Response to Referee 1 Comments on "Stratospheric geoengineering impacts on El Niño/Southern Oscillation," by C. J. Gabriel and A. Robock.

Referee comments are in black. Responses are in blue.

Editor response:

1) The concern is that the revised version lacks substantial results of interest.

2) Expand presentation of "results on changes in the mean and interannual variability of Niño 3.4 temperatures" as they "would be of interest.

3) Address referee 1's new comments in your responses.

Referee 1:

1) Firstly, my previous comment 1) still stands. The authors have confirmed what I suspected, that the ENSO changes cannot be assessed with this dataset yet ignored the thrust of this criticism, i.e. what is the point of dedicating the entire results section to displaying a null result?

We have added Figure 10 showing the linear trend in the Niño 3.4 index for each 40 year period (in K decade-1) in each simulation. This shows the temperature trend in this key tropical Pacific "center of action" region during each experimental or control period. The results provide a measure of tropical Pacific changes over a relatively long period of time. We have also added Table 3, which shows the difference in linear trend between each experiment and control. Please also see section **3.4 Long term behavior of Niño 3.4 under geoengineering** for presentation of those results as well as discussion of those results on pp. 30-31, lines 711-732.

2) I suggested that other analysis be pursued in place of the many figures that show no change (Figures 3-9). Reviewer 2 made a similar suggestion that alternative analyses be made but I believe the authors misinterpreted our suggestions. Rather than suggesting that the authors use different analyses to assess ENSO, I was suggesting that the authors do different analyses of the Pacific Ocean in general, i.e. ones where concrete results would be forthcoming.

Please see above and Figure 10, which shows how the tropical Pacific will evolve under geoengineering relative to under global warming. Also, we have calculated the standard deviation of the sea surface temperature anomaly (SSTA) over Niño 3.4. This is a metric recommended by the CLIVAR Pacific panel to assess strength of ENSO events. Also, since ENSO is the dominant source of interannual variability in this region, the standard deviation of SSTA is a metric of interannual variability in the region. Importantly, standard deviation of SSTA is an alternative measure for ENSO strength independent of our amplitude and frequency based analysis. We find that the standard deviations of SSTA in Niño 3.4 are in strong agreement for all experiments G1-4 with values near 0.7 K for each. We also investigate SSTA for each model and see that the standard deviation of Niño 3.4 SSTA varies considerably from model to model. Figures 1 (a) and (b) of Bellenger et al. (2013) show SSTA for Niño 3 and Niño 4 for 32 CMIP5 models, including all of the models we use in this experiment. Our calculations were substantially in line with that figure and displaying that additional data in the form of a figure would not be new.

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.

3) The authors' response that the SST response to SRM is well studied does not tally with my understanding of the literature; few studies assess the oceanic response to SRM, most do so only very briefly, and this is an area deserving of greater attention.

We were referring to the literature available related to the ocean response to volcanic eruptions. The reviewer is correct that there is little research regarding the impacts of SRM on the ocean and that SRM impacts on the ocean require greater attention. We have clarified our statement regarding studies about the oceanic response to SRM by adding the second sentence below (*italicized*).

“No modeling study has ever evaluated the impact of long term solar dimming or continuous stratospheric sulfate injections on ENSO. *Additionally, little work has been done to assess the oceanic response to SRM.*” (p. 6, line 146 added).

Also, we point to Figure 10 in the revised manuscript, which shows the different impacts of G1-G4 as well as global warming and historical scenarios on tropical Pacific SSTs. It is of particular interest that the two physically plausible geoengineering scenarios (G3/4) that seek to offset RCP 4.5 only offset a modest portion of the linear trend in warming seen in RCP 4.5.

4) Secondly, in my previous points 2), 7) and 9) I noted that more should be done to assess and quantify what would be needed to detect changes in ENSO.

We agree. See our response to question 5 for a detailed explanation of how we assess and quantify the sensitivity of our detection methods.

5) To make a useful contribution to this issue the authors should make a quantitative assessment of what would be required to detect changes in ENSO of various magnitudes. In other words the article ought to have addressed a specific quantitative question along these lines: How long (or how many 100 year ensemble members) ought it to take to detect a 10% increase in ENSO frequency or average intensity in one model?

We agree that a quantitative assessment of what would be required to detect changes in ENSO variability would be useful. We did not obtain robust significant results in our geoengineering comparisons, but we were able to determine and report the minimum detectability threshold for our comparisons. Rather than framing the answer in terms of 100 or 100+ year simulations that were not available for this experiment, we provide the minimum detectability threshold in the following terms: what percentage increase in event amplitude or frequency would be detectable in comparisons between groups of 11 40-year ensembles within a particular model. We performed this analysis for CanESM, CSIRO and GISS models. (See below). We report the minimum change in value of each ENSO property that would have to exist in the data for our method (described in our answer to item 9 below) to detect it. (Please see Discussion pp. 26-27, lines 620-646.) Although we do not have 100+ year geoengineering simulations, we would anticipate that applying these same methods to analyzing longer simulations would yield lower detectability thresholds.

6) Are the results different in different models?

The results are similar across models for event amplitude, but there is some variability in the detection threshold for event frequency. (See quoted text in our response to question 5).

7) How do these figures relate to findings of observational studies?

Satellite observations of Niño 3.4 only go back to 1980. In situ observations increased by an order of magnitude and became sufficiently dense during the 1960s. While observations date back to around 1857, we are more confident in the more recent portion of the observational record. However, studies that have extended the ENSO record back as far as 700 years have revealed distinct periods of muted ENSO variability (late 17th century) and enhanced ENSO variability (late 20th century). Many of these studies have detected only very substantial changes in ENSO frequency and intensity, documenting changes on the order of 30%-40% (Gergis and Fowler 2006; Li 2013).

Gergis J.L., Fowler, A.M.: How unusual was late twentieth century El Nino Southern Oscillation? Adv. Geosci.. 6, 173–79. 2006.

Li, J., Xie, S.-P., Cook E. R., Morales, M., Christie, D., Johnson, N., Chen, F., D'Arrigo, R., Fowler, A., Gou, X. and Fang, K.: El Niño modulations over the past seven centuries. Nature Climate Change, 3, 822–826. 2013.

8) In comment 8) I noted that there are longer pre-industrial simulations that the authors could use that could form the basis of such an assessment. This remains the case – all CMIP5 models ran extended pre-industrial spin-up simulations

We have already shown with the runs we have the distinct ENSO responses of the different models. CanESM features frequent large ENSO events. GISS and HadGEM feature frequency that is about the same as observations and slightly muted variability, whilst IPSL and MPI are very close. Longer runs could investigate longer term variability of ENSO, but that is beyond the scope of this paper.

9) I don't believe the authors have adequately addressed the concern about treating results from models with different ENSO behavior as though they form part of the same set for their statistical tests. That Cai et al. 2015 used this method does not seem justification enough.

Our decision to create a multi-model ensemble composed of simulations from different models was initially motivated by the methods employed by Cai et al. (2015) among many other studies. However, the above comment has given us the opportunity to carefully explain our methodology in creating and comparing our large multi-model ensembles. Our results come from two sample t-tests assuming unequal variance. To validly use a t-test, several conditions must be met. In our case, we need to determine whether or not the members in the large ensembles we create constitute normal distributions before we apply parametric statistics (a t-test in this case). To determine the normality of the distribution, we evaluate the parameters used in the widely-used formal D'Agostino's K2 test for goodness of fit. Based on the analysis described in the Methods section, we determined that the goodness of fit is sufficient to treat each experiment as a normal distribution for the purposes of applying parametric statistics. Please see pp. 10-11, lines 253-275).

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.

D'Agostino, R.B., Belanger A., Ralph B.: A suggestion for using powerful and informative tests of normality. The American Statistician 44, 316–321. 1990.

Decarlo, L.T.: On the meaning and use of kurtosis. Psychological Methods 2, 292-307. 1997.

10) The approach used to generate the 90% confidence statements on page 31, line 1024 is unexplained. Are the authors using inter-member variance in their grand ensemble? Is it then appropriate to assume these are independent measures of some underlying sample?

The reviewer is correct that the approach should be explained. As described in our answer to question 9, we were able to test and determine that each sample could be treated as a normal distribution. We then matched each experimental sample with the appropriate control sample and compared the means using a two sample t-test assuming unequal variance. This statistical test is most appropriate for the data. It is robust for small sample sizes so long as the data is not sufficiently skewed. (Please see our answer to question 9 for an explanation of our test for goodness of fit that we performed before the t-test.) By comparing the difference between the means with the difference between means that would be statistically significant, the t-test generates a p-value. This value was then compared to a *p*-value of 0.1 rather than 0.05. If $p < 0.1$ we tentatively rejected the null hypothesis. In the two cases in which we rejected the null hypothesis, we performed a simple bootstrapping technique (described in Methods section) to determine if our decision to reject the null hypothesis was valid.

11) Which years are used for 4xCO2? Is a linear fit appropriate?

We use years 11-50 for 4xCO2. The 4xCO2 experiment results in a total warming of about 6 K globally. However, the full response takes several hundred years to occur. After about 100 years, only about 75% of the warming occurs. The fast response, about 55% of the warming, happens in the first ten years of the simulation (Caldeira and Myhrvold, 2013). Therefore, between years 11-50, we have already passed the period of abrupt transient warming and are on a gradual curve, with no more than about 10% - 15% of the warming happening gradually between years 11 and 50.

In our experiment, years 11-50 are characterized by a relatively smooth warming of around 0.2 K decade^{-1}. The linear fit is appropriate for the data. Additionally, if there were dramatic warming in the early part of years 11-50 with gradual warming thereafter, the trend would create spurious El Niño events in the early part of the time series.

Caldeira, K., and N. P. Myhrvold: Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. Environ. Res. Lett., 8, 034039. 2013.

12) Page 29, line 982: “we find reasonable agreement between all aggregated model data and the observations during all historical time intervals” – Not explained how this is found

The reviewer is correct in pointing out that the statement is not explained. We have omitted that statement and substantially rewritten the discussion section. Additionally, in light of several comments, including this comment and the next comment (13), as well as earlier comments from reviewer 3 we reevaluated our finding pertinent to the historical models.

