our response to the Major comment related to Figure 7, which raises the same question.

- Line 127: It would be better to start a new paragraph when starting the description of the transient experiments

Updated.

- Line 130: It would be good to explain why the numerical setup leads to the following statement: "The low-frequency SST modes are out of phase across these ensemble members"

Please see the response to the major comment about low frequency modes for a clarification of this statement.

- Line 135: please remove "variance". Could you tell the reader for which purposes momentum fluxes are plotted here.

We would like to retain the use of “variance,” as the plots shown in Fig. 2c,d are the mean of U′V′, which represents the mean of the 6-hourly variance of the momentum flux associated with the vortex wind jet. This illustrates how the variability and stability of the vortex in our model simulations compare to the variability of the observed state as estimated from reanalysis. This analysis was added in response to the previous review, and we would like to retain it because it demonstrates how our model simulates the vortex itself, as explained in lines 141–151.

- Line 136: "calculated from sub-monthly fields" is not clear enough. What does that mean ? a subtraction to the monthly mean ?

Model output from our simulations included 6-hourly prognostic fields - these variances are computed as the deviation of the 6-hourly values from the monthly mean, and so we refer to these variances as “computed from sub-monthly fields”. Text updated: 145-146.

- Line 147: Is the definition of the primes the same as in Line 135-136 ?

Yes

- Line 148: "overbar is the time average (DJF seasonal mean)". Is the DJF seasonal mean a climatological mean over all DJF months of all the years or is it dependent on the year ?

The overbar is the mean over all the years. We have updated the text.

- Line 156: please suppress "of"

updated

- Line 160: how do you select the samples ?

These are random samples, and we have updated the text to mention it.

- Lines 190-192: It would be good to refer to the dots in Fig.1b related to the results of the perpetual year experiments.

We have added text to mention the consistency between these two sets of experiments. Line number: 204-205.

- Line 194: "expected radiative effects of increasing CO2" please add a reference

Updated.

- Line 214: "three sets of simulations" to be changed by "three sets of perpetual-year simulations."

Updated.

- Lines 215-216: please replace "result" by "simulation"

Regarding the comment for Lines 215–216: We have checked the text and the word 'result' does not appear in these lines. The sentence reads “To articulate the role of the dynamic tendency terms on the NH stratospheric polar temperature we therefore begin by analyzing the eddy heat flux from the three sets of simulations. ”

- Line 247: Please explain how the simulation design helps to discard the effect of low-frequency climate variability"

Please see our response to major comments on the same topic.

- Figure 7, caption: "dotted lines show individual ensembles" to be replaced by "dotted lines show individual simulations"

Updated

- Line 260-261: The end of the sentence is not clear. What is meant by "propagation speed of the IPO" ?

We have updated the text to say “correct phase” instead of "propagation speed”.

- Line 287: as mentioned above, it is strange that GEOS-MITgcm performs better than the NASA GISS high top model because I think GEOS-MITgcm has a less well resolved stratosphere.

Please see our response to major comments on the same topic.

- Line 319: I am not sure to understand if ODS is the forcing in the simulations presented in the paper or if the model is forced by ozone concentrations

As described in the methods section, only Ozone concentrations were varied for the respective years.

- Line 325: Some hypotheses could be however provided. It is a bit weird to finish a paper without offering possible mechanisms or approaches to tackle that question.

As the revised text in the summary section now states more clearly, we have indeed proposed a mechanism that could explain the warming then cooling behavior seen in reanalyses, our model simulations and the GISS CMIP6 high-top simulation. We suggest a possible indirect impact of the Pinatubo eruption, whereby the emissions from the eruption had an impact on ozone chemistry and locally increased ozone concentrations. This additional ozone generated additional tropical heating and the subsequent increase in eddy heat transport to the polar stratosphere. The summary has been rephrased to make this suggested mechanism more noticeable. Line numbers: 336-344.