Response to Referees of "Stratospheric geoengineering impacts on El Niño/Southern Oscillation," by C. J. Gabriel and A. Robock

Referee comments are in black. Responses are in blue.

Referee #2

1) The authors find no significant difference in ENSO frequency or amplitude under geoengineering scenarios, but do not investigate any other ocean related variables that might more easily show significant changes, like surface winds.

The most easily detected variable in this regard is sea surface temperature. Other variables are not as well simulated by climate models, and analysis of them would make detection of a signal even more difficult. For example, we looked at the Southern Oscillation Index, the atmospheric component of ENSO, and found (p11-12 lines 301-320) that it was not simulated in such a way that it could be used as a proxy for SST, as the area of strong ocean-atmosphere coupling was in the right place, but didn't cover enough area.

We additionally discuss why we did not use surface winds and precipitation. (Please see Section 2 Methods page p11-14 lines 301-370)

2) A more extensive discussion on the reasons why the models do not simulate well ENSO might help in reestablishing confidence**.**

Additional considerations of the difficulties the models we used have in simulating ENSO are now extensively discussed in the new discussion. We discuss the tendencies of each model as seen in our experiment, and as viewed in the context of results about model processes in other papers (Please see p22-23 line 697-722)

3) Given that the latitudinal distribution of the radiative forcing in G3 and G4 will be different from G1 and G2, couldn't these two classes of experiments have very different effects on ENSO? Shouldn't they therefore be kept separately, given the different experiment design? Do the authors assume a-priori that there will be no large difference among experiments?

We agree that the experiments generate climates that are sufficiently distinct so as to negate the value of aggregation. Each group of experiments that we aggregate must be neatly distinct. The simulated future climates must be similar in each aggregated experiment, or the result will not be robust. We have disaggregated G1-G4 and now consider each of the four scenarios separately. We can only justify aggregation of G3 and G4. The following text has been added to the methods section to clarify our criteria for aggregating experiments. (Please see bottom page 9 through middle of page 10 lines 249-266).

"Clearly, detection of changes in future ENSO variability under different scenarios is challenging. As we are limited in both the length and number of geoengineering simulations, we aggregate geoengineering experiments, when appropriate, in order to increase sample size. We combine experiments only when the aggregated experiments form a group that is neatly distinct from its matching comparison group. Aggregated experiments must simulate a future climate that both starts from a similar mean climate and follows a similar trend, or lack of a trend, throughout the experimental period. After applying this standard, we are able to aggregate G1 and G2, since the experiments both initialize from a preindustrial climate and the anthropogenic warming imposed is fully offset by the solar dimming. We are also able to aggregate G3 and G4, since both initialize from a year 2020 climate and follow trajectories in which RCP 4.5 is either fully (G3) or largely (G4) offset by constant sulfur dioxide injections during the experimental period. Application of this standard for aggregation of experiments precludes the aggregation of all GeoMIP experiments G1-G4 into a single ensemble, as the experiments initialize from different climates and follow independent trajectories thereafter. This standard is also applied when we consider aggregating control experiments. Since each control experiment – instantaneous quadrupling of CO_2, 1% yr^{-1} CO_2 increase runs and RCP 4.5 – depicts climates that are distinct from each other, no aggregation of control experiments is performed.

4) The authors show that "disagreement between models was far more significant than that between different experiments and scenarios" Wouldn't this suggest that a multi-model mean is not the best way to proceed, and the authors should rather analyze ENSO changes among different scenario in each single model?

After reviewing the results of our comparisons between models, we find that the dominant contribution to this disagreement between models came from the comparison of either CanESM or CSIRO with the other models. Had we excluded CanESM and CSIRO, the agreement between models would have been reasonably good. However, we choose not to discard either for two reasons. First, CanESM in active periods and CSIRO in relatively quiescent periods capture ENSO variability that is within range over certain intervals from the observational record. Second, both CSIRO and CanESM produce a physically plausible ENSO. The models excluded did not produce physically plausible ENSOs and the variability in those models was entirely outside the range of what has been documented in the observational record. We start 3.3 Comparison Between Models (p20 starting at line 497) with this:

"There was significant disagreement between two of the models. The CanESM depicted ENSO frequency and amplitude that would be more in line with ENSO variability during more active slices of the observational record. The CSIRO model depicted ENSO variability that was more in line with less active slices of the observational record. We choose to include both models as they both simulate plausible ENSO events and variability that is in line with some portion of the observational record."