A re-review of the literature shows strong evidence that ENSO variability was in fact enhanced in the late 20th century, relative to the rest of an extended proxy record lasting either 478 or 700 years (Gergis and Fowler 2006; Li et al. 2013). Also, by synthesizing a large number of proxies spanning over 400 years McGregor et al. (2013) finds that ENSO was more active between 1979-2009 relative to any other 30 year period between 1600-1900. Using multiple coral records, Cobb et al. (2013) find that late 20th century variability was 42% greater than the 7,000 year average

As with other comparisons, we compared the ENSO frequency and amplitude statistics from 40 year slices of the historical simulation record dating back to 1850 using a two sample t-test assuming unequal variance. For example, a comparison between 40 year periods from the late 19th century, mid 20th century and late 21st century showed that differences in ENSO behavior between the three periods were not detectable or not present, except in the case of warm event frequency from 1966-2005, which is elevated relative to the 1916-1955 and 1966-1905 comparison periods.

Observations from the period 1966-2005 were almost identical to the mean of the models for that period. Both the models and observations show that the 1966-2005 period is elevated relative to the other historical periods (90% confidence).

In light of the correct identification of this finding, the portion of section **3.2 Analysis** that deals with analysis of historical and observational data has been substantially rewritten. Please see pp. 18-20, lines 444-478.

Cobb, K.M.; Sayani, H.R.; Di Lorenzo, E.; Westphal, N.; Watson, J.T.; Charles, C.D.; Cheng, H.; Edwards, R.L.. Science 339 (6115), 67-70, 2013.

Gergis J.L., Fowler, A.M.: How unusual was late twentieth century El Nino Southern Oscillation? Adv. Geosci.. 6, 173–79. 2006.

Li, J., Xie, S.-P., Cook E. R., Morales, M., Christie, D., Johnson, N., Chen, F., D'Arrigo, R., Fowler, A., Gou, X. and Fang, K.: El Niño modulations over the past seven centuries. Nature Climate Change, 3, 822–826. 2013.

McGregor, S., Timmermann, A., England, M. H., Elison Timm, O., and Wittenberg, A. T.: Inferred changes in El Niño–Southern Oscillation variance over the past six centuries, Clim. Past, 9, 2269-2284, 2013.

13) Page 30, line 1001: “However, given only 90% confidence, the relatively large number of comparisons made and the agreement between cold event frequency and amplitude in those same comparisons, the result is most likely by chance.” – There is no justification for why this is the case.

The reviewer is correct that the dismissal of the results was not justified. The dismissal was based on a misreading of the literature. We believed that the evidence pointed squarely toward the late 20th century exhibiting ENSO variability that was indistinguishable from the rest of the historical record. We erroneously believed that the published evidence directly contradicted a non-null finding. Because of this contradiction, we discussed and reported the finding, but attributed it to chance. However, further review of the paleoclimate literature reveals a tendency towards strong ENSO warm events in recent decades. For example, a 478-year reconstruction of ENSO variability reveals that up to 30% of anomalously strong ENSO events occurred in the late 20th century (Gergis and

Fowler, 2006). A 700-year tree ring record also reveals anomalously strong ENSO variability in the late 20th century (Li et al., 2013). A synthesis of multiple proxies over the past 400 years shows that the period 1979-2009 was more active than any 30 year period from 1600-1900 (McGregor et al. 2013). Additionally, the comparison between 1966-2005 and 1916-1955 comes very close to 95% confidence (p = 0.056). Given the corroboration of the statistical finding with the findings of other research, we choose to report and discuss the result.

The paragraph that contains the statement copied in reviewer's question 13 as well as the paragraph that follows in the manuscript has been substantially rewritten to reflect the fact that both the literature and the results from this experiment indicate anomalously strong El Niño events in the late 20th century. Additionally, this finding reveals that our methods were sensitive enough to detect both a 22% and 21% increase in ENSO warm event frequency given two samples of 18 ensembles. While our sensitivity analysis is confined to comparisons between samples from the same model, this historical period detectability threshold is in line with what we reported in the comparisons within models. Please see pp. 18-19, lines 444-478.

Gergis J.L., Fowler, A.M.: How unusual was late twentieth century El Nino Southern Oscillation? Adv. Geosci.. 6, 173–79. 2006.

Li, J., Xie, S.-P., Cook E. R., Morales, M., Christie, D., Johnson, N., Chen, F., D'Arrigo, R., Fowler, A., Gou, X. and Fang, K.: El Niño modulations over the past seven centuries. Nature Climate Change, 3, 822–826. 2013.

McGregor, S., Timmermann, A., England, M. H., Elison-Timm, O., and Wittenberg, A. T.: Inferred changes in El Niño–Southern Oscillation variance over the past six centuries, Clim. Past, 9, 2269-2284, 2013.

14) "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." – surely the conclusion ought to be that any changes in ENSO are below the detection threshold of your approach.

We agree that noting the detectability limitations in our method should be included in this conclusive statement. We have changed the first line of the discussion. Please see pp. 23-24, lines 608-611.

"We conclude that changes in ENSO event frequency and amplitude in a geoengineered world relative to the historical record, global warming simulations, and the observational record are either not present, or not large enough to be detectable using the approach employed in this experiment and described in the methods section above."

15) [Your approach] which has not been quantified). [What is the approach]

We believe you are referring to our approach for analyzing the data. We have elaborated on our approach to analyzing the data. Please see our answers to questions 9 and 10 above. Also, please see section **2. Methods** in the revised manuscript.

16) "The absence of a significant result, showing differences in ENSO variability between geoengineering and AGW may well be a valid finding." – This seems to suggest that the changes were too small to detect in this study and may well be too small to detect in the real world but it's unclear.

The reviewer is correct that the meaning of this statement is vague. It has been removed. The discussion section has been substantially rewritten.

17) Finally, the structure is very muddled with many of the sections containing text that seems wholly out of place. Additionally, where the authors have made efforts to revise their text, it reads as if they are responding to unasked questions.

We have reread the entire paper and made efforts to improve the structure of the paper. We agree that certain sections did seem a bit out of context and this revised version should be more organized.

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

Revised, September 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 Representative Concentration Pathway 4.5 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 19^{th} century. Finally, while warming of the Niño3.4 region in the tropical Pacific is fully offset in G1 and G2 during the 40 year simulations, the region continues to warm significantly in G3 and G4, which both start from a present day climate.

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 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 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? In addition to an exploration of changes in ENSO frequency and amplitude under different scenarios, we seek to determine how sea surface temperatures in the tropical Pacific will evolve under geoengineering, relative to historical and global warming scenarios over the entire length of the simulations.

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. Additionally, little work has been done to assess the oceanic response to SRM.

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 the Representative Concentration Pathway 8.5 (RCP8.5) scenario (Meinshausen et al., 2011). 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.

It is difficult to draw robust conclusions about future ENSO variability. There is no unambiguous signal of how ENSO may change under global warming in CMIP5. However, several recent studies have been able to detect statistically significant changes in ENSO. For example, Cai et al. (2015) shows a statistically significant increase 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 anthropogenic global warming (AGW) scenarios, such as RCP 6.0 and RCP 8.5 may improve detectability. Given that detecting an ENSO change in a 200-year record with 21 different participating GCMs is not straightforward, 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 to 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 in a comparison between one experiment and its control. 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 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% yr^{-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 RCP4.5 for G3 and G4. We compare the means of each sample by applying a two independent sample t-test assuming unequal variance. To apply this test, the populations making up the two samples being compared must both follow a normal distribution and the two populations must be measured on an equal-interval scale. In our case, we must establish normality to move forward to performing a valid t-test. If the samples were a bit larger ($n > 30$, where n is the size of the sample), the central limit theorem would likely make analysis of the normality of the respective samples moot.

There are many ways to assess normality. An important, but partially qualitative, first step in determining whether a sample is normally distributed is to create a histogram of values for each sample. This was done for each sample and the distribution appeared roughly normal.

Skewness and kurtosis are the properties of a distribution that serve as the basis for calculation of the widely used formal D'Agostino's K^2 test for goodness of fit. Conceptually, the K^2 test concurrently examines whether a sample is skewed (to the left or right) or peaked (or squished) relative to a normal distribution (D'Agostino, 1990). Skewness is a measure of symmetry around the sample mean, while kurtosis assesses if a distribution is sharply peaked or flattened relative to a normal distribution (DeCarlo, 1997). A perfect normal distribution has a skewness value of zero and kurtosis value of three. The kurtosis value of three for a normal distribution is equivalent to an excess kurtosis value of zero.

Skewness values of less twice that of $(6/n)^{0.5}$ are consistent with a symmetric distribution. Kurtosis values of less twice that of $(24/n)^{0.5}$ are consistent with a normal distribution. No metric evaluated in the experiments showed either skewness or kurtosis values exceeding the limits of what is consistent with a normal distribution. Based on this analysis, we are comfortable proceeding to use a two independent sample t-test. We are forced to assume unequal variance due to somewhat different variances within the samples being compared. We choose to use 90% confidence intervals to enhance detectability. However, by narrowing the confidence intervals, we are forced to supplement the finding of a significant result by either subsequently applying a bootstrapping method for an original finding pertinent to geoengineering and AGW, or to consult the appropriate studies to establish the veracity of a significant finding that matches the findings in other work, such as in a comparison between control or historical runs.

Even with carefully applied methods for analysis, 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, 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 RCP4.5 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 RCP4.5 – 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 Niño3.4 index was generated. We chose Niño3.4 over Niño 3 or Niño 4 because we find that the Niño3.4 region remains the center of action for ENSO variability both in observations and models. The Niño3.4 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 Niño3.4 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 Niño3.4 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.

Before we proceeded with the approach described above, concerns about the detectability of changes in ENSO variability during the period of modeled geoengineering compelled us to also 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 thresholds 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 sea level pressure (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, 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.

These endeavors to utilize a non-SST based Niño3.4 index are now set aside due in large part to the robust observational record of SST and the noisy nature of atmospheric variables. Next, we turn to defining what will constitute an ENSO event in our experiments. 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.

In addition to seeking to identify changes in ENSO variability, we attempt to describe the evolution of Niño3.4 SSTs during a period of geoengineering as compared to AGW or the historical record. To do this, we calculate the linear trend in the Niño3.4 index during the period 2030-2069 for G3, G4 and RCP4.5, 1966-2005 for historical simulations and Years 11-50 for G1, G2, +1% CO_2 and 4xCO2. Could imposition of a geoengineering regime partially or fully offset warming in Niño3.4 over a 40 year period?

