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.

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."

Referee #3

1) The manuscript seeks to identify changes in ENSO frequency and amplitude under various historical, projected, and geoengineering scenarios. The paper is well motivated and clearly written. That said, there are major caveats associated with limitations in available experiments that go under-appreciated in the text and greatly limit what can be done. I find the results therefore somewhat unconvincing. I recommend strongly that the manuscript provide a more direct appreciation for the inherent limits of the simulations used and provide better context for what is needed to really address this questions with greater certainty.

We agree. We have added a discussion of the inherent limitation of our experimental design to the introduction. We hope that the additional context provided will not only clarify our results, but provide a road map for what types of future simulations would be most useful in detecting potential changes in ENSO variability under various geoengineering regimes.

"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.

2) The experiments used are useful but fail to provide a very tight constraint on the null hypothesis being posed because of the facts that 1) so few ensemble members are available for each model, and 2) ENSO is so poorly and variably simulated across the models (as alluded to in 3.3)- contributing to large error bars and hence coarse detectability of any potential change. There is considerable uncertainty associated with the changing mix models across the metrics being computed that is not adequately dealt with. It would seem to be essential to me to make this weighting constant across any comparisons being made.

In the revised manuscript, each model included in each comparison group is weighted equally. Additionally, we have eliminated the aggregation of multiple experiments that produce large ensembles, but also mix experiments with dissimilar climates into the same ensemble. Please see the revised section 3.2. In all comparisons, each model is weighted equally in both the experimental and control runs. If there are three experimental ensembles for a particular model, there are also three control ensembles for that model.

3) I recommend that manuscript start with the most simple question: in a single scenario and for a single model, do any provide enough ensemble members to detect a change in ENSO? I presume the answer is 'no', since it is not dealt with - but pointing this out would be useful for motivating the need to create multi-model ensemble metrics. In cases where significant differences are identified - what is the role of changes in the model mix?

We agree that pointing this out would be valuable. We have added this discussion to start section 2 Methods (p9 lines 226-231):

"We begin with the simple question of whether or not, in a single GeoMIP participating model that simulates ENSO well, a difference in ENSO amplitude or frequency is evident. Unsurprisingly, given the large inherent variability in ENSO, such a change is not detectable in one model. Given that, we adopt an approach in which we use output from nine GeoMIP-participating GCMs, each running between one and three ensemble members, of each experiment G1-G4. "

4) Please put +- 1-sigma values on the model-mean numbers. I think this provides essential context

± sigma values have been added.

5) Detection of only two significant results in the context of the large # that have been done is at the limits of what may be expected by chance. An associated caveat should be added here.

As mentioned above, we have eliminated comparisons in which we combine experiments that depict climates that are dissimilar. This limits the number of comparisons performed. Only two significant results remain and they are at 90% confidence. However, a simple resampling with replacement technique revealed that the significant result was likely actually the result of chance.

6) Why would you put more significance on the RCP4.5 finding that ones that have assessed the question in a broader array of models?

We have rewritten the abstract so that it does not call attention to RCP 4.5. Not mentioning the other experiments with equal weight in the abstract was likely an oversight by the author. Please see section 3.2, as it has been rewritten. Any special emphasis on the RCP 4.5 finding is no longer present.

7) Doesn't the fact that SOI is not a useful ENSO proxy speak to the inherent deficiency of using a given model for this type of analysis? How can one expect to get a reasonable bearing on the dynamical-thermostat mechanism or other dynamical links of forcing to ENSO if the SOI relationship is so poor since essentially the dynamic component of ENSO (SOI) also so poor? Shouldn't this be an additional constraint on which models to use?

The area of high correlation (> 0.5) is suppressed in models relative to observations. However, the region of highest correlation is in the correct location. The spatial pattern is similar, but the value of the correlation coefficient is muted. Also, as can be seen in Figure 3, the temporal relationship between SST and SOI is realistically simulated. The strong ocean-atmosphere coupling that is evident in the heart of the immediate equatorial central and eastern Pacific shows that the models are depicting a plausible ENSO cycle, albeit over a smaller area. In models that were excluded, the SOI SST correlation was plainly unrealistic, rather than just muted in spatial extent. Further, we sampled several SST-SOI correlations from our analysis, looked up the maximum value of the correlation, and found that it was around 0.8 in both observations and the models sampled. Therefore, it seems that the most robust ocean-atmosphere coupling was occurring, just over too small an area. This distortion of the spatial pattern was substantial enough go that SOI could not simply stand-in for SST, but it was not cause to discredit the depiction of ENSO dynamics in the models we used.

8) How can one establish confident ENSO statistics from such a short duration/limited ensemble of runs? Model runs suggest that robust statistics of ENSO (particularly at its low frequency tails) require records of over a century. What has been done here (to group all of them together) might be justified if they all had the same ENSO statistics but clearly they do not.

We agree that detectability of changes in ENSO may be inhibited by the unavailability of more lengthy simulations. We have analyzed 150 years of historical simulations for all ensembles within each model. We have also now done further analysis to determine if differences in ENSO variability between geoengineering and control runs during the 40 year period is greater than or less than the differences between selected 40 year periods in the historical data. This has allowed us to determine that ENSO variability between experiments is less than that seen when comparing 40 year slices of the historical period with each other. Therefore, based on our findings, ENSO variability under geoengineering as compared to under AGW would not exceed the variability found within the historical record.

9) The fact that some models have unrealistic ENSO behavior is hardly a new result and I don’t think it requires 2 figures. A mere sentence in the text would suffice. Moreover, internal variability of ENSO could lead to periods of such low variability even with a reasonable ENSO and thus I’d base any such statement on multiple ensemble members or an extended control.

The results shown in the two figures mentioned are selected because they are typical of what is seen across most of the models used. The purpose is to show that the ocean-atmosphere coupling is muted in terms of its spatial extent, but similar to observations in terms of the maximum value of SST/SOI correlation. This negates the possibility of using SOI in place of ENSO 3.4 SST as an ENSO index. Also, the SST/SOI correlation result is relevant to how ENSO is depicted in the simulations used. The figures are intended to place the results about ENSO variability in the context of the underlying physical mechanism. We do not assert that the SST/SOI correlation demonstrates for the first time that some models depict unrealistic ENSO behavior.

Additionally, the figures augment the discussion of the ocean-atmosphere coupling that is an essential part of ENSO. A visual depiction of the dynamics of ENSO as provided in Figures 2 and 3 may clarify the discussion for some readers.

10) Maximum event magnitude (e.g. Fig 9) doesn't seem like a very robust metric to use given the limited length of these runs. Why not use total variance?

Maximum event amplitude, mean event amplitude and total variance were all considered for use in the figures. The results pertinent to event amplitude would have ranked the models and experiments similarly. The readers are likely familiar with the magnitude of the strongest and weakest ENSO events in the observational record and reporting results in terms of maximum event magnitude is done for clarity.

11) On the discussion: We already knew changes in ENSO were inconsistent across models (e.g. Guilyardi et al 2012). This is not new It is likely that additional model runs should have been rejected based on the importance of dynamics in the science questions being posed and the lack of SOI fidelity. It seems odd that the authors used this as a basis for rejecting the SOI rather than the models! Perhaps a dynamical validation combined with a power spectrum validation would be a more appropriate way to screen models.

The area of high correlation (> 0.5) is suppressed in models relative to observations. However, the region of highest correlation is in the correct location. The spatial pattern is similar, but the value of the correlation coefficient is muted. Also, as can be seen in Figure 3, the temporal relationship between SST and SOI is realistically simulated. The strong ocean-atmosphere coupling that is evident in the heart of the immediate equatorial central and eastern Pacific shows that the models are depicting a plausible ENSO cycle, albeit over a smaller area. In models that were excluded, the SOI/SST correlation was plainly unrealistic, rather than just muted in spatial extent.

12) The question of whether the 1966-2005 period is really adequate to validate modeled ENSO is never addressed but needs to be considered. ENSO statistics varied considerably through the course of the 20th C.

We have now analyzed the full historical period of approximately 150 years for each model. We took a number of 40-year time slices from the 150-year historical record and created ensembles to test the variability of 40 year ENSO statistics. Variability between 40 year periods in the historical record was at least as large as that seen between geoengineering and AGW simulations. (Please see 3.2 analysis, which has been rewritten) None of the time slices were significantly different from each other at a 95% confidence level on any of the metrics tested. ENSO frequency in the 1966-2005 period was very similar to that seen in the full 150-year record. However, the 1966-2005 time period failed to capture the strongest warm and cold events in many of the 150-year historical periods. It is somewhat reassuring that the maximum warm and cold event amplitudes in the observational record was almost identical to the average maximum warm and cold event amplitudes in the simulations.

Given the paucity of observations of SST over the Niño regions prior to 1960, we see less value in matching pre-1960 historical simulations to observations.

Stratospheric Geoengineering Impacts on El Niño/Southern Oscillation

Corey J. Gabriel and Alan Robock
Department of Environmental Sciences, Rutgers University, New Brunswick, NJ, USA

Submitted to *Atmospheric Chemistry and Physics*
GeoMIP Special Section

March 2015

Revised, July 2015

Corresponding Author:

Corey J. Gabriel
Department of Environmental Sciences
14 College Farm Road
Rutgers University
New Brunswick, NJ 08901
USA
Email: cjgabriel7@gmail.com

Abstract

To examine the impact of proposed stratospheric geoengineering schemes on the amplitude and frequency of El Niño/Southern Oscillation (ENSO) variations we examine climate model simulations from the Geoengineering Model Intercomparison Project (GeoMIP) G1-G4 experiments. Here we compare tropical Pacific behavior under anthropogenic global warming (AGW) using several scenarios: an instantaneous quadrupling of the atmosphere's CO_2 concentration, a 1% annual increase in CO_2 concentration, and the representative concentration pathway resulting in 4.5 W m^{-2} radiative forcing at the end of the 21st Century, the RCP4.5 scenario, with that under G1-G4 and under historical model simulations. Climate models under AGW project relatively uniform warming across the tropical Pacific over the next several decades. We find no statistically significant change in ENSO frequency or amplitude under stratospheric geoengineering as compared with those that would occur under ongoing AGW, although the relative brevity of the G1-G4 simulations may have limited detectability of such changes. We also find that the amplitude and frequency of ENSO events do not vary significantly under either AGW scenarios or G1-G4 from the variability found within historical simulations or observations going back to the mid 19th century.

