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 #1

1) Given that no statistically significant change in ENSO frequency or strength is found it somewhat of a missed opportunity to dedicate the entire results section to this finding. The paper would be much richer if this negative result were supported by analysis of more-easily-detected changes in the Pacific which would influence ENSO behavior or be indicative of changes in ENSO behavior.

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 that it was not sufficiently well simulated in most of the climate models to serve as an SST proxy. We further discuss our attempts to consider SLP, zonal winds, thermocline depth and upwelling strength and why each of those approaches introduced thorny detectability issues. (Please see section 2 Methods beginning on page 12 line 301 to p14 line 370).

2) The authors also missed an opportunity in the discussion to help clarify what would be needed in future SRM studies to detect changes in ENSO event properties.

We agree. We have added the following caveat to the end of the introduction section of the manuscript: (Please see p8 lines 204-230)

"We are acutely aware of the challenges inherent in attempting to draw robust conclusions about future ENSO variability. However, 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. Given that detecting an ENSO change in a 200 year record with 21 different participating GCMs is not straight forward, we anticipate that detecting changes in ENSO by analyzing GeoMIP may be difficult."

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.

3) Given the difficulty of detecting changes in ENSO event properties it is understandable that the authors have followed in the footsteps of others and tried to draw on as wide an array of models and experiments as possible (as Cai et al. 2014 do). However, the way in which the authors have done so is problematic and may have led to spurious reports of statistical significance. There are two major problems – the aggregation of results from very different experiments and the uncoordinated nature of the large number of comparisons.

We agree. Section 3.2 has been rewritten to discuss only the comparisons between neatly distinct groups of simulations.

We have eliminated all comparisons that involve aggregation of experiments which do not depict sufficiently similar future climates. For example, since G1 does not produce a linear warming trend, it is inappropriate to combine G1 with G3 or G4. We now only aggregate G3 and G4 for the purpose of comparison. In terms of non-geoengineering runs that are used for comparison, only RCP 4.5 and 1% annual carbon dioxide concentration increase runs are similar enough to combine during the 2030-2069 comparison period. We have added the following text to the introduction. (Also, please see p10 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 of this study attempt the same kind of aggregation however their groupings often exhibit very large intra-group differences. Importantly the authors do not justify why they've chosen the groupings they've investigated or discuss the properties of these groupings.

Please see response to previous comment.

5) In another comparison, they split all experiments into two groups, all GeoMIP (geoengineered) and all non-GeoMIP experiments, but these groups are not neatly distinct in the same way as the control and warm groups used by Cai et al. 2014. Both groupings have a mix of warmer and cooler climates with some warming and some in steady-state, all of which would be expected to affect ENSO properties. It is unclear what would be demonstrated if there were statistically significant differences between these groupings as there doesn't seem to be a consistent approach for choosing the constituents. Similar issues arise for other groupings that the authors make.

Please see response to previous comment.

6) The authors also test so many different groupings of models and experiments that it is unsurprising that they detect statistically significant changes at the 5% level. How many comparisons were made and how many false positive results would be expected if all of the data were drawn from the same sample? No effort is made to test whether such spurious statistical significant results were likely nor is any note made of this basic problem. The authors could consider some form of bootstrapping resampling approach which may help to give an idea of how robust the statistics derived from these groupings are.]

We agree. After applying the condition that each grouping be distinct, we only obtain two significant differences in the remaining comparisons. We qualify the robustness of these results with a simple resampling technique. After performing the resampling, it appears that the significant results were obtained by chance. (Please see section 3.2 Analysis, which has been completely rewritten)

7) Putting these issues to one side, the main finding of this study is that the effect of SRM on ENSO are not detectable in 40 year records of ENSO 3.4 temperatures. However, was this not obvious beforehand? ENSO behavior is notoriously variable and would be expected to show substantial variations in 40-year statistics. The authors should do more to explain and investigate the challenges of detecting a change in ENSO behavior.