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 Niño3.4 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 an initial test of model performance, we first evaluated agreement between the models used and the observational record. We found good agreement between 150 years of model data and the full observational record dating back 150 years. 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.

We also compared several selected 40-year periods from the historical simulations with historical simulations and observations from different 40-year periods in order to assess ENSO variability within the historical record. Are there 40-year periods in the historical record where ENSO variability is different than other 40-year periods, and, if so, is our detection method sensitive enough to detect the difference? The statistically significant differences (90% confidence) found are for comparisons made between warm event frequency between 1966-2005 and warm event frequency between 1866-1905 (p = 0.07) or 1916-1955 (p = 0.06). There is good agreement between models and observations throughout the record and this includes extremely close agreement between 1966-2005 historical simulations and the 1966-2005 observations in terms of ENSO warm event frequencies being elevated relative to the rest of the period. This also lends support to the validity of the 1966-2005 enhanced warm event finding. While the overall fit between historical models and observations is excellent, there is a good deal of model spread (see Figures 7 and 8).

We believe this finding to be robust in part because it is buttressed by the results of numerous studies using various combinations of observations and proxy records and historical modeling to reconstruct past ENSO behavior. A number of studies show similar findings about enhanced ENSO variability in the late 20th century. For example, Gergis and Fowler (2006) show that late 20th century El Niño frequency and intensity is significantly greater than it had been at any point since 1525. They further demonstrate that the post-1940 period accounts for between 30% and 40% of extreme and protracted El Niño events (Gergis and Fowler, 2006). Li et al. (2013) used 700 years of tree ring records from multiple locations to show that ENSO activity has been unusually high in the late 20th century. Additionally, a synthesis of multiple proxies over the past 400 years showed that the period 1979-2009 was more active than any 30 year period from 1600-1900 (McGregor et al. 2013). Based on 7,000 coral records, late 20th century ENSO is unusually strong, 42% greater than the 7,000 year average (Cobb et al., 2013).

However, we cannot conclude that the unusually strong ENSO variability in the late 20th century is the result of anthropogenic forcing, as there are other periods in the extended record where ENSO is either significantly enhanced or suppressed.

This finding from the historical record may not be germane to geoengineering, but it does test the limits of our method's detectability threshold and also demonstrates that ENSO behavior has exhibited significantly different properties during distinct portions of the historical record. A formal analysis of what percentage increase or decrease in ENSO event amplitude or frequency would be detectable given a particular sample size of 40-year runs is provided in the discussion section below.

We now turn back to the thrust of this paper, attempting to detect whether or not ENSO variability under a regime of geoengineering is distinct from ENSO variability under AGW. To do this, we perform a series of comparisons. 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. We find that 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. Because only six CMIP5 GeoMIP models produce a reasonable ENSO, 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 many data as possible 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 RCP4.5 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 detectability issues will be covered in greater detail in the discussion section.

3.3 Comparisons between 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.

3.4 Long term behavior of Niño3.4 under geoengineering

While ENSO is a dominant source of interannual variability both in the tropical Pacific and globally, the evolution of conditions in the Niño3.4 region on time scales much longer than that of ENSO are also of importance to regional and global climate. Individual ENSO events are modulated by the complex interaction of positive and negative feedbacks. The long-term trend in SSTs in this region is heavily influenced by other sources natural variability on decadal time scales. However, over a 40-year period of global warming or geoengineering, the SST trend in this region will be largely dependent on anthropogenic forcing or the combination of geoengineering and anthropogenic forcing. In order to examine the change in Niño3.4 over the full duration of each experiment, we calculate the linear trend of warming or cooling in the

Niño3.4 index over the applicable 40-year period. The linear trends are calculated over 2030-2069 for G3, G4 and RCP4.5, and over years 11-50 for G1, G2, 4xCO2 and +1% CO_2 yr^{-1} (Table 3, Figure 10).

The objective of geoengineering is to cool Earth's surface sufficiently so as to offset some of the negative impacts of global warming. The amount of temperature change expected over an extended period of geoengineering as opposed to under AGW alone is an obvious indicator of the potential efficacy of geoengineering in a particular region. SSTs over Niño3.4 region change considerably under global warming and under some geoengineering scenarios. Under G1 and G2, the linear trend of SSTs in Niño3.4 is negative, and very close to zero. In both experiments, a CO_2 increase is imposed on a steady-state preindustrial climate concurrently with a fully offsetting solar constant reduction. G1 and G2 generally depict little change in global mean temperature. In terms of variability between models simulations of Niño3.4 linear trend, the coarsest error bars can be found with the G1 and G2 experiment. Climate was fully stabilized in most runs, but some runs exhibited evidence of either a warming or cooling trend.

G3 and G4 initialize from a warming climate and seek to offset RCP4.5 with SO_2 injections. In both experiments, the Niño3.4 region continues to warm despite the SO_2 injections. In G3 the warming trend is 0.07 K $decade^{-1}$. However, the G3 experiment is designed so that the RCP4.5 warming is fully offset to generate zero net forcing. Offsetting radiative forcing by injecting a layer of stratospheric aerosols does not prevent continued warming in the ocean. Put simply, the ocean has a huge thermal mass. Rising or falling air temperatures take time before their impact is felt in the ocean.

For G4, the initial SO_2 forcing is fully offset at first, but during the experiment, the amount of SO_2 being injected into the atmosphere does not change, even as the RCP4.5 forcing grows. Therefore, it is unsurprising that Niño3.4 continues to warm under G4. However, the trend of 0.13 K decade^{-1} is rather robust and is only a 23% weaker than the trend observed from 1966-2005 of 0.16 K decade^{-1}.

Only the geoengineering scenarios that are physically implausible (G1-G2) fully offset global warming induced SST changes in Niño3.4. The more realistic scenarios (G3 and G4) are able to significantly reduce the magnitude of warming under RCP4.5. The RCP4.5 warming trend of 0.21 K decade^{-1} is three times stronger than G3 and 62% stronger than the warming trend seen in the G4 experiment, which is the experiment that best reflects the process by which stratospheric geoengineering would actually be deployed in the real world.

The warming trends are 0.22 K decade^{-1} in +1% CO_2 yr^{-1} and 0.17 K decade^{-1} in 4xCO2. Our results about the evolution of Niño3.4 warming under 4xCO2 are in line with the extensive analysis of the global 4xCO2 response described by Caldeira and Myhrvold (2013). They show that about 55% of the 4xCO2 warming occurs in the first 10 years and then a relatively slow trend develops for the next 50 years during which another approximately 15% of the total warming occurs. After about 60 years the warming flattens out substantially and it takes several hundred years for the full extent of the warming to be evident. If the total 4xCO2 warming is 6 K, we would expect about 10% to 15% (0.6 K to 0.9 K) of that to occur over our experimental period.

The key finding in our experiment with regard to the long term behavior of the Niño3.4 Index is that under RCP4.5 the warming trend will be 62% stronger between 2030-2069 than if we imposed geoengineering as simulated by G4 beginning in 2020.

4 Discussion

We conclude that changes in ENSO event frequency and amplitude in a geoengineered world relative to the historical record, global warming simulations, and the observational record are either not present, or not large enough to be detectable using the approach employed in this experiment and described in the methods section above. However, this conclusion comes with a number of very strong caveats, including the relatively brief simulation length and the considerable model spread. The absence of a detectable change in ENSO amplitude and frequency in our experiments is fine. However, it is of interest to explicitly state the conditions under which changes in ENSO variability could have been detected. Therefore, we assess the sensitivity of our method for identifying differences in ENSO frequency and amplitude between an experiment and its control. We begin by calculating the minimum increase in event frequency or amplitude that would be detectable.

Since GISS is a relatively well-performing model, and the standard deviation of event amplitude and frequency is relatively low, we address the minimum detectability issue with the GISS model first. We take the GISS runs and randomly assign each simulation into one of two groups of 11. The random assignment to each group is repeated many times. The amplitude and frequency statistics of the two groups are then compared using a two-sample t-test assuming unequal variance. We select the comparison that generates a probability of wrongly rejecting the hypothesis that there is no difference between the means that is closest to but not greater than 0.05 (0.03-0.05 in all cases). This corresponds to a significant result at the 95% level. The difference between these two means, expressed as a percentage increase, is then reported as the threshold value of detectable change in ENSO frequency or amplitude.

In the GISS model, to detect a 31% increase in El Niño frequency, two groups of 11 ensembles of 40 year simulations are required. A 26% increase in La Niña event frequency is required for detectability. In terms of amplitude, an 18% increase in El Niño amplitude was detectable, as was a 17% increase in La Niña frequency.

We then use the same approach to test the minimum detectability threshold in the CSIRO model; we expect to find increased sensitivity to changes in event frequency, but not necessarily to amplitude. In fact, a 25% increase in warm event frequency and a 21% increase in cold event frequency are both detectable. The ENSO event amplitude detectability threshold in CSIRO is close to that found in the GISS model: 17% for warm events and 16% for cold events.

Lastly, we examine the CanESM model, which featured the largest and most frequent ENSO events, but also featured close agreement between ensemble members. The thresholds for detectability for the CanESM are increases of 16% for both warm and cold event frequency, 18% for an increase in warm event amplitude, and 15% for an increase in cold event amplitude. A priority in subsequent experiments would be to reduce the detectability thresholds. For the purposes of improving detectability, steady-state simulations that simulate 100 years or more of geoengineering in which a large CO_2 forcing is offset would likely be of most interest for additional detailed analysis, especially if a large number of simulations were available. This has been proposed for future GeoMIP experiments (Kravitz 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 apply our detection methodology and 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 in depiction of future climate between 4xCO2, G1 and the preindustrial control, this would likely be the comparison most likely to show a difference sufficient to exceed our detectability threshold.

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. We would also recommend comparing the extended GeoMIP simulation results to the late 20th century ENSO record. 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.

We now turn to another important issue that hinders detectability of changes in ENSO variability: the ability of current generation GCMs to accurately reproduce ENSO. 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.

Since the ENSO mechanism is so vulnerable to small perturbations, subtle differences in how forcing is applied could impact the ENSO mechanism in the models. Specifically, emissions are imposed differently in the historical runs from how they are imposed in RCP4.5. RCP4.5 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 RCP4.5 ENSO variability could seek to determine if the interannual variability in SSTs is muted enough by this smoothing in RCP4.5 to potentially alter the evolution of ENSO events.