1 Introduction

1.1 Background

The warming of Earth in the Industrial Age is unequivocal, and it is extremely likely that the warming since 1950 is primarily the result of anthropogenic emission of heat trapping gases rather than natural climate variability (IPCC, 2013). Ice core records from the European Project for Ice Coring in Antarctica (EPICA) reveal that current concentrations of the heat trapping gases carbon dioxide and methane are higher now than at any time during the past 650,000 years. (Siegenthaler et al., 2005). All realistic emissions scenarios utilized in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report reveal that the modeled global mean temperature in 2100 will exceed the full distribution of global mean temperature in proxy reconstructions of global temperature over the past 11,300 years of the Holocene (Marcott et al., 2013). Ongoing warming is unprecedented in human history both in magnitude and rate of change.

The realization that weathering the impacts of this warming may be beyond human adaptive capacity has generated many proposed mitigation techniques, which focus on limiting emission or increasing storage of heat-trapping gases such as carbon dioxide. Implementation costs and economic, political and societal factors limit societies' will and ability to impose mitigation measures. This has forced recent consideration of geoengineering – intentional manipulation of global-scale physical processes (Crutzen, 2006). Specifically, a form of solar radiation management (SRM) known as stratospheric geoengineering has been proposed. Annual sulfate injections into the tropical stratosphere have the potential to create a long-lasting, well-mixed sulfate aerosol layer, which could reduce incoming shortwave radiation, in an attempt to offset the warming by the excess heat-trapping gases (Robock, 2008). The cost of implementing stratospheric geoengineering is most likely not prohibitive (Robock et al., 2009). Any decision about the implementation would likely be based on substantive issues of risk and feasibility of governance (Caldeira et al., 2013).

Assessment of the efficacy and risk profile of stratospheric geoengineering is underway in a series of standardized climate modeling experiments as part of the Geoengineering Model Intercomparison Project (GeoMIP) (Kravitz et al., 2011). Any assessment of the impact of geoengineering on climate must include analysis of how geoengineering could alter patterns of natural climate variability and how geoengineering could change the mean climate state in such a way that natural climate variability would evolve differently in an intentionally-forced world.

1.2 Research question and motivation

Here we seek to examine whether stratospheric geoengineering would have any impact on the frequency or amplitude of El Niño/Southern Oscillation (ENSO). More specifically, will ENSO amplitude and frequency be different under a regime of geoengineering from that in a global warming scenario? ENSO is the most important source of interannual climate variability. Its amplitude, frequency and the attendant teleconnection patterns have critical consequences for global climate patterns. (McPhaden, 2006). ENSO exhibits a 2-7 year periodicity with warm (El Niño) and cold (La Niña) events each lasting 9-12 months and peaking during the DJF season.

The possibility of a connection between warm ENSO events subsequent to stratospheric aerosol loading via volcanism has been explored both in proxy records and model simulations. Despite its relative simplicity, the Zebiak-Cane (ZC) model (Zebiak and Cane 1987) possesses an exceptional ability to describe the coupled ocean-atmosphere dynamics of the tropical Pacific. By forcing ZC with the calculated radiative forcing from each eruption in the past 1000 years, Emile-Geay et al. (2007) showed that El Niño events tend to occur in the year subsequent to major tropical eruptions, including Tambora (1815) and Krakatau (1883). A strong enough cooling by a volcanic event is likely to cause warming in the eastern Pacific over the next one to two years (Mann et al., 2005). The dynamical "ocean thermostat" describes the mechanism underlying differential heating in the eastern and western Pacific. In the presence of a global strong negative radiative forcing, the western Pacific will cool more quickly than the eastern Pacific. This is because the western Pacific mixed layer's heat budget is almost exclusively from solar heating, while, in the east, both horizontal divergence and strong upwelling contributes to the mixed layer heat budget. Therefore, a uniform solar dimming is likely to result in a muted zonal sea surface temperature (SST) gradient across the equatorial tropical Pacific (Clement et al., 1996). A diminished SST gradient promotes a weakening of trade winds, resulting in less upwelling and an elevated thermocline, further weakening the cross-basin SST gradient. This "Bjerknes feedback" describes how muting of the SST gradient brought on by negative radiative forcing alone is exacerbated by ocean-atmosphere coupling (Bjerknes, 1969). Following the initial increase in El Niño likelihood, La Niña event probability peaks in the third year post-eruption (Maher et al., 2015).

Trenberth et al. (1997) placed the likelihood of an ENSO event in a given year at 31%. Using 200 ZC simulations lasting 1000 years each, Emile-Geay et al. (2008) showed that the probability of an El Niño event in the year after the simulated volcanic forcing never exceeded 43% absent negative (volcanic) radiative forcing of greater than 1 W m^{-2}, with modeled next year El Niño probabilities clustered around 31%. Volcanic events with radiative forcing ranging from –1 W m^{-2} to –3.3 W m^{-2} fit into a transition regime, with the number of events approaching or exceeding the 43% probability maximum. For all modeled volcanic events with radiative forcing exceeding –3.3 Wm^{-2}, the probability of a next year El Niño exceeded 43%. This is a forced regime – negative radiative forcing applied to the ZC model forced El Niño likelihood out of a free regime and into a regime where enhanced variability would be more likely (Emile-Geay et al. 2007). In the transition and forced regimes, increased El Niño amplitude is also simulated following moderate to strong volcanic events.

Geoengineering schemes simulated in current general circulation models (GCMs) introduce long-lasting radiative forcing of the magnitude found in the transition regime. This means that while radiative forcing of that magnitude does not force the probability of a next year El Niño event to exceed the 43% free oscillation maximum threshold, instead the radiative forcing applied does fit into a range in which the 43% threshold is exceeded during the next year in some simulations. Therefore, we ask whether solar dimming lasting many years, as a proxy for sulfate injections, or sulfate injections lasting many years as simulated by models may also alter El Niño or La Niña event frequency and amplitude. Rather than using the ZC model, we use various geoengineering experiment designs in modern, state-of-the-art GCMs to determine whether forcing from stratospheric aerosol injections, added continuously, will load the deck in favor of El Niño events in the succeeding year. No modeling study has ever evaluated the impact of long term solar dimming or continuous stratospheric sulfate injections on ENSO.

Since our comparison is between El Niño and La Niña amplitude and frequency under a geoengineering regime and under a scenario of unabated global warming, the evolution of ENSO behavior under global warming, independent of geoengineering, is also of interest. Overwhelming evidence from climate model experiments shows that geoengineering could effectively reduce or offset the surface temperature increase resulting from global warming by limiting the amount of incoming shortwave radiation, compensating for global warming (Jones et al. 2013, Robock et al. 2008). An alternative theory for why ENSO amplitude and frequency may be different in the future, under a geoengineering regime, than under global warming is based on the fact that ENSO events may evolve differently from a warmer tropical Pacific mean state under global warming than if a geoengineering scheme were imposed.

Kirtman and Schopf (1998) showed that tropical Pacific mean-state changes on decadal timescales are more responsible than atmospheric noise for changes in ENSO frequency and predictability. This does not imply any external cause for the changes in ENSO, but does imply that a uniform warming of the tropical Pacific may cause changes in ENSO. Despite the lack of a robust multi-model ENSO signal in the Coupled Model Intercomparison Project 5 (CMIP5) models (Taylor et al., 2012), there are suggestions that strong El Niño events may become far more likely under global warming, specifically in a multi-model ensemble experiment using RCP8.5. As global warming continues, background state tropical Pacific SSTs are expected to warm faster along the equator than off the equator, and faster in the east than in the west – the inverse of the ocean dynamical thermostat mechanism (Held et al., 2010). With the weaker zonal SST gradient in the tropical Pacific, there will be more occurrences of higher SSTs in the eastern Pacific, promoting large scale organization of convection further to the east, with twice as many strong El Niño events over 200 years of RCP8.5 runs (Cai et al., 2014). We will not seek to replicate the RCP8.5 results. No physically plausible geoengineering experiment would seriously attempt to offset RCP 8.5 with solar dimming or sulfate injections. Therefore, we use RCP4.5 as the control in GeoMIP experiments, and will attempt to identify if the long term mean state changes generate divergent ENSO frequency under geoengineering and global warming.

1.3 Representation of the Tropical Pacific in CMIP

The ability to detect subtle differences in the tropical Pacific under global warming vs. geoengineering requires sufficiently skilled models. Proper depiction of ENSO in a GCM is confounded by the fact that ENSO is a coupled ocean-atmospheric phenomenon, generated by the interaction of many processes, each occurring on one of several different time scales. Nearly all CMIP3 models were able to produce an ENSO cycle, but significant errors were evident (Guilyardi et al., 2009). Analysis of CMIP5 models has shown significant improvement, but the improvement has not been revolutionary. Such a comparison is facilitated by standardized "metrics developed within the CLIVAR [Climate and Ocean: Variability, Predictability and Change] Pacific Panel that assess the tropical Pacific mean state and interannual variability" (Bellenger et al., 2013). The following metrics were used in the CLIVAR CMIP3/CMIP5 comparison: ENSO amplitude, structure, spectrum and seasonality. Some process-based variables were also studied, including the Bjerknes feedback.