We extensively discuss the impacts of volcanic eruptions on ENSO. These papers describe ENSO behavior in the immediate aftermath (< 5 years) of the eruption. We test the hypothesis that long-lasting solar dimming or sulfate injections could produce a signal similar to the volcanism signal, but operating on a longer time scale. We did think it unlikely that a difference between geoengineering ENSO and global warming ENSO would be evident after comparing experiments and controls from one model. However, we thought that using multi-model ensembles, and aggregating experiments where appropriate, might make such changes detectable.

Additionally, please see additional discussion on p8 line 205-224 in the final paragraph of the introduction for discussion of the challenge of evaluating ENSO behavior.

"We are acutely aware of the challenges inherent in attempting to draw robust conclusions about future ENSO variability. However, 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. Given that detecting an ENSO change in a 200 year record with 21 different participating GCMs is not straight forward, we anticipate that detecting changes in ENSO by analyzing GeoMIP may be difficult. Further, we recognize that even if significant differences between ENSO in a geoengineered world as opposed to an AGW world are evident, a large number of comparisons will have be made, and further analysis of significant results will need to be performed to determine whether or not the result is robust. Despite these substantial caveats, it would be irresponsible for geoengineering research to progress without consideration of how a geoengineering regime could alter ENSO.

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.

8) The authors could tackle this detection issue directly by analyzing the ENSO 3.4 behavior in the long pre-industrial steady-state simulations for each of the 6 models used. This could then make clear how variable 40-year ENSO statistics are. Alternatively the authors could refer to existing results or theoretical considerations from the literature to develop these expectations. There are studies, which investigate variations in ENSO statistics in those long pre-industrial simulations in CMIP5 which should be referred to.

There are 150 years of historical simulations available for each ensemble member of each model. We have analyzed all of the available historical simulations. We have compared the variability within the historical simulations with that seen in comparing the geoengineering and AGW experiments. The analysis shows, if anything, more variability during the historical period than during the experimental period. The historical runs were chosen because the rate of warming throughout the period is relatively modest when compared to future global warming scenarios and because the vast majority of our simulations were not steady state. It seems more relevant to squarely address the question of whether ENSO variability under AGW or geoengineering would be different than what is evident in observations and in historical simulations. We though it more appropriate to constrain the variability between future ENSO scenarios with the variability seen in the historical record. This allows us to more squarely address the question of whether ENSO will be fundamentally different from what it has been under geoengineering and/or future AGW.

We also discuss the methods employed by Cai et al. (2014, 2015), a very sophisticated approach, for which a great deal of data was available to the authors. The description of the Cai et al. experiments in our introduction is a strong caveat, limiting ab initio what the limits detectability might be in this experiment.

Please also see the discussion for suggestions of what might be needed to detect changes in future geoengineering experiments.

9) [I]t would have greatly enriched the study if there had been some discussion of what requirements would be needed for an experiment to investigate the effects of SRM on ENSO…
a.) What averaging periods and number of ensemble members would be required?

Replicating the approach of Cai et al. (2015) and selecting from an array of 32 models, each providing a 100+ year simulation length is not possible for the SRM experiments so far conducted. However, one first step would be to select the single best performing GeoMIP model in simulating ENSO. We could then eliminate the termination of SRM in 2070 and allow the model to continue to run for another 100 years. We could then evaluate the question of whether, in a particular model that does well in simulating ENSO, there is evidence of a difference in ENSO variability between the SRM experiment and the appropriate control.

Future GeoMIP experiments plan model simulations of 100 years of geoengineering (Kravitz et al., 2015) and they would not only allow for analysis of changes in ENSO, but of the potential impacts of geoengineering on other sources of natural climate variability.