In terms of how to proceed, we began with the basic question of whether, in one relatively well-performing model, ENSO variability is detectably different under a regime of geoengineering as opposed to global warming as usual. Not surprisingly, such a difference was not evident. Originally, 9 models were used, however, three did not produce a realistic ENSO and they were excluded. (Please see 2 Methods section p10 lines 244-266 for criteria for exclusion). Among the remaining models, some simulated more robust ENSO variability than others, although each of the six models used in the comparisons simulate ENSO variability realistically as compared to observations.

5) Do the authors think that prolonging the simulations might lead to significant changes?

Yes. We have revised the introduction (p10 lines 249-266):

"Detecting changes in ENSO variability is notoriously difficult. The use of lengthy simulations, multiple models, and ensembles is often employed. Cai et al. (2015) were able to detect a statistically significant change in the frequency of extreme La Niña events under RCP 8.5 as compared to a non-global warming control scenario. They selected 21 of 32 available CMIP5 models, because of their ability to accurately simulate processes associated with extreme ENSO events. Each model simulation lasted for a period of 200 years. The detectability of changes in ENSO variability in future SRM modeling experiments will likely be buoyed by the availability of more models and longer simulations. Additionally, future SRM experiments that attempt to offset or partially offset more extreme AGW scenarios, such as RCP 6.0 and RCP 8.5 improve detectability."

Cai, W., Wang, G., Santoso, A., McPhaden, M., Wu, L., Jin F-F., Timmermann, A., Collins, M., Vecchi, G., Lengaigne, M., England, M., Dommenget, D., Takahashi, K. and Guilyardi, E.: Increased frequency of extreme La Niña events under greenhouse warming. Nature Climate Change 5, 132-137, 2015.

6) Is there any trend or change in the ENSO index during the 40 years analyzed?

No trend was evident in our results. The time period analyzed is too short to expect to find a robust trend.

7) The authors state that one statistically significant result the ENSO frequency in RCP4.5 simulations is diminished with respect to historical simulations. Couldn't this also be because of changes in forcing?

After adding ensemble members to each group and repeating the calculations, the RCP 4.5 result is no longer significant. Regarding the changes in forcing, we have added this to the discussion (p24-25 lines 720-726):

"Emissions are imposed differently in the historical runs from how they are imposed in RCP 4.5. RCP 4.5 emissions are imposed decadally, while historical models incorporate gridded monthly data. There is likely a modest amount of interannual variability in tropical Pacific SST that is omitted from RCP4.5 simulations due to the decadal smoothing of emissions. Subsequent research focused on RCP 4.5 ENSO variability could seek to determine if the interannual variability in SSTs is muted enough by this smoothing in RCP 4.5 to potentially alter the evolution of ENSO events."

8) RCP4.5 emissions should be decadal, and I do not think that they contain any interannual variability. Do historic runs include interannual variability of emissions, which might have an effect on SSTs in the Pacific?

Yes, the historical CMIP5 runs incorporate gridded monthly emissions, while RCP4.5 emissions are interpolated to decadal intervals and then scaled globally. The issue of whether interannual variability of emission causes a tropical Pacific SST pattern that is distinct enough from the tropical Pacific SST pattern that would result upon decadal smoothing of emissions to cause a difference in how ENSO events would evolve is worth exploring. The only large shorter-term forcing variability in the historical runs comes from episodic volcanic eruptions, and we discuss those in the paper. We will add the following statement after the RCP Historical result caveat added above in response to item 6 (p24-25 lines 720-726):

"Emissions are imposed differently in the historical runs from how they are imposed in RCP 4.5. RCP 4.5 emissions are imposed decadally, while historical models incorporate gridded monthly data. There is likely a modest amount of interannual variability in tropical Pacific SST that is omitted from RCP4.5 simulations due to the decadal smoothing of emissions. Subsequent research focused on RCP 4.5 ENSO variability could seek to determine if the interannual variability in SSTs is muted enough by this smoothing in RCP 4.5 to potentially alter the evolution of ENSO events."