Key results included that 65% of CMIP5 models produce ENSO amplitude within 25% of observations as compared to 50% for CMIP3. Other results included improved seasonal phase-locking and the proper spatial pattern of SSTs at the peak of ENSO events. Despite the improvement in these result-based variables, analysis of process-based variables, such as the Bjerknes feedback, showed less consistent improvement. This gives rise to the possibility that the bottom-line improvement in ENSO depiction was at least partially the result of error cancellation, rather than clear improvements in parameterization and simulation of physical processes (Yeh et al., 2012; Guilyardi et al., 2012; Bellenger et al., 2013). A particularly striking area of divergence between modeling and observations is in the absence of a shift from a subsidence regime to a convective regime in the equatorial central Pacific during evolution of El Niño events. Many models maintained a subsidence regime or convective regime at all times over the equatorial central Pacific (Bellenger et al., 2013). This error likely led to the muting of the negative shortwave feedback in many models, leading to muted damping of ENSO events in those models.

Both the improvement in depiction of ENSO amplitude and seasonality from CMIP3 to

CMIP5 and the ability to understand the simulation of key process-based variables motivate an analysis of ENSO and geoengineering using CMIP5 GCMs.

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.

2 Methods

We begin with the simple question of whether or not, in a single GeoMIP participating model that simulates ENSO well, a difference in ENSO amplitude or frequency is evident. Unsurprisingly, given the large inherent variability in ENSO, such a change is not detectable in one model. Given that, we adopt an approach in which we use output from nine GeoMIP-participating GCMs, each running between one and three ensemble members of each experiment G1-G4. The simulations are then are analyzed. These GeoMIP experiments are described by Kravitz et al. (2011). See Figure 1 for schematics of GeoMIP experiments G1-G4 and Tables 1 and 2 for details about the GCMs used in these experiments. The G1 experiment – instantaneous quadrupling of CO_2 coupled with a concurrent fully offsetting reduction of the solar constant – was designed to elicit robust responses, which then facilitate elucidation of physical mechanisms for further analysis. We compared G1 output to a control run in which the atmospheric carbon dioxide concentration is instantaneously quadrupled. The G2 experiment combines a 1% $year^{-1}$ CO_2 increase with a fully offsetting reduction in the solar constant. The G3 experiment combines RCP4.5 with a fully offsetting sulfur dioxide injection. The G4 experiment – stratospheric loading of 25% the SO_2 mass of the 1991 Mt. Pinatubo volcanic eruption (5 Tg) each year concurrent with RCP4.5, with top-of-atmosphere radiation balance not fixed at zero – attempts to replicate a physically and politically plausible large scale geoengineering deployment scenario.

Each experiment (G1–G4) is compared to its respective control scenario: 4xCO2 for G1, 1% annual CO_2 increase for G2 and RCP 4.5 for G3 and G4. 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.

To identify and analyze ENSO variability and amplitude, absent the contamination of the signal induced in the immediate aftermath of application of initial solar dimming or stratospheric aerosol forcing, the first 10 years of each geoengineering model run were removed. The relevant comparison periods become either "Years 11-50" in G1 and 2030-2069 in G2-G4. Initial forcing is applied in "Year 1" in G1 and in 2020 in G2-G4. This 40-year interval is then compared to RCP4.5 2030-2069 and historical 1966-2005 for each respective model and to observations. We used the Kaplan (1998) SST data set because it is well-documented and used in many of the referenced papers. Differences between the Kaplan data and other available data sets are trivial during the period of data used.

We used several SST-based indices to quantify the amplitude and phase of the ENSO cycle. For each ensemble member of each model, a time series of the Niño 3.4 index was generated. We choose Niño 3.4 over Niño 3 or Niño 4 because we find that the Niño 3.4 region remains the center of action for ENSO variability both in observations and models. The Niño 3.4 region is the area 120°W-170°W and 5°N-5°S. The Niño 3 region misses a good deal of the Modoki ENSO-type variability, while Niño 4 misses a good deal of canonical ENSO-type variability. We define an ENSO event as a departure of the 5-month running mean Niño 3.4 index (computed over 5°S-5°N, 120°W-170°W) of greater than 0.5 K from the 2030-2069 climatology, with the linear trend removed from the 2030-2069 climatology before anomalies are calculated. Cold and warm events have the same definition, just with opposite sign. Anomalies in the historical data and the observational record are calculated relative to a 1966-2005 climatology, which is also detrended before anomalies are calculated. G1 output is analyzed absent detrending, as there is no trend in the data. We used skin temperature (T_S) anomalies rather than SST anomalies to build the Niño 3.4 time series for the BNU, IPSL and MPI, and models, because they were available on a regular grid. For the purpose of computing an anomaly-based index, the variable T_S is an excellent SST proxy variable, which is interchangeable with SST.

Mindful of the potentially weak detectability of changes in ENSO variability during the period of modeled geoengineering, we consider non-SST related measures of changes in the tropical Pacific. Might detectability of changes in ENSO be more evident from analyzing changes in non-SST based ENSO indices? First, we considered the Southern Oscillation Index (SOI), which is a standardized index based on the atmospheric pressure difference between Darwin, Australia, and Tahiti, because climate change does not produce SOI trends, except for a trivial increase as a result of increased water vapor concentration in a warmer world. Ideally, using SOI as a proxy for SST or T_S would allow inspection of the data absent the complications of dealing with a trend. Unfortunately SOI simulations show somewhat muted variability when compared with SST and T_S based indexes in the GCMs.

While the muted variability prevents the use of SOI as a proxy for the SST-based Niño3.4 index without redefining the requisite SOI threshholds for warm and cold ENSO events, we see that the ocean-atmosphere coupling, as reflected in the SOI, follows a realistic spatial structure. However, the spatial extent of the SST/SOI correlation is suppressed. Also, the magnitude of the correlation is realistic in the area that is the heart of ENSO variability. The area of prominent ENSO variability covers a smaller spatial area than that seen in observations, but the center of action is located in the same place as in observations. Additionally, the maximum values of the SST/SOI correlation in the historical models and the observations over the same period are both approximately the same ($r = 0.8$).

Figure 2 shows a spatial comparison between observed SOI-SST correlation and that modeled in a representative GISS historical run spanning the same time interval as the observations, and Figure 3 for corresponding time series. These examples are representative of the other CMIP5 GCMs used in this study. Although, the somewhat muted SOI variability prevented SOI analysis from being used in our study, the CMIP5 GeoMIP GCMs do produce plausible ocean-atmosphere coupling, albeit not extending as far eastward or away from the Equator as seen in observations.

We also explored changes in zonal surface winds and the possibility of a detectable weakening trend in the Walker Circulation and its relationship with ENSO. Vecchi (2006) identified a weakening of the ascending branch of the Walker circulation over equatorial southeast Asia. This change likely occurred as a result of increased precipitation. Precipitation increases much more slowly than humidity as a result of global warming. Therefore, the circulation weakens to maintain a balance of transport of water vapor out of the areas under the ascending branch that features extensive convection (Held and Soden 2006). These changes in the Walker Circulation were evident in the spatial pattern and trend of tropical Pacific SLP both in models that applied an anthropogenic change in radiative forcing over the historical period 1861-1990 and in the 21st century.

Unfortunately, considering changes in zonal wind and SLP is prevented by both the large inherent variability of SLP and zonal winds in the tropical Pacific, and the difficulty in deconvoluting the possible Walker Circulation weakening and ENSO change signals. In the observational record, 30-50 year changes in the Walker Circulation can occur concurrently with extended periods of more frequent ENSO warm events (Power and Smith, 2007). Hence, the observed weakening of the Walker Circulation during a period of somewhat more frequent ENSO warm events is not necessarily the result of a anthropogenically forced change in the Walker Circulation, but is instead convoluted by increased ENSO warm events and other inherent variability in the Tropical Pacific. The period of time required to robustly detect and attribute changes in the tropical Pacific Walker Circulation is found to be up to 130 years (Vecchi, 2006 and 2007) and no less than 60 years (Tokinaga et al., 2012). Because we cannot deconvolute the two signals in such a 40-year interval, we reject using zonal wind or SLP spatial pattern or trend as a proxy for ENSO. Additionally, as mentioned earlier, Cai et al. (2015) showed a robust weakening of the Walker Circulation under RCP8.5 counterintuitively co-occurs with a period of anomalously strong La Niña events as a result of increased heating over the Maritime continent. Therefore, while changes in atmospheric Walker Circulation over the tropical Pacific can be impacted by ENSO on decadal time scales, the changes may also be entirely unrelated to ENSO variability.

Lastly, based on the mechanism underlying the ocean component of ENSO, we conjecture that changes in thermocline depth or upwelling strength in the eastern and central Pacific might constitute a helpful, non-SST based indicator of changes in ENSO variability. However, given the somewhat difficult time CMIP5 models have simulating ENSO, it would be preferable to not consider these metrics, since they cannot be evaluated in the models with a long observational record.

Presently the National Oceanic and Atmospheric Administration Climate Prediction Center defines the climatological base period from which we calculate the departure from the current value and define an ENSO event as 1981-2010. We depart from this definition due to the robust warming trend in tropical SST in the Pacific both during the 1966-2005 comparison, and in 2030-2069 model runs, which show continued warming of the tropical Pacific. The detrended 40-year average produces a more realistic assessment of the base climate from which a particular ENSO event would evolve. This avoids the trap of identifying spurious ENSO events toward the end of the time series, which are really artifacts of the warming trend. Ideally, a climatological period in a rapidly changing climate would span less than 40 years. However, longer term natural trends in Pacific SST variability, including extended ENSO warm or cold periods, force use of a lengthy climatological base period to avoid comparing variability against a climatology that also includes that same variability.

The ENSO parameters evaluated are amplitude and frequency. El Niño amplitude is defined here as the peak anomaly value (in K) found during each El Niño event in the time series. La Niña amplitude is defined here as the mean negative peak anomaly value (in K) found during each ENSO event in the time series. Frequency is counted as the number of warm and cold events in each 40-year time slice. These parameters are chosen because ENSO frequency and amplitude have particular importance as global climate drivers.

The ENSO frequency and amplitude calculated in each ensemble member of each experiment (G1-G4) are compared 1) to other ensemble members from the same model for the same experiment, if available, 2) to their respective control runs, 3) to runs from other models with the same experimental design and 4) with different experimental designs, 5) to historical model runs, and 6) to observations. From this we seek to identify significant differences between model output from geoengineering scenarios, global warming, and historical runs compared to each other and to observations, as well as differences between models running the same G1-G4 experiment. Not only do we seek to analyze differences in ENSO amplitude and frequency between different scenarios, but we also seek to identify ENSO tendencies specific to particular models. The discussion below includes the successes and limitations of CMIP5 GCMs in depicting ENSO.

3 Results

3.1 Data excluded from final comparison

Although great strides have been made in the modern GCMs' ability to depict a realistic ENSO cycle, not all models are yet able to simulate a realistic ENSO cycle. Prior to further analysis, we applied two simple amplitude-based filters to exclude unreasonable ENSO time series data. The BNU-ESM output was excluded, because runs from more than one of several experiments found unrealistic 40-year ENSO time series where the substantial portion of warm and cold events maximum amplitude exceeded 3 K. Some BNU-ESM events exceeded 4 K, nearly a factor of 2 greater than the largest amplitude warm or cold events in the observational record. The model also produced nearly annual swings from implausibly strong warm events to cold events and back, implying an almost constant non-neutral state (Figure 4).

The MIROC-ESM and MIROC-ESM-CHEM output were also both excluded. Runs from more than one experiment in those models resulted in unrealistic 40-year ENSO time series without a single positive anomaly of more than 1 K of the Niño 3.4 Index. Negative anomalies were similarly suppressed in simulations from both models (Figure 5).

3.2 Analysis

We considered output from six GeoMIP participating GCMs: CanESM, CSIRO, GISS, HadGEM, IPSL and MPI (Tables 1 and 2). As a threshold question, we first evaluated agreement between the models used and the observational record for the metrics that we evaluate. To do this, we compared 150 years of observations with multiple 150-year ensembles of simulations from the six models. We also compared several selected 40-year periods from the historical simulations with observations from the same 40-year period. We find reasonable agreement between all aggregated model data and the observations during all historical time intervals. The strong agreement between simulations and observations includes the period after 1960, when the spatial and temporal density of Niño3.4 in situ observations increased dramatically. Specifically, the 1966-2005 observational record shows nine warm events, eight cold events, a maximum warm amplitude of 2.3 K, and maximum cold amplitude of 1.9 K. A multi-model ensemble of historical simulations of the same period shows 9.0 (±1.9) warm events, 8.5 (±1.7) cold events, maximum warm amplitude of 1.9 (±0.5) K and maximum cold amplitude of 1.7 (±0.6) K. Therefore, in this subset of GeoMIP CMIP5 models, the number of warm and cold events, warm amplitude, and cold amplitude are similar to observations, although some models outperform others across all experiments (See Figures 6 and 7).

Next, we seek to determine the variability of ENSO statistics in a 40-year period. Historical simulations for each of the six models start in 1850. The observational record dates back to 1857. We selected three periods, 1866-1905, 1916-1955, and 1966-2005 and built ensembles of these periods. Next, we compared ENSO amplitude and frequency in the simulations and observational record between the three periods. Although simulations of the 1966-2005 period show more ENSO warm events as compared to both the 1866-1905 and 1916-1955 periods (90% confidence), the periods are otherwise similar. Had the ENSO warm event frequency results reached 95% confidence, we would have performed a simple resampling technique to better detect if the significant result was robust, or reached by chance. However, given only 90% confidence, the relatively large number of comparisons made and the agreement between cold events frequency and amplitude in those same comparisons, the result is most likely by chance.

Since ENSO variability during the 150-year historical period, as assessed by comparing ensembles of simulations from distinct time intervals, does not contain statistically distinct periods in which ENSO frequency or amplitude is different from other periods, we turn to attempting to detect whether or not ENSO variability under a regime of geoengineering is distinct from ENSO variability under AGW. First, each experiment G1-G4 is matched with and compared to its respective control simulation, to the 1966-2005 historical period (during which observations are spatially and temporally dense) and to the full 150 years of available historical simulations. There are no statistically significant differences in ENSO frequency or amplitude between G1-G4 and their respective controls, or from the observations or historical simulations. As mentioned, only six CMIP5 GeoMIP models produce a reasonable ENSO. Therefore, we have a limited number of ensemble members available with which to perform comparisons. This limits our ability to detect differences and leads us to make some suggestions in the discussion section about future GeoMIP experiments.

The criteria used for aggregating experiments are provided in the methods section above. The purpose of aggregation is to construct the largest possible ensemble of simulations, which can then be compared. The inherent variability and usually subtle character of changes in ENSO compels the use of as much data as possible in order to filter out all of the internal variability and detect the ENSO change attributable to a particular forcing. Based on the aggregation criteria, we are only able to aggregate G1 with G2 and G3 with G4. In the G1/2 comparisons with 4xCO2, we see significantly more frequent (90%confidence) La Niña events - 8.32 ± 2.5 per 40 years for G1/2 and 6.71 ± 1.7 for 4xCO2. We also see more frequent (90% confidence) La Niña events in the G1/2 ensemble than in the 1% annual CO_2 increase ensembles - 8.32 ± 2.5 per 40 years for G1/2 and 7.37 ±1.7 for 1% annual CO_2 increase.

Since a number of comparisons were made and confidence in these results being significant is only 90%, we decided to apply a simple resampling technique to test the robustness of these results. First, we chose a sample, with replacement, from the G1/2 ensembles. Next, we chose a sample, with replacement, from 4xCO2. After calculating the median of each of the two samples, we repeated this process 500 times. Next, we calculated the differences between the medians in each of the 500 samples. This gives us an array of 500 integers, which are the differences between medians. The 25 highest and lowest differences between medians are stored, and the remaining 450 integers form a 90% confidence interval. Since the difference between the means in the G1/2 ensemble and the 4xCO2 ensemble fall within the 90% confidence interval of differences between means that we obtained via resampling, we conclude that the result showing increased ENSO frequency in G1/2 relative to control was likely obtained by chance. The same process was carried out for the G1/2 comparison with 1% annual CO_2 increase. Although the original comparison showed a significant result (90% confidence), after resampling, the difference between the G1/2 and 1% annual CO2 means was shown to be within the 90% confidence interval. Therefore, despite the initial presentation of results, based on a simple resampling technique, which allows for replacement, we find that there are no significant differences between G1/2 and the applicable controls.

Next, we turn to the final aggregated comparison, G3/4 and RCP4.5. Among all experiments and control simulations RCP4.5 simulations showed both the strongest and most frequent ENSO events. However, error bounds are large due to relatively small sample size (n=21) for RCP 4.5 and G3/4, and no difference in ENSO frequency or amplitude was detected in this comparison.

These results show the absence of a significant difference between GeoMIP experimental runs and AGW runs. However, the comparisons were limited by a number of factors. First, we excluded simulations from several GeoMIP participating modeling groups due to an implausible ENSO in the models. Second, current generation GeoMIP runs may not be long enough to detect changes in ENSO. Third, the signal-to-noise ratio in RCP4.5 is rather low. A geoengineering experiment that seeks to offset a stronger forcing may improve our chances detecting potential changes in future ENSO. These issues will be covered in greater detail in the discussion section.

3.3 Comparisons between models

There was significant disagreement between two of the models. The CanESM simulated 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 chose to include both models as they both simulate plausible ENSO events and variability that is in line with some portion of the observational record. This is unsurprising given the wide range of ENSO predictions among CMIP5 models. One of the features most readily apparent across all models was the confinement of the most robust coupling between the ocean and the atmosphere too close to the Equator and not extending as far eastward as in observations. The ENSO center of action was over a small area in the central Pacific, whereas the center of action extended further into the equatorial eastern extent of the basin in the observational record (Figure 2).

Next we make model vs. model comparisons - comparing all ensemble members of runs from each model against each other. Figures 8 and 9 show ENSO event amplitude and frequency simulated by each model. Of the models not excluded, the CanESM results diverged far more than the other models from the overall mean on all four parameters evaluated. CanESM depicts ENSO warm and cold events that are both more frequent and stronger than that documented in the Kaplan SST observational record. The CSIRO model depicts the lowest number of both cold and warm events, as well as event amplitudes that are on the low end. The GISS, HadGEM, IPSL and MPI models are not in close agreement on all four parameters, but they agree more than the CSIRO and CanESM. The best agreement between models existed between the GISS, HadGEM, MPI and IPSL, which agreed on all but cold event amplitude. Had we excluded the CanESM (most frequent and strongest ENSO events) and CSIRO models (least frequent ENSO events), agreement between the remaining four models would have been reasonable. However, both the CanESM and CSIRO produce a physically plausible ENSO. During especially active periods, ENSO has behaved in line with the CanESM results. During relatively quiescent periods in the observational record, the CSIRO results are not out of line with observations.

4 Discussion

We conclude that ENSO event frequency or amplitude in a geoengineered world will not differ from the observational record, historical model runs, or under global warming. However, this conclusion comes with a number of very strong caveats, which result from the relatively brief simulation length and the considerable model spread. Simulations that depict 100 years or more of geoengineering would likely be amenable to more detailed analysis. During a 40-year experiment, it becomes nearly impossible to devise a non-SST based ENSO index. It is also difficult to detect a subtle signal within an ensemble of ENSO time series, as the Niño 3.4 index exhibits a great deal of inherent variability. Precipitation and surface winds are exceptionally noisy, and the observational data set of SST is far more solid than that for wind or precipitation. Also, over periods of less than 100 years, it becomes difficult to attribute changes in tropical Pacific zonal wind to long-term trends in ENSO, or to the weakening of the Walker circulation. Further, a weakening of the Walker circulation can – counter intuitively - occur concurrently with a period of frequent extreme La Niña events (Cai et al., 2015).

The next generation of GeoMIP experiments that will be part of CMIP6 will extend G1 simulations to 100 years and also provide simulations in which a more extreme AGW scenario – RCP6 or RCP8.5 – is offset by constant stratospheric sulfate injections or solar dimming. These new experimental designs result from the need to understand extreme precipitation and temperature events, changes in regional climate and to examine modes of internal variability (Kravitz et al., 2015).

A potential future GeoMIP experiment could proceed as follows. Compare the 100 year 4xCO2 scenario with 100 years of preindustrial control and G1 (G1extended; Kravitz et al., 2015). Given the length of this simulation and the large differences between 4xCO2, G1 and the preindustrial control, this would likely be the place to start. Should a signal be detectable in G1extended, the G6solar or G6sulfur experiments, which will run for at least 80 years without termination, offsets either RCP6.0 or RCP8.5 with continuous sulfate injections or solar dimming, should be evaluated next. The presence of a signal in the steady-state G1extended experiment would be more likely than in the transient G6 experiments. However, the G6solar and G6sulfur scenarios are far more like real climate, and although detecting a difference in G6 would likely be more difficult than in G1extended, a G6 signal would speak more directly to how the tropical Pacific might evolve under plausible future geoengineering scenarios. Even after the next generation of GeoMIP simulations has been released, the ability for those experiments to potentially detect future changes in ENSO variability will still be limited by how well each model performs.

Returning to the current experiment, to buttress the robustness of the negative result, we offer a discussion of why the participating models behaved as they did. Substantial work has been done to determine why many models have difficulty simulating an ENSO that matches observations in amplitude and frequency. The Climate and Ocean: Variability, Predictability and Change (CLIVAR) ENSO working group's analysis of process-based variables in CMIP 3/5 – which quantifies a GCM's ability to simulate key ENSO processes – is underway (Guilyardi, 2012). That work is more squarely focused on determining exactly how the simulation of both ENSO events and the underlying ENSO processes can be improved in GCMs. However, we can evaluate the ENSO behavior we saw in this geoengineering experiment in the context of the process based variable analysis conducted by CLIVAR.

In the CSIRO model we see lower variability than in other models and slightly dampened amplitude. We also noticed that the center of action in terms of ENSO variability is shifted somewhat westward. This is in line with Bellenger et al. (2012), who found that the standard deviation of SST in Niño4 is far greater than that seen in Niño3. In the GISS and other models ocean-atmosphere coupling as reflected by the SST-SOI correlation was very robust in Niño3.4 at the Equator, but somewhat muted elsewhere. The area in which coupling between the ocean and atmosphere is most robust did not extend throughout Niño3.4, and therefore areas farther away from the Equator and further east were not contributing as much to the Niño index. Therefore we see slightly lower Niño3.4 amplitude values in GISS and HadGEM than we do in other models. The presence of this tendency is bolstered by Guilyardi (2012), who showed that the standard deviation of SST in both Niño3 and Niño4 is toward the low end of the range for CMIP5 models. In our experiment, the IPSL model performs very well in terms of a realistic ENSO frequency and amplitude. This tendency is also reinforced by Guilyardi (2012), who showed that the root mean square error is among the lowest of CMIP5 models for both SST and surface wind stress. Even though different models struggle with various ENSO processes, the tendencies of each model are relatively well understood and each model generated an ENSO that is plausible, albeit not necessary exactly fit to what is seen in the observational record.

However, the story for the models excluded is quite different. A representative BNU G1 run showed a cold event with a maximum anomaly of –4.8 K during a relatively short duration La Niña event. Similar non-geoengineering runs revealed equally extreme amplitude and short lasting events. The observational record fails to hint that such an event has occurred, and none of the other models simulate anything like such an event. Future work could explore why simulated BNU ENSO events are so strong, short and frequent relative to observations and other models.

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.

The absence of a significant result, showing differences in ENSO variability between geoengineering and AGW may well be a valid finding. However, detectability of changes over a relatively short interval of time is only possible if the differences seen between groups are extremely large. To evaluate more subtle differences, longer simulations are likely required. Additionally, ENSO is complicated, and GCMs are still developing in terms of their ability to produce an extremely accurate ENSO. However, work is currently underway to determine what processes in each particular model lead to a less than ideal depiction of ENSO. The results of this analysis will likely be factored into CMIP6, and it will be interesting to examine ENSO behavior in CMIP6.

Returning to geoengineering, the next generation of GeoMIP experiments will produce longer simulations, with more robust forcing in the case of G1 extended. It is imperative that we understand potential changes in extreme event frequency under geoengineering and potential changes in modes of internal variability. Any future contemplation of large-scale deployment of geoengineering would require confidence in model predictions about potential changes in natural variability and the frequency and nature of extreme events.

Acknowledgments. We thank all the climate modeling groups for conducting the GeoMIP simulations and making their output available. We thank Ben Kravitz for his overall leadership of GeoMIP. We also thank three anonymous reviewers for their thoughtful and constructive feedback. This work is supported by NSF grants AGS-1157525 and GEO-1240507.

References

Arora, V.K., Scinocca, J.F., Boer, G.J., Christian, J.R., Denman, K.L., Flato, G.M., Kharin, V.V., Lee, W.G. and Merryfield, W.J,: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophys.Res. Lett., 38, L05805, 2011.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M. and Vialard, J.: ENSO representation in climate models: from CMIP3 to CMIP5. Clim. Dyn., 42, 1999-2018, 2013.

Bjerknes, J.: Atmospheric teleconnections from the equatorial Pacific. Mon. Wea. Rev., 97, 163–172. 1969.

Caldeira, K., Bala, G. and Cao, L.: The science of geoengineering. Ann, Rev, Earth Planetary Sci., 41, 231-256. 2013.

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.: Increasing frequency of extreme El Niño events due to greenhouse warming. Nature Climate Change, 4, 111-116. 2014.

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.

Clement, A.C., Seager, R., Cane, M.A. and Zebiak, S.E.: An ocean dynamical thermostat. J. Climate, 9, 2190-2196. 1996.

Collins, W.J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C.D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A. and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051–1075. 2011.

Crutzen, P.: Albedo enhancement by stratospheric sulfur injections: A contribution to solve a policy dilemma? Climatic Change, 77, 211–219, 2006.

Dai, Y., Zeng X.B., Dickinson R.E., Baker I., Bonan, G.B., Bosilovich, M.G., Denning A.S., Dirmeyer P.A., Houser P.R., Niu G.Y., Oleson K.W., Schlosser C.A., Yang Z.L.: The Common Land Model (CLM), Bull. Am. Meteorol. Soc., 84, 1013–1023. 2003.

Dufresne, J.L., Foujols, M,.A., Denvil, S., Caubel, A., Marti, O., Aumont, O, Balkanski, Y., Bekki, S , Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F , Codron, F , Cozic, A , Cugnet, D., de Noblet, N., Duvel, J,P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Clim. Dynamics , 40, 2123-2165, 2013.

Emile-Geay, J., Seager, R., Cane, M.A., Cook, E.R., Haug, G.H.: Volcanoes and ENSO over the past millennium, J. Climate, 21(13), 3134–3148. 2007.

Giorgetta, M. A., Jungclaus, J. H., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W. A., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J. and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. J. Advances in Modeling Earth Systems, 5, 572-597. 2013.

Guilyardi E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Oldenborgh, G.J.V., Stockdale, T.: Understanding El Niño in ocean-atmosphere general circulation models : progress and challenges. Bull. Amer. Met. Soc., 90, 325-340. 2009.

Guilyardi E., Bellenger, H., Collins, M., Ferrett, S., Cai, W. and Wittenberg, A.: A first look at ENSO in CMIP5. CLIVAR Exchanges, 58, 29-32. 2012.

Held, I.M., and Soden, B.J.: Robust responses of the hydrological cycle to global warming, J. Climate, 19, 5686–5699, 2006.

Held, I.M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. and Vallis, G.K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. J. Climate, 23, 2418–2427. 2010.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Jones, A., Haywood, J., Boucher, O., Kravitz, B. and Robock, A: Geoengineering by stratospheric SO_2 injection: Results from the Met Office HadGEM2 climate model and comparison with the Goddard Institute for Space Studies ModelE. Atmos. Chem. Phys., 10, 5999-6006. 2010.

Kirtman, B.P., and Schopf, P.S.: Decadal variability in ENSO predictability and prediction. J. Climate, 11, 2804–2822, 1998.

Kravitz, B., Robock, A., Boucher, O., Schmidt, H., Taylor, K.E., Stenchikov, G. and Schulz, M.: The geoengineering model intercomparison project (GeoMIP), Atm. Sci. Lett., 12, 162-167. 2011.

Maher, N., McGregor, S., England, M.H., Gupta, A.S.: Effects of volcanism on tropical variability. Geophys. Res. Lett., in press.

Mann, M.E., Cane, M.A., Zebiak, S.E., Clement, A.: Volcanic and Solar Forcing of the Tropical Pacific Over the Past 1000 Years, J. Climate, 18, 447-456. 2005.

Marcott, S.A., Shakun, J.D., Clark, P.U. and Mix, A.C.: A reconstruction of regional and global temperature for the past 11,300 years. Science 339 (6124): 1198–1201. 2013.

McPhaden, M.L., Zebiak, S.E. and Glantz, M.H.: ENSO as an integrating concept in earth science. Science, 314, 1739–1745. 2006.

Phipps, S. J., Rotstayn, L.D., Gordon, H.B., Roberts, J.L., Hirst, A.C. and Budd, W.F.: The CSIRO Mk3L climate system model version 1.0 - Part 1: Description and evaluation, Geoscientific Model Development, 4, 483-509. 2011.

Phipps, S. J., Rotstayn, L.D., Gordon, H.B., Roberts, J.L., Hirst, A.C. and Budd, W.F.: The CSIRO Mk3L climate system model version 1.0 - Part 2: response to external forcings, Geoscientific Model Development, 5, 649-682. 2012.

Power, S.B. and Smith, I.N.: Weakening of the Walker Circulation and apparent dominance of El Niño both reach record levels, but has ENSO really changed? Geophys. Res. Lett., 34, L18702, 2007.

Robock, A.: Whither geoengineering? Science, 320, 1166-1167. 2008.

Robock, A., Marquardt, A., Kravitz, B., and Stenchikov, G.: The benefits, risks, and costs of stratospheric geoengineering. Geophys Res. Lett., 36, L19703, doi:10.1029/2009GL039209. 2009.

Schmidt, G.A., Ruedy, R., Hansen, J.E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Del Genio, A., Faluvegi, G., Friend, A.D. Hall, T.M., Hu, Y., Kelley, M., Kiang, N.Y., Koch, D., Lacis, A.A., Lerner, J., Lo, K.K., Miller, R.L., Nazarenko, L., Oinas, V., Perlwitz, J.P., Perlwitz, J., Rind, D., Romanou, A., Russell, G.L., Sato, M., Shindell, D.T., Stone, P.H., Sun, S., Tausnev, N., Thresher, D. and Yao, M.S.: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite and reanalysis data, J. Clim., 19, 153–192. 2006.

Siegenthaler, U., Stocker, T. F., Monnin, E., Luthi, D., Schwander J., Stauffer, B., Raynaud, D., Barnola, J. M., Fischer, H., Masson-Delmotte, V., and Jouzel, J.: Stable carbon cycle-climate relationship during the late Pleistocene, Science, 310, 1313–1317. 2005.

Taylor K.E., Stouffer R.J., Meehl, G.A .: An overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., 93, 485-498. doi:10.1175/BAMS-D-11-00094.1. 2012

Tilmes, S., Fasullo, J., Lamarque, J.F., Marsh, D.R., Mills, M., Alterskjær, K., Muri, H., Kristjánsson, J.E., Boucher, O., Schulz, M., Cole, J.N.S., Curry, C.L., Jones, A., Haywood, J., Irvine, P.J., Ji, D., Moore, J.C., Karam, D.B., Kravitz, B., Rasch, P.J., Singh, B., Yoon, J.-H., Niemeier, U., Schmidt, H., Robock, A., Yang, S. and Watanabe, S.: The hydrological impact of geoengineering in the Geoengineering Model Intercomparison Project (GeoMIP), J. Geophys. Res. Atmos., 118, 11,036-11,058. 2013.

Tokinaga, H., Xie, S.P., Deser, C., Kosaka, A., Okumura, Y.M.: Slowdown of the Walker Circulation driven by tropical Indo-Pacific warming," Nature, 491, 439-443 , 2012.

Trenberth, K. E., and Hoar, T.J.: El Niño and climate change. Geophys. Res. Lett., 24, 3057–3060. 1997.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev., 4, 845–872, 2011.

Vecchi, G.A., Soden, B.J., Wittenberg, A.T., Held, I.M., Leetmaa, A. and Harrison, M.J.: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing, Nature, 7089, 73-76. 2006.

Vecchi, G.A., and Soden, B.J. Global warming and the weakening of the tropical circulation, J. Climate, 20, 4316–4340. 2007.

Yeh S.W., Ham, Y.-G., and Lee, J.-Y.: Changes in the tropical Pacific SST trend from CMIP3 to CMIP5 and its implication of ENSO. J Climate, 25, 7764–7771. 2012.

Zebiak, S. E., Cane M.L.: A model El Niño/Southern Oscillation. Mon. Wea. Rev., 115, 2262–2278. 1987.

Table 1. The names of the climate models used in this study, with short names and references. Asterisks indicate that the models were excluded from comparison due to unrealistic ENSO variability.

Model	Model short name	Reference
*BNU-ESM	BNU	Dai et al. (2003)
CanESM2	CanESM	Arora et al. (2011)
CSIRO-Mk3L	CSIRO	Phipps et al. (2011,2012)
GISS-E2-R	GISS	Schmidt et al. (2006)
HadGEM2-ES	HadGEM	Collins et al. (2011)
IPSL-CM5A-LR	IPSL	Dufresne et al. (2012)
*MIROC-ESM	MIROC	Watanabe et al. (2011)
*MIROC-ESM-CHEM	MIROC-C	Watanabe et al. (2011)
MPI-ESM-LR	MPI	Giorgetta et al. (2012)

Table 2. Models analyzed in each experiment. Asterisks indicate that the models were excluded from comparison due to unrealistic ENSO variability. The number of ensemble members for each experiment is given in parenthesis after the model name.

a. Models in G1	**b. Models in G2**
*BNU-ESM (2) CanESM2 (3) CSIRO-Mk3L (3) GISS-E2- (3) HadGEM2-ES (1) IPSL-CM5A-LR (1) *MIROC-ESM (1) MPI-ESM-LR (1)	*BNU-ESM (3) CanESM2 (3) CSIRO-Mk3L (3) GISS-E2-R] (3) HadGEM2-ES (3) IPSL-CM5A-LR (1) *MIROC-ESM (1) MPI-ESM-LR (1)
c. Models in G3	**d. Models in G4**
*BNU-ESM (1) GISS-E2-R (3) HadGEM2-ES (2) IPSL-CM5A-LR (1) MPI-ESM-LR (3)	*BNU-ESM (2) CanESM2 (3) CSIRO-Mk3L (3) GISS-E2-R (3) HadGEM2-ES (1) IPSL-CM5A-LR (1) *MIROC-ESM (1) *MIROC-ESM-CHEM (1) MPI-ESM-LR (1)

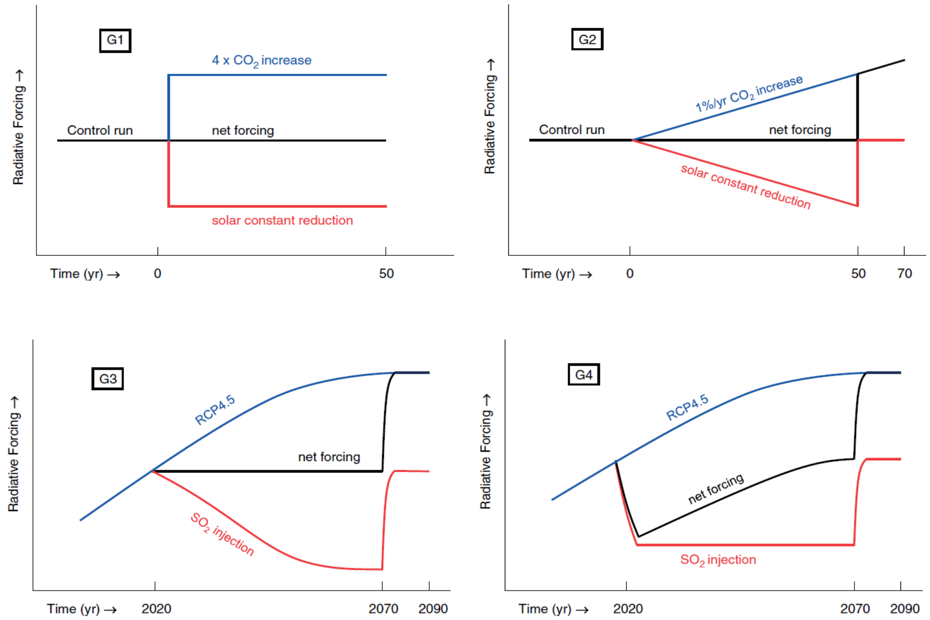


Figure 1. GeoMIP G1-G4 experiment designs. Figures 1-4 from Kravitz et al. (2011).

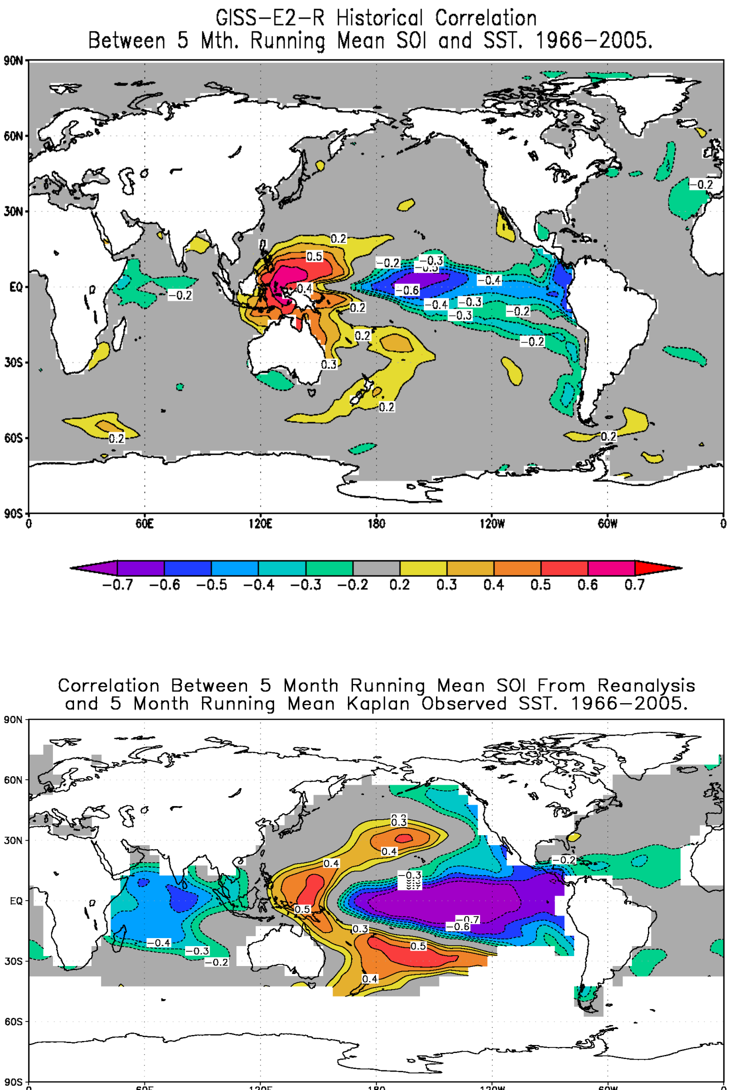


Figure 2. Top panel shows spatial correlation between GISS historical sea surface temperature (SST) and the Southern Oscillation Index (SOI). The area of strong negative correlation is confined to a small region in the central Pacific, relative to the broad area of strong negative correlation in the observations in the bottom panel, which shows the spatial correlation between NCEP SLP reanalysis and the Kaplan SST observations data set.

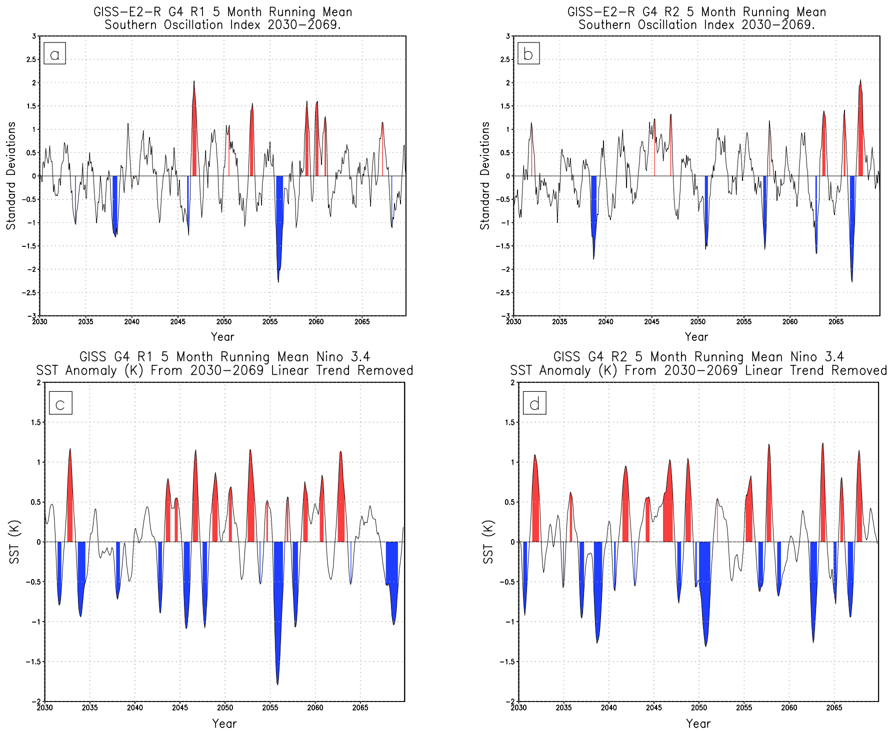


Figure 3. Time series of normalized Southern Oscillation Index (SOI) for (a) GISS G4 Run 1 and (b) GISS G4 Run 2. In the context of SOI, ENSO events are defined as departures of 0.5 standard deviations from zero. SOI warm events are highlighted in red, while cold events are highlighted in blue. No highlight is applied during an ENSO neutral phase. Time series of SST in Niño 3.4 region for (c) GISS G4 Run 1 and (d) GISS G4 Run 2. The SST-based index in the bottom panel depicts more realistic ENSO variability, and therefore SOI is not used as an SST proxy.

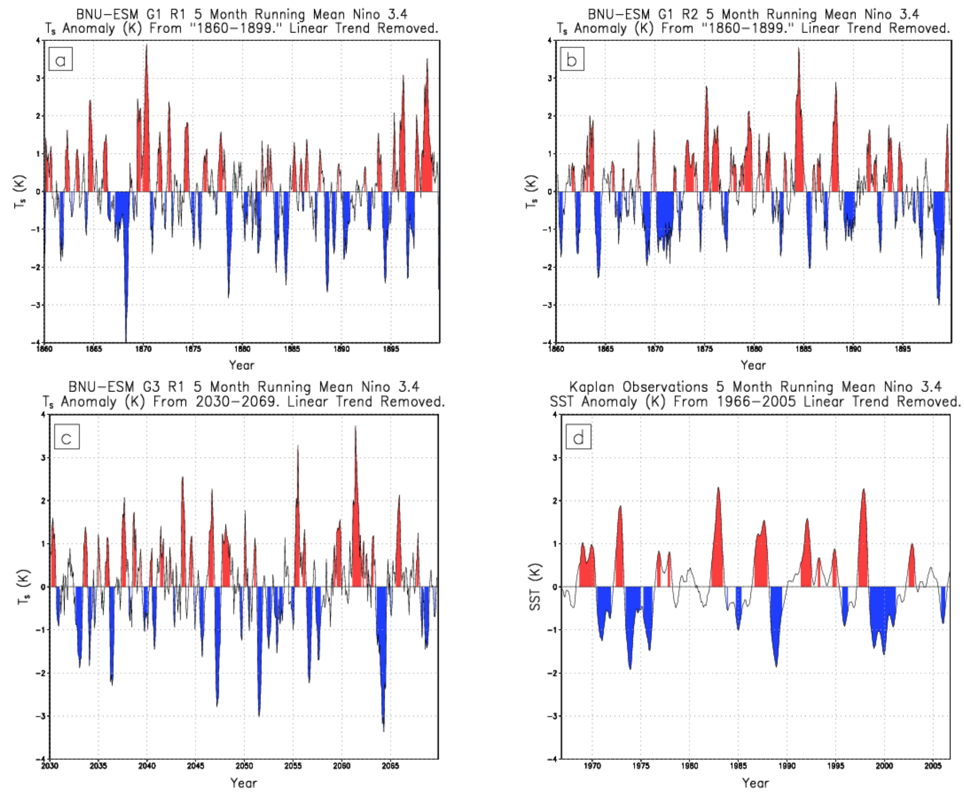


Figure 4. Time series of Niño 3.4 anomalies from three experimental runs, (a) G1 run 1, (b) G1 run 2, and (c) G3 run 1, of the BNU-ESM model compared to observations (d). Red coloring indicates ENSO warm events, while blue shading indicates ENSO cold events. The model is excluded due to unrealistic variability and amplitude.

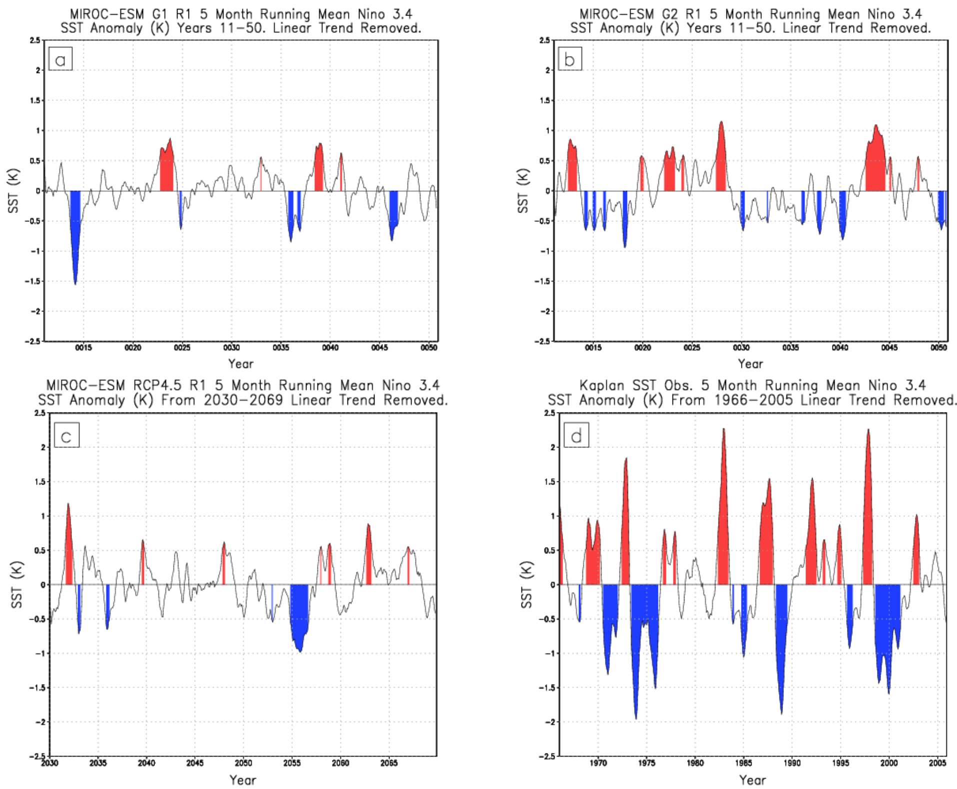
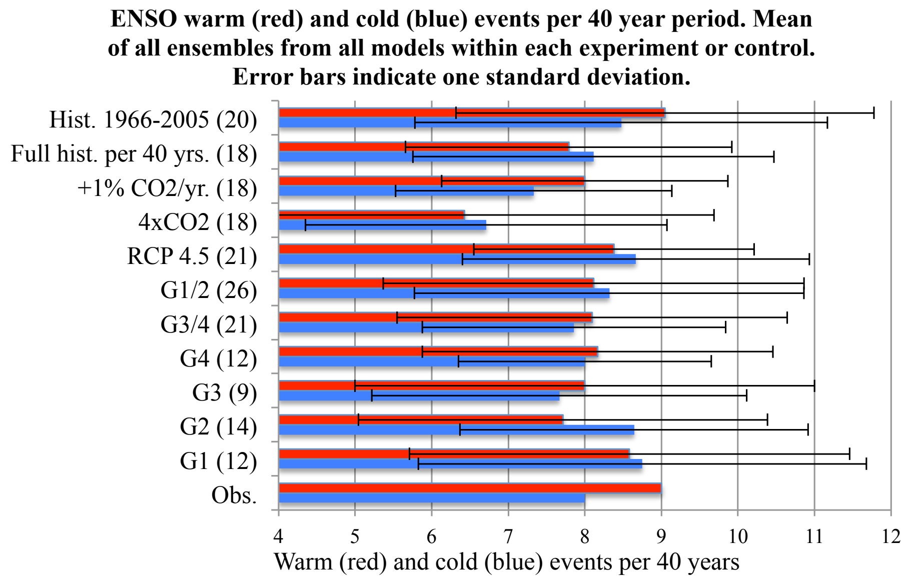
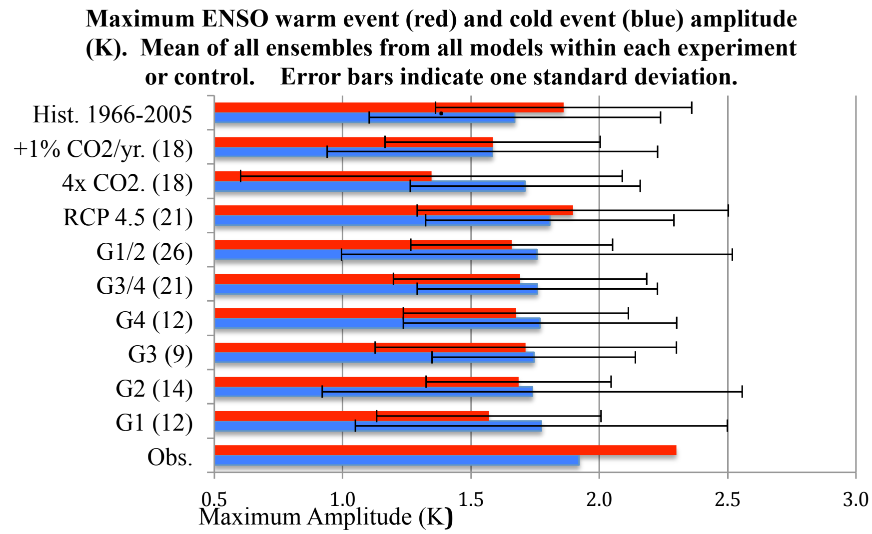


Figure 5. Niño 3.4 anomalies for MIROC-ESM. Time series of (a) G1, (b) G2 and (c) RCP4.5 from each model all show significantly muted variability and amplitude compared to (d) observations, with few, if any, warm events exceeding a 1 K anomaly. All other MIROC family experiments showed the same muted variability. Cold event amplitude is essentially muted, with no 1 K or greater departures. The inability of the MIROC-ESM to depict a plausible ENSO cycle is also seen in the MIROC-ESM-CHEM. Therefore, both sets of model output were excluded.



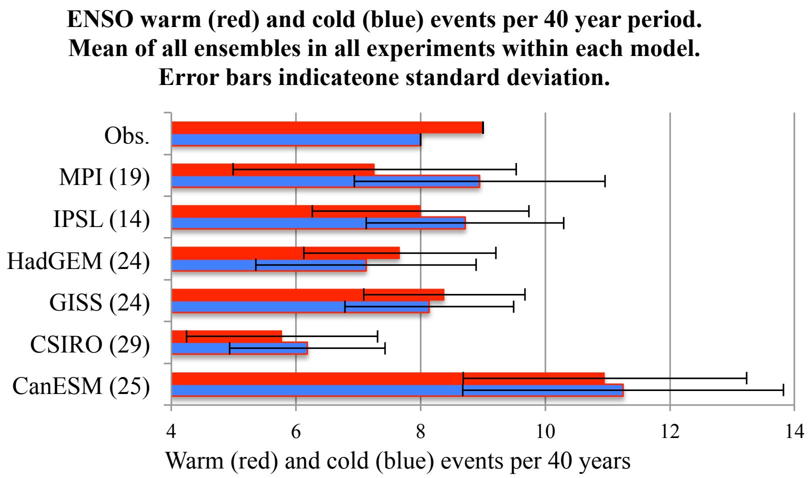
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2 /yr. (18)	1850-2005 /40 yr. (18)	1966-2005 (18)
Warm events	9.00	8.58	7.71	8.00	8.17	8.10	8.12	8.38	6.43	8.00	7.79	9.05
Cold events	8.00	8.75	8.64	7.67	8.00	7.86	8.32	8.67	6.71	7.33	8.11	8.48

Figure 6. Number of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G1-G4, RCP4.5, 1% annual CO_2 increase and 4x CO2, and 1966-2005 for historical simulations and observations for the CanESM, CSIRO, GISS, HadGEM, IPSL and MPI models. The full historical record spans 1850-2005 and the number of events reported for this period is the per 40 year frequency of warm or cold events in this full record. Values in parentheses are the number of ensemble members for each experiment or family of experiments. Error bars represent plus or minus one standard deviation of ENSO events relative to the experiment mean. A table of values is provided under the graph.



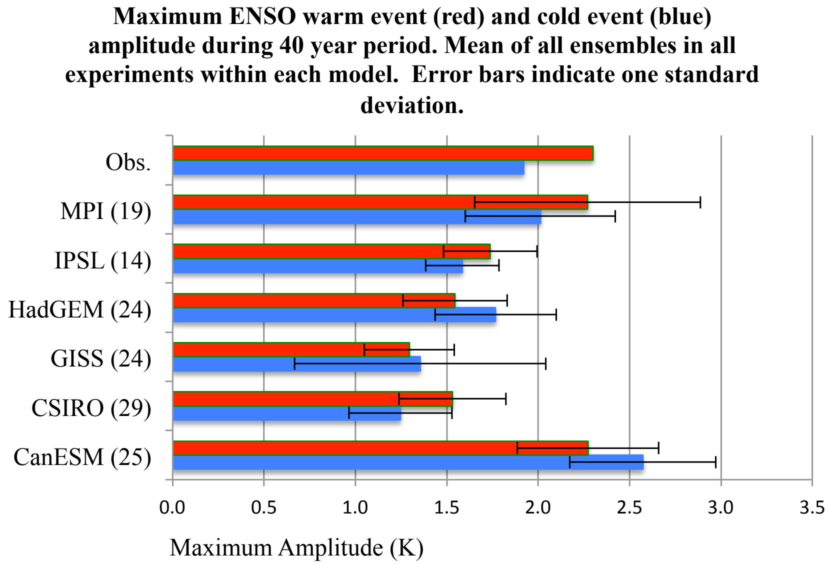
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2/yr. (18)	1966-2005 (18)
Warm amp.	2.30	1.57	1.69	1.71	1.68	1.69	1.66	1.90	1.35	1.58	1.86
Cold amp.	1.92	1.77	1.74	1.74	1.77	1.76	1.76	1.81	1.71	1.58	1.67

Figure 7. Maximum amplitude (K) of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G1-G4, RCP4.5, 1% annual CO_2 increase and 4x CO2, and 1966-2005 for historical simulations and observations . Values in parenthesis following y-axis (model name) labels indicate the number of ensemble members, inclusive of all experiment designs, run by the particular model. Error bars show plus or minus one standard deviation relative to the model mean. A table of values is provided under the graph.



	Obs.	CanESM (25)	CSIRO (25)	GISS (24)	HadGEM (24)	IPSL (14)	MPI (19)
Warm Events	9.00	10.96	5.78	8.38	7.67	8.00	7.26
Cold Events	8.00	11.25	6.19	8.14	7.13	8.71	8.95

Figure 8. Number of ENSO warm (red) or cold (blue) events observed or simulated in the applicable 40-year comparison period for the CanESM, CSIRO, GISS, HadGEM, IPSL and MPI models. Values in parentheses are the number of ensemble members for each experiment or family of experiments. Error bars represent plus or minus one standard deviation of ENSO events relative to the experiment mean. A table of values is provided under the graph.



	Obs.	CanESM (25)	CSIRO (25)	GISS (24)	HadGEM (24)	IPSL (14)	MPI (19)
Warm amp.	2.30	2.27	1.53	1.29	1.54	1.74	2.27
Cold amp.	1.92	2.57	1.25	1.35	1.77	1.58	2.01

Figure 9. Maximum amplitude (K) of ENSO warm (red) or cold (blue) events observed or simulated in the applicable 40-year comparison period. Values in parenthesis following y-axis (model name) labels indicate the number of ensemble members, inclusive of all experiment designs, run by the particular model. Error bars show plus or minus one standard deviation relative to the model mean. A table of values is provided under the graph.

des 7/15/15 11:24 PM
Deleted: ... [1]