(Please see the first part of section 4 Discussion p21 line 668-673)

Kravitz, B., Robock, A., Tilmes, S., Boucher, O., English, J. M., Irvine, P. J., Jones, A., Lawrence, M. G., MacCracken, M., Muri, H., Moore, J. C., Niemeier, U., Phipps, S. J., Sillmann, J., Storelvmo, T., Wang, H., and Watanabe, S.: The Geoengineering Model Intercomparison Project Phase 6 (GeoMIP6): simulation design and preliminary results, *Geosci. Model Dev. Discuss.*, *8*, 4697-4736, 2015.

b.) Are steady-state simulations to be recommended over transient simulations?

Probably extreme steady-state simulations would be the place to look. If there is nothing there, then transient runs with smaller forcings would not be likely to show much. In GeoMIP6 (see previous answer), 4xCO2 will be balanced by solar dimming for 100-year runs (experiment G1ext). We think comparing those results (4xCO2, geoengineered climate, and control climate) will let us see if there are any potential signals. (Please see the first few paragraphs 4 Discussion p21-22 lines 674-686)

c.) What recommendations do the authors have for SRM deployment strength in such experiments?

As discussed above, large offsets from global warming would be the best. Similar experiments using more realistic sulfate aerosols in the stratosphere would be preferred to solar dimming, but those are not planned with such large forcing. But comparisons of the proposed G6solar and G6sulfur runs will allow us to quantify how much difference this makes. (Please see the first few paragraphs of p21-22 lines 674-686)

d.) And what do these recommendations suggest about the possibility of detecting changes in ENSO in the next generation of GeoMIP experiments?

See previous answer.

10) No statistically significant results were found for ENSO which is an important result but given this I was surprised that this result was not supported by some statistically significant results on other related measures of change in the Pacific. The authors identify a number of important factors which will influence ENSO behavior which would seem to be more readily detectable: Pacific zonal SST gradient, the strength of trade winds, upwelling strength, thermocline depth, etc. (see Guilyardi et al. 2012 for more). In addition, whilst not necessarily definitive, results on changes in the mean and interannual variability of ENSO 3.4 temperatures would be far easier to analyze and would be of interest.

ENSO 3.4 SST was selected as a preferred indicator of ENSO behavior for several reasons. We found that using variables related to surface wind were less reliable than SST. The historical records of thermocline depth and upwelling strength are not robust. Pacific zonal SST gradient changes could be difficult to attribute fully to ENSO. We have added extensive discussion reflecting on why were forced to work with SST as our primary variable. (Please see section 2 Methods p12 -14 lines 301-370 discussing our rationale)

While changes in ENSO 3.4 mean temperature over the 40 year period simulated is important, we feel that dedicating a figure or much discussion to this would be a bit redundant given the extensive body of literature that exists showing trends in SST both under both geoengineering and global warming scenarios.

11) The authors hypothesize in essence that if the volcanic forcing were persistent (as in SRM simulations) there would be a permanent shift in ENSO behavior. An alternative that goes uninvestigated is that this change in ENSO behavior is solely a transient phenomenon, one which may also be evident in the initial phases of some of the GeoMIP simulations. However, the authors discard the initial 10 years and so such transient phenomena are removed. I'm unsure an investigation of this transient behavior would be fruitful but G1 and G4 should produce a transient cooling that may give rise to the same phenomenon.

We provide a discussion of the literature on the impact of the pulse forcing produced by volcanoes on ENSO. While this is not what we focus on in the manuscript, our review of the literature leads us to conclude that the impact of the pulse forcing brought on by commencing SRM would be similar to that produced by a volcano that produces radiative forcing equal to the radiative forcing produced by commencing SRM. Most proposals for implementing SRM are for a gradual implementation, so we would not expect such a pulse response. In fact, the technology to instantaneously inject the same amount of sulfur as a large volcanic eruption does not exist, and such a capability has not been proposed. Therefore we confine our focus to investigating the possibility of a permanent shift in ENSO behavior, as that issue has not yet been addressed.