While our results pertaining specifically to changes in ENSO frequency and amplitude reveal that detectability of changes is difficult, the overall trend in Niño3.4 is clear under each potential future scenario. The most important conclusion from analysis of the long-term trend in Niño3.4 is that the warming trend in RCP4.5 would be 62% stronger than in G4, the most realistic geoengineering scenario. Changes in SSTs in Niño3.4 are important because the mean climate of this region is an important factor in enhancing weather and climate trends on multiple time scales. For example, the 0.9 K warming in Niño3.4 over 40 years under RCP4.5 would greatly enhance the amount of precipitable water in the region. Meridional transport of water vapor out of the tropics occurs through relatively narrow regions in the atmosphere. These so-called atmospheric rivers advect moisture into the baroclinic zone and are responsible for many extreme precipitation events in North America and other places.

Superposing a 0.9 K warmer Niño3.4 mean onto a strong El Niño event could easily result in more extreme remote ENSO impacts, such as flooding. Additionally, increased SSTs results in warmer surface air temperatures and a steeper lapse rate. This promotes broad areas of enhanced convection over warm ocean areas, which induce poleward propagating wave motion in the atmosphere, which alter the general circulation and change weather patterns around the world. Would a 0.9 K warming of Niño3.4 cause generally enhanced convection over the Niño3.4 region and would this convection induce detectable changes in the general circulation? Would the warming of only 0.5 K under G4 result in different regional and global warming impacts than the RCP4.5 warming? Certainly it is worthwhile to study how long term trends in particular regions such as Niño3.4 may be closely related to other long term trends in remote places and how these relationships may differ under geoengineering as opposed to AGW.

Lastly, 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, 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.

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.

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

Caldeira, K., and Myhrvold, N. P.: Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. Environ. Res. Lett., 8, 034039, 2013.

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

Cobb, K. M., Westphal, N., Sayani, H. R., Watson, J. T., Di Lorenzo, E., Cheng, H., Edwards, R. L., and Charles, C. D.: Highly variable El Niño-Southern Oscillation throughout the Holocene, Science 339 (6115), 67-70, 2013.

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.

D'Agostino, R. B., Belanger A., and Ralph B.: A suggestion for using powerful and informative tests of normality. The American Statistician 44, 316–321, 1990.

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., and Yang Z.-L.: The Common Land Model (CLM), Bull. Am. Meteorol. Soc., 84, 1013–1023, 2003.

Decarlo, L.T.: On the meaning and use of kurtosis. Psychological Methods 2, 292-307, 1997.

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., and 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., and Haug, G.H.: Volcanoes and ENSO over the past millennium, J. Climate, 21(13), 3134–3148, 2007.

Gergis J. L., and Fowler, A.M.: How unusual was late twentieth century El Niño-Southern Oscillation? Adv. Geosci., 6, 173–79, 2006.

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. Adv. Modeling Earth Sys., 5, 572-597, 2013.

Guilyardi E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Oldenborgh, G. J. V., and 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.

Kaplan, A., Cane, M., Kushnir, Y., Clement, A., Blumenthal, M., and Rajagopalan, B.: Analyses of global sea surface temperature 1856-1991, J. Geophys. Res., 103, 18,567-18,589, 1998.

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.

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. Disc., 8, 4697-4736, doi:10.5194/gmdd-8-4697-2015, 2015.

Li, J., Xie, S.-P., Cook E. R., Morales, M., Christie, D., Johnson, N., Chen, F., D'Arrigo, R., Fowler, A., Gou, X., and Fang, K.: El Niño modulations over the past seven centuries. Nature Climate Change, 3, 822–826, 2013.

Maher, N., McGregor, S., England, M. H., and Gupta, A. S.: Effects of volcanism on tropical variability. Geophys. Res. Lett., 42(14):6024-6033, 2015.

Mann, M. E., Cane, M. A., Zebiak, S. E., and 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.

McGregor, S., Timmermann, A., England, M. H., Elison Timm, O., and Wittenberg, A. T.: Inferred changes in El Niño–Southern Oscillation variance over the past six centuries, Clim. Past, 9, 2269-2284, 2013.

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

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extension from 1765 to 2300, Climatic Change, 109, 213-241, doi:10.1007/s10584-011-0156-z, 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 1: Description and evaluation, Geoscientific Model Dev., 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 Dev., 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. Climate, 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., and 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., and 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., and Cane M. L.: A model El Niño/Southern Oscillation. Mon. Wea. Rev., 115, 2262–2278, 1987.

1183

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

*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)

b. Models in G2

*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

*BNU-ESM (1)
GISS-E2-R (3)
HadGEM2-ES (2)
IPSL-CM5A-LR (1)
MPI-ESM-LR (3)

d. Models in G4

*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)

Table 3. Difference in linear trend of Niño3.4 Index (5°S-5°N, 120°W-170°W) between each experiment G1-G4 and its control.

G1 minus 4xCO2	-0.18 K $decade^{-1}$
G2 minus +1% CO_2 yr^{-1}	-0.24 K $decade^{-1}$
G3 minus RCP4.5	-0.14 K $decade^{-1}$
G4 minus RCP4.5	-0.07 K $decade^{-1}$

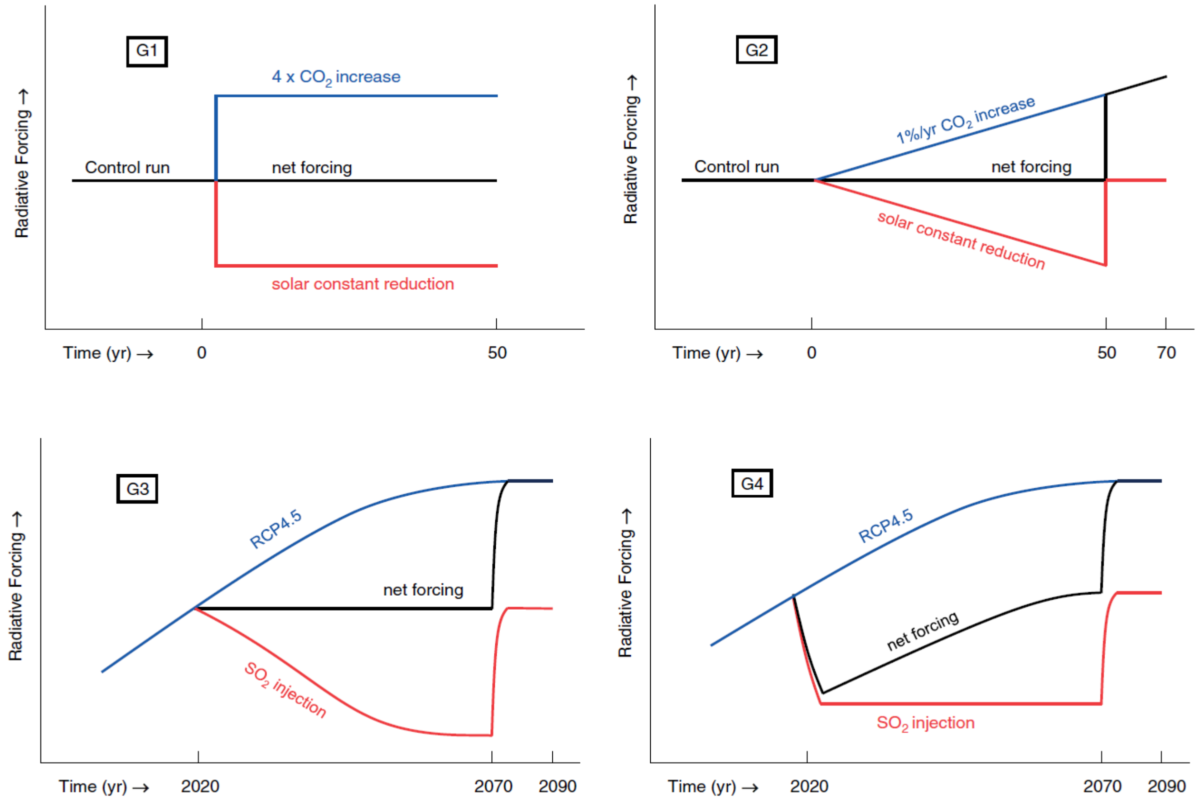
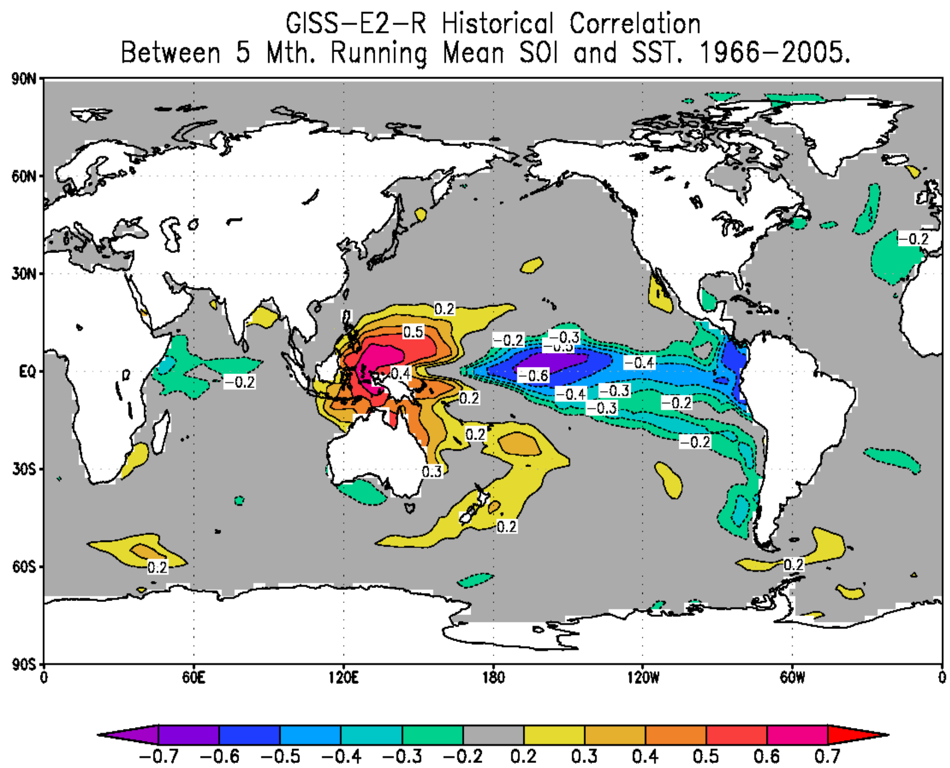


Figure 1. GeoMIP G1-G4 experiment designs. Figures 1-4 from Kravitz et al. (2011).



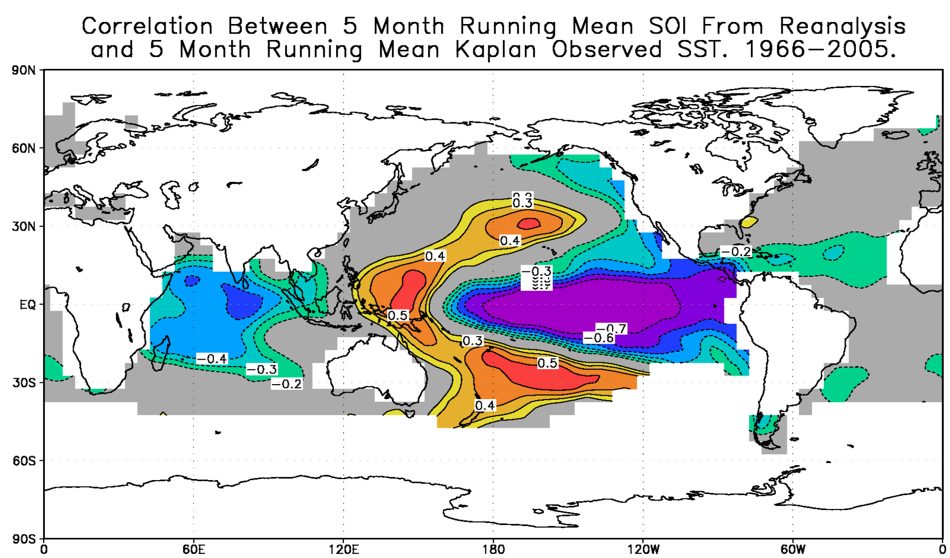


Figure 2. Top panel shows spatial correlation between GISS historical sea surface temperature (SST) and the Southern Oscillation Index (SOI). The area of strong negative correlation is confined to a small region in the central Pacific, relative to the broad area of strong negative correlation in the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

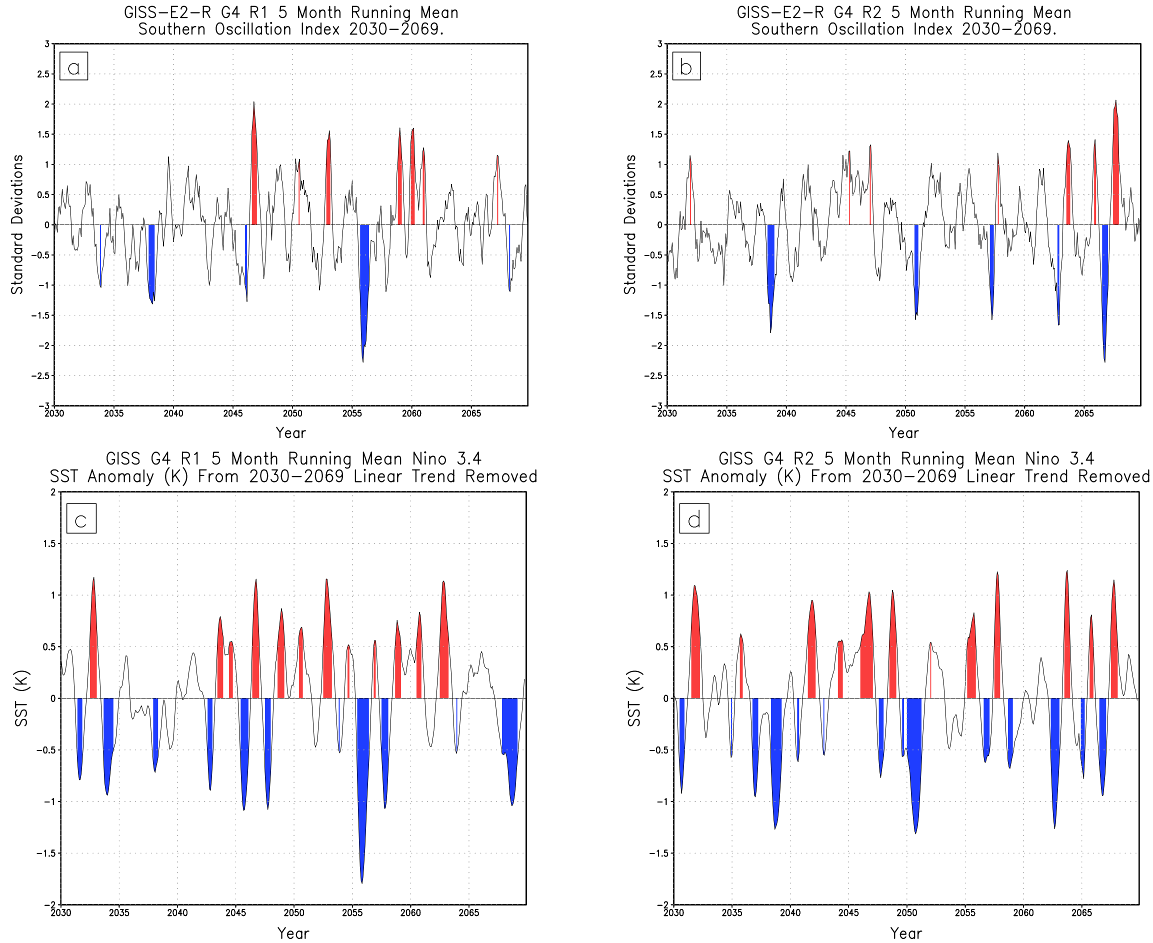


Figure 3. Time series of normalized Southern Oscillation Index (SOI) for (a) GISS G4 Run 1 and (b) GISS G4 Run 2. In the context of SOI, ENSO events are defined as departures of 0.5 standard deviations from zero. SOI warm events are highlighted in red, while cold events are highlighted in blue. No highlight is applied during an ENSO neutral phase. Time series of SST in Niño3.4 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.

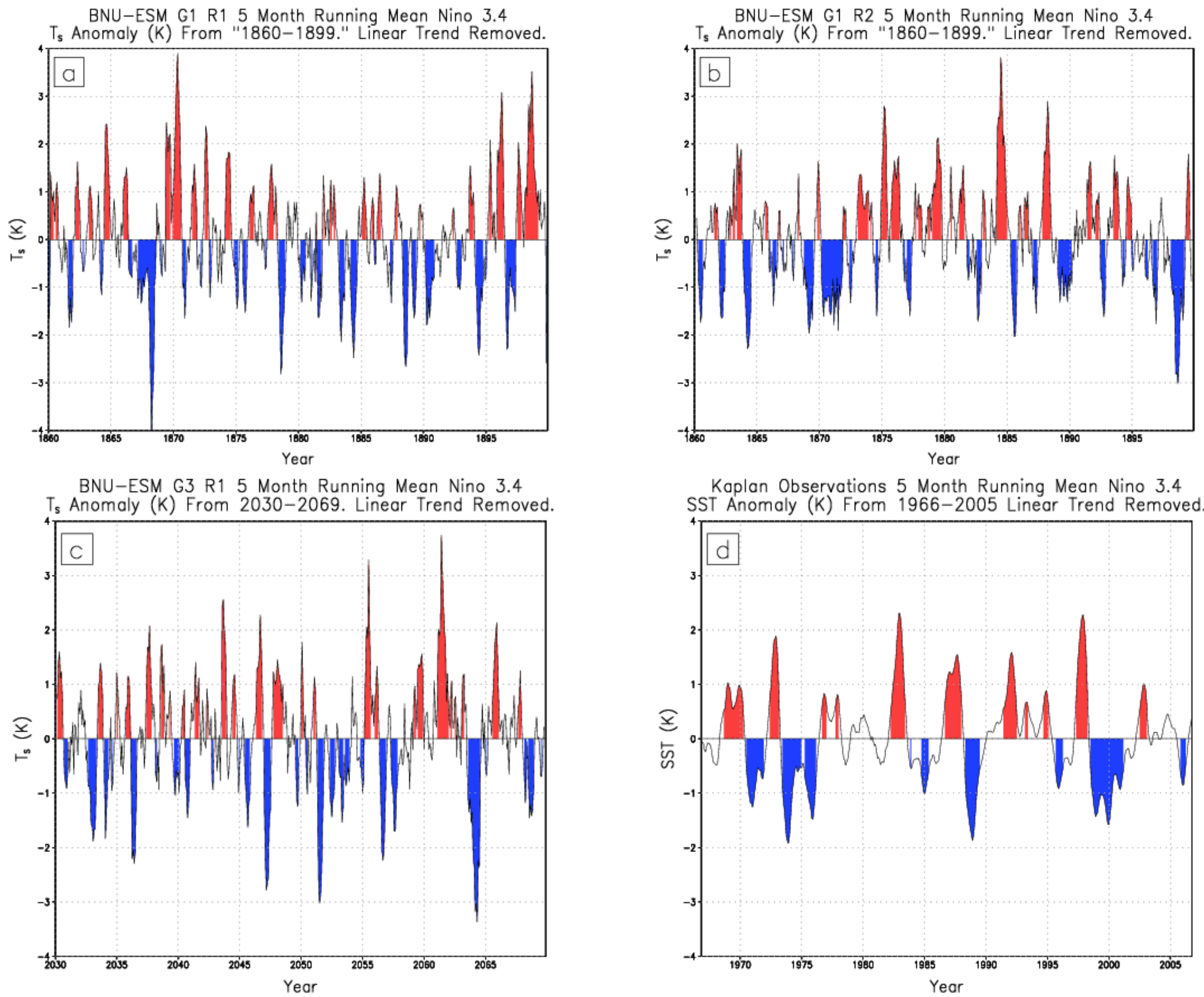


Figure 4. Time series of Niño3.4 anomalies from three experimental runs, (a) G1 run 1, (b) G1 run 2, and (c) G3 run 1, of the BNU-ESM model compared to observations (d). Red coloring indicates ENSO warm events, while blue shading indicates ENSO cold events. The model is excluded due to unrealistic variability and amplitude.

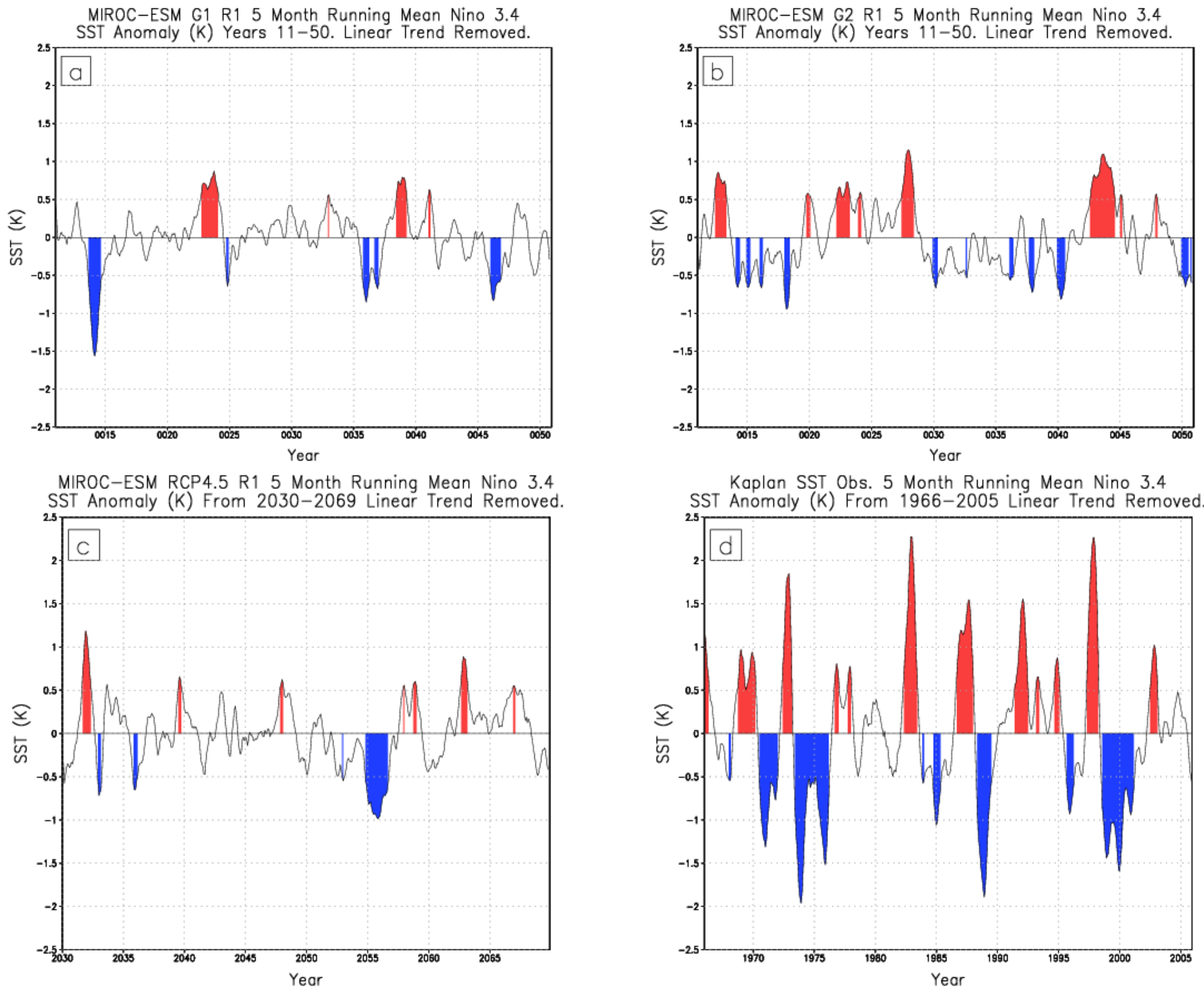
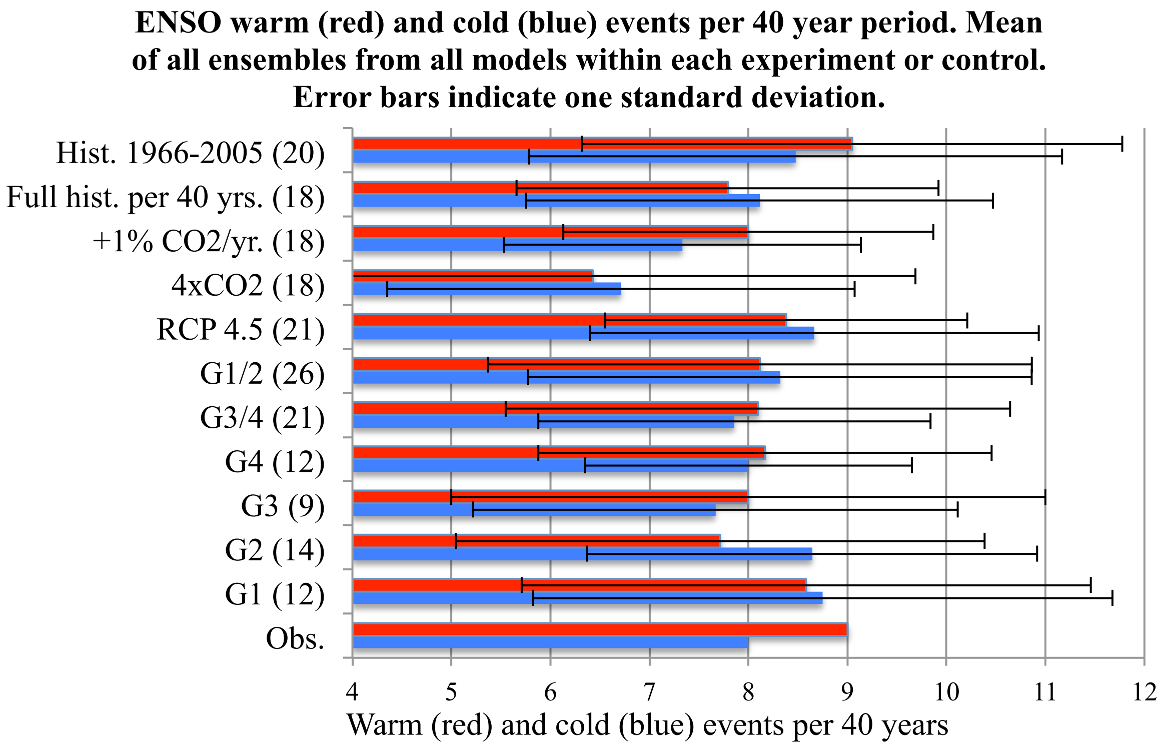
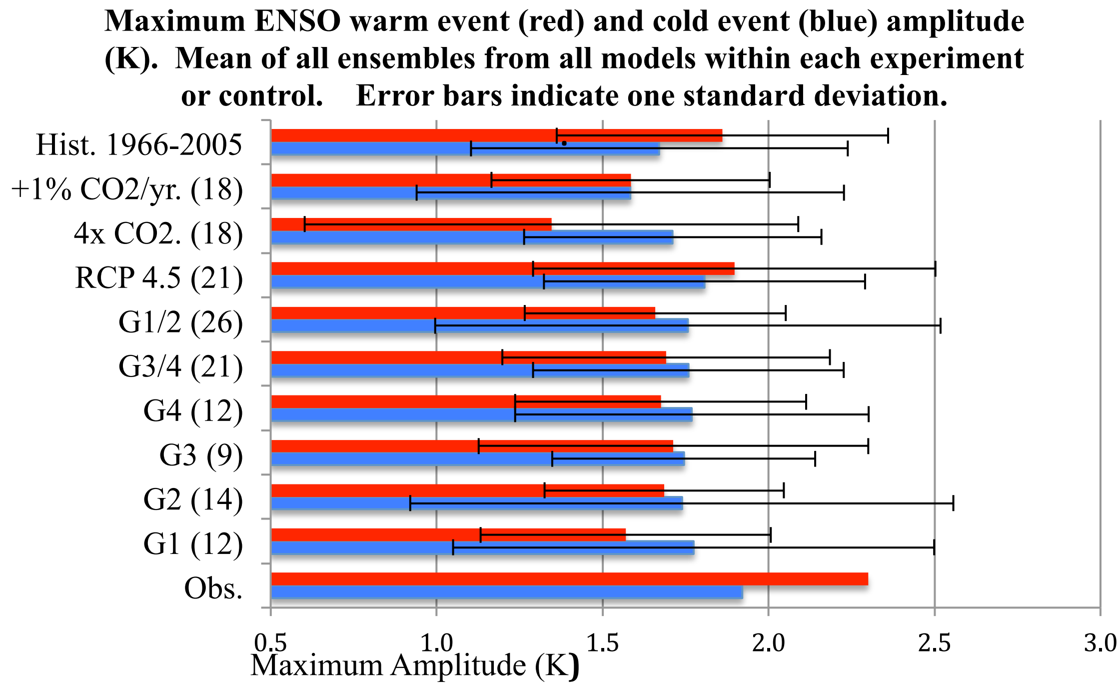


Figure 5. Niño3.4 anomalies for MIROC-ESM. Time series of (a) G1, (b) G2 and (c) RCP4.5 from each model all show significantly muted variability and amplitude compared to (d) observations, with few, if any, warm events exceeding a 1 K anomaly. All other MIROC family experiments showed the same muted 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.



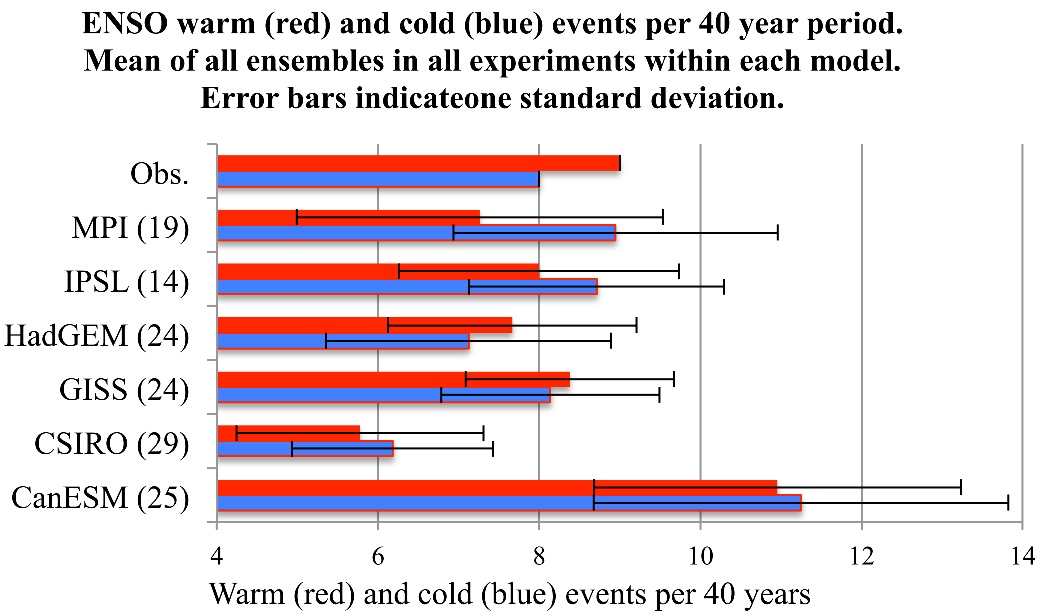
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2 /yr. (18)	1850-2005 /40 yr. (18)	1966-2005 (18)
Warm events	9.00	8.58	7.71	8.00	8.17	8.10	8.12	8.38	6.43	8.00	7.79	9.05
Cold events	8.00	8.75	8.64	7.67	8.00	7.86	8.32	8.67	6.71	7.33	8.11	8.48

Figure 6. Number of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G3, G4, RCP4.5, Years 11-50 for G1, G2, +1% CO_2 yr^{-1} 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.



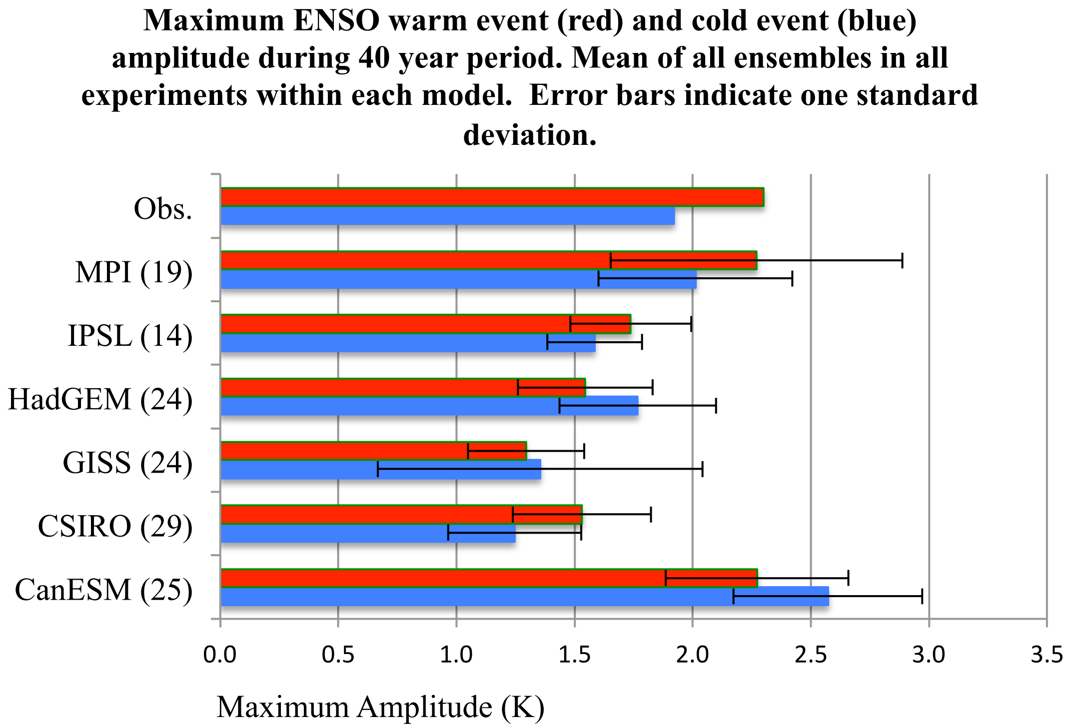
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2/yr. (18)	1966-2005 (18)
Warm amp.	2.30	1.57	1.69	1.71	1.68	1.69	1.66	1.90	1.35	1.58	1.86
Cold amp.	1.92	1.77	1.74	1.74	1.77	1.76	1.76	1.81	1.71	1.58	1.67

Figure 7. Maximum amplitude (K) of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G3, G4, RCP4.5, Years 11-50 for G1, G2, +1% CO_2 yr^{-1} 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.

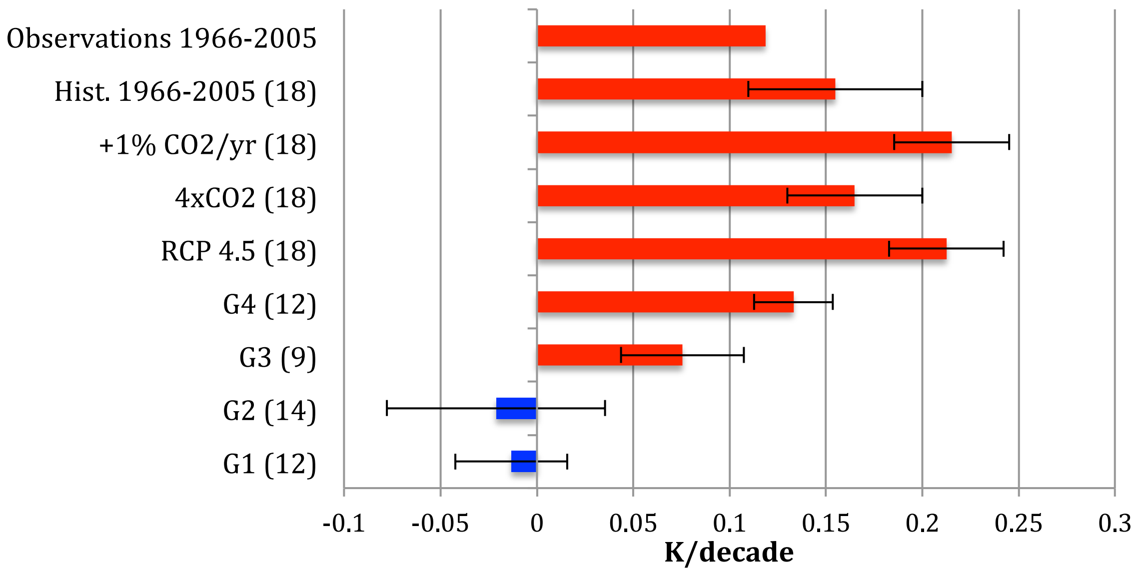


Figure 10. Linear trend in Niño3.4 Index (5°S-5°N, 120°W-170°W) during the applicable 40-year comparison periods. The applicable comparison periods are 1966-2005 for historical simulations; years 11-50 for G1, G2, +1% CO_2 yr^{-1} and 4xCO2; 2030-2069 for G3, G4 and RCP4.5; and 1966-2005 for historical simulations and observations. The values in parenthesis are the number of ensemble members for each experiment. Red bars indicate a warming trend. Blue bars indicate a cooling trend. Error bars indicate one standard deviation.

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

Revised, September 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 Representative Concentration Pathway 4.5 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. Finally, while warming of the Niño3.4 region in the tropical Pacific is fully offset in G1 and G2 during the 40 year simulations, the region continues to warm significantly in G3 and G4, which both start from a present day climate.

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 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 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? In addition to an exploration of changes in ENSO frequency and amplitude under different scenarios, we seek to determine how sea surface temperatures in the tropical Pacific will evolve under geoengineering, relative to historical and global warming scenarios over the entire length of the simulations.

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. Additionally, little work has been done to assess the oceanic response to SRM.

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 the Representative Concentration Pathway 8.5 (RCP8.5) scenario (Meinshausen et al., 2011). 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.

It is difficult to draw robust conclusions about future ENSO variability. There is no unambiguous signal of how ENSO may change under global warming in CMIP5. However, several recent studies have been able to detect statistically significant changes in ENSO. For example, Cai et al. (2015) shows a statistically significant increase 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 anthropogenic global warming (AGW) scenarios, such as RCP 6.0 and RCP 8.5 may improve detectability. Given that detecting an ENSO change in a 200-year record with 21 different participating GCMs is not straightforward, 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 to 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 in a comparison between one experiment and its control. 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 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% yr^{-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 RCP4.5 for G3 and G4. We compare the means of each sample by applying a two independent sample t-test assuming unequal variance. To apply this test, the populations making up the two samples being compared must both follow a normal distribution and the two populations must be measured on an equal-interval scale. In our case, we must establish normality to move forward to performing a valid t-test. If the samples were a bit larger ($n > 30$, where n is the size of the sample), the central limit theorem would likely make analysis of the normality of the respective samples moot.

There are many ways to assess normality. An important, but partially qualitative, first step in determining whether a sample is normally distributed is to create a histogram of values for each sample. This was done for each sample and the distribution appeared roughly normal.

Skewness and kurtosis are the properties of a distribution that serve as the basis for calculation of the widely used formal D'Agostino's K^2 test for goodness of fit. Conceptually, the K^2 test concurrently examines whether a sample is skewed (to the left or right) or peaked (or squished) relative to a normal distribution (D'Agostino, 1990). Skewness is a measure of symmetry around the sample mean, while kurtosis assesses if a distribution is sharply peaked or flattened relative to a normal distribution (DeCarlo, 1997). A perfect normal distribution has a skewness value of zero and kurtosis value of three. The kurtosis value of three for a normal distribution is equivalent to an excess kurtosis value of zero.

Skewness values of less twice that of $(6/n)^{0.5}$ are consistent with a symmetric distribution. Kurtosis values of less twice that of $(24/n)^{0.5}$ are consistent with a normal distribution. No metric evaluated in the experiments showed either skewness or kurtosis values exceeding the limits of what is consistent with a normal distribution. Based on this analysis, we are comfortable proceeding to use a two independent sample t-test. We are forced to assume unequal variance due to somewhat different variances within the samples being compared. We choose to use 90% confidence intervals to enhance detectability. However, by narrowing the confidence intervals, we are forced to supplement the finding of a significant result by either subsequently applying a bootstrapping method for an original finding pertinent to geoengineering and AGW, or to consult the appropriate studies to establish the veracity of a significant finding that matches the findings in other work, such as in a comparison between control or historical runs.

Even with carefully applied methods for analysis, 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, 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 RCP4.5 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 RCP4.5 – 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 Niño3.4 index was generated. We chose Niño3.4 over Niño 3 or Niño 4 because we find that the Niño3.4 region remains the center of action for ENSO variability both in observations and models. The Niño3.4 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 Niño3.4 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 Niño3.4 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.

Before we proceeded with the approach described above, concerns about the detectability of changes in ENSO variability during the period of modeled geoengineering compelled us to also 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 thresholds 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 sea level pressure (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, 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.

These endeavors to utilize a non-SST based Niño3.4 index are now set aside due in large part to the robust observational record of SST and the noisy nature of atmospheric variables. Next, we turn to defining what will constitute an ENSO event in our experiments. 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.

In addition to seeking to identify changes in ENSO variability, we attempt to describe the evolution of Niño3.4 SSTs during a period of geoengineering as compared to AGW or the historical record. To do this, we calculate the linear trend in the Niño3.4 index during the period 2030-2069 for G3, G4 and RCP4.5, 1966-2005 for historical simulations and Years 11-50 for G1, G2, +1% CO_2 and 4xCO2. Could imposition of a geoengineering regime partially or fully offset warming in Niño3.4 over a 40 year period?

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 Niño3.4 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 an initial test of model performance, we first evaluated agreement between the models used and the observational record. We found good agreement between 150 years of model data and the full observational record dating back 150 years. 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.

We also compared several selected 40-year periods from the historical simulations with historical simulations and observations from different 40-year periods in order to assess ENSO variability within the historical record. Are there 40-year periods in the historical record where ENSO variability is different than other 40-year periods, and, if so, is our detection method sensitive enough to detect the difference? The statistically significant differences (90% confidence) found are for comparisons made between warm event frequency between 1966-2005 and warm event frequency between 1866-1905 (p = 0.07) or 1916-1955 (p = 0.06). There is good agreement between models and observations throughout the record and this includes extremely close agreement between 1966-2005 historical simulations and the 1966-2005 observations in terms of ENSO warm event frequencies being elevated relative to the rest of the period. This also lends support to the validity of the 1966-2005 enhanced warm event finding. While the overall fit between historical models and observations is excellent, there is a good deal of model spread (see Figures 7 and 8).

We believe this finding to be robust in part because it is buttressed by the results of numerous studies using various combinations of observations and proxy records and historical modeling to reconstruct past ENSO behavior. A number of studies show similar findings about enhanced ENSO variability in the late 20th century. For example, Gergis and Fowler (2006) show that late 20th century El Niño frequency and intensity is significantly greater than it had been at any point since 1525. They further demonstrate that the post-1940 period accounts for between 30% and 40% of extreme and protracted El Niño events (Gergis and Fowler, 2006). Li et al. (2013) used 700 years of tree ring records from multiple locations to show that ENSO activity has been unusually high in the late 20th century. Additionally, a synthesis of multiple proxies over the past 400 years showed that the period 1979-2009 was more active than any 30 year period from 1600-1900 (McGregor et al. 2013). Based on 7,000 coral records, late 20th century ENSO is unusually strong, 42% greater than the 7,000 year average (Cobb et al., 2013).

However, we cannot conclude that the unusually strong ENSO variability in the late 20th century is the result of anthropogenic forcing, as there are other periods in the extended record where ENSO is either significantly enhanced or suppressed.

This finding from the historical record may not be germane to geoengineering, but it does test the limits of our method's detectability threshold and also demonstrates that ENSO behavior has exhibited significantly different properties during distinct portions of the historical record. A formal analysis of what percentage increase or decrease in ENSO event amplitude or frequency would be detectable given a particular sample size of 40-year runs is provided in the discussion section below.

We now turn back to the thrust of this paper, attempting to detect whether or not ENSO variability under a regime of geoengineering is distinct from ENSO variability under AGW. To do this, we perform a series of comparisons. 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. We find that 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. Because only six CMIP5 GeoMIP models produce a reasonable ENSO, 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 many data as possible 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 RCP4.5 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 detectability issues will be covered in greater detail in the discussion section.

3.3 Comparisons between 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.

3.4 Long term behavior of Niño3.4 under geoengineering

While ENSO is a dominant source of interannual variability both in the tropical Pacific and globally, the evolution of conditions in the Niño3.4 region on time scales much longer than that of ENSO are also of importance to regional and global climate. Individual ENSO events are modulated by the complex interaction of positive and negative feedbacks. The long-term trend in SSTs in this region is heavily influenced by other sources natural variability on decadal time scales. However, over a 40-year period of global warming or geoengineering, the SST trend in this region will be largely dependent on anthropogenic forcing or the combination of geoengineering and anthropogenic forcing. In order to examine the change in Niño3.4 over the full duration of each experiment, we calculate the linear trend of warming or cooling in the

Niño3.4 index over the applicable 40-year period. The linear trends are calculated over 2030-2069 for G3, G4 and RCP4.5, and over years 11-50 for G1, G2, 4xCO2 and +1% CO_2 yr^{-1} (Table 3, Figure 10).

The objective of geoengineering is to cool Earth's surface sufficiently so as to offset some of the negative impacts of global warming. The amount of temperature change expected over an extended period of geoengineering as opposed to under AGW alone is an obvious indicator of the potential efficacy of geoengineering in a particular region. SSTs over Niño3.4 region change considerably under global warming and under some geoengineering scenarios. Under G1 and G2, the linear trend of SSTs in Niño3.4 is negative, and very close to zero. In both experiments, a CO_2 increase is imposed on a steady-state preindustrial climate concurrently with a fully offsetting solar constant reduction. G1 and G2 generally depict little change in global mean temperature. In terms of variability between models simulations of Niño3.4 linear trend, the coarsest error bars can be found with the G1 and G2 experiment. Climate was fully stabilized in most runs, but some runs exhibited evidence of either a warming or cooling trend.

G3 and G4 initialize from a warming climate and seek to offset RCP4.5 with SO_2 injections. In both experiments, the Niño3.4 region continues to warm despite the SO_2 injections. In G3 the warming trend is 0.07 K $decade^{-1}$. However, the G3 experiment is designed so that the RCP4.5 warming is fully offset to generate zero net forcing. Offsetting radiative forcing by injecting a layer of stratospheric aerosols does not prevent continued warming in the ocean. Put simply, the ocean has a huge thermal mass. Rising or falling air temperatures take time before their impact is felt in the ocean.

For G4, the initial SO_2 forcing is fully offset at first, but during the experiment, the amount of SO_2 being injected into the atmosphere does not change, even as the RCP4.5 forcing grows. Therefore, it is unsurprising that Niño3.4 continues to warm under G4. However, the trend of 0.13 K decade^{-1} is rather robust and is only a 23% weaker than the trend observed from 1966-2005 of 0.16 K decade^{-1}.

Only the geoengineering scenarios that are physically implausible (G1-G2) fully offset global warming induced SST changes in Niño3.4. The more realistic scenarios (G3 and G4) are able to significantly reduce the magnitude of warming under RCP4.5. The RCP4.5 warming trend of 0.21 K decade^{-1} is three times stronger than G3 and 62% stronger than the warming trend seen in the G4 experiment, which is the experiment that best reflects the process by which stratospheric geoengineering would actually be deployed in the real world.

The warming trends are 0.22 K decade^{-1} in +1% CO_2 yr^{-1} and 0.17 K decade^{-1} in 4xCO2. Our results about the evolution of Niño3.4 warming under 4xCO2 are in line with the extensive analysis of the global 4xCO2 response described by Caldeira and Myhrvold (2013). They show that about 55% of the 4xCO2 warming occurs in the first 10 years and then a relatively slow trend develops for the next 50 years during which another approximately 15% of the total warming occurs. After about 60 years the warming flattens out substantially and it takes several hundred years for the full extent of the warming to be evident. If the total 4xCO2 warming is 6 K, we would expect about 10% to 15% (0.6 K to 0.9 K) of that to occur over our experimental period.

The key finding in our experiment with regard to the long term behavior of the Niño3.4 Index is that under RCP4.5 the warming trend will be 62% stronger between 2030-2069 than if we imposed geoengineering as simulated by G4 beginning in 2020.

4 Discussion

We conclude that changes in ENSO event frequency and amplitude in a geoengineered world relative to the historical record, global warming simulations, and the observational record are either not present, or not large enough to be detectable using the approach employed in this experiment and described in the methods section above. However, this conclusion comes with a number of very strong caveats, including the relatively brief simulation length and the considerable model spread. The absence of a detectable change in ENSO amplitude and frequency in our experiments is fine. However, it is of interest to explicitly state the conditions under which changes in ENSO variability could have been detected. Therefore, we assess the sensitivity of our method for identifying differences in ENSO frequency and amplitude between an experiment and its control. We begin by calculating the minimum increase in event frequency or amplitude that would be detectable.

Since GISS is a relatively well-performing model, and the standard deviation of event amplitude and frequency is relatively low, we address the minimum detectability issue with the GISS model first. We take the GISS runs and randomly assign each simulation into one of two groups of 11. The random assignment to each group is repeated many times. The amplitude and frequency statistics of the two groups are then compared using a two-sample t-test assuming unequal variance. We select the comparison that generates a probability of wrongly rejecting the hypothesis that there is no difference between the means that is closest to but not greater than 0.05 (0.03-0.05 in all cases). This corresponds to a significant result at the 95% level. The difference between these two means, expressed as a percentage increase, is then reported as the threshold value of detectable change in ENSO frequency or amplitude.

In the GISS model, to detect a 31% increase in El Niño frequency, two groups of 11 ensembles of 40 year simulations are required. A 26% increase in La Niña event frequency is required for detectability. In terms of amplitude, an 18% increase in El Niño amplitude was detectable, as was a 17% increase in La Niña frequency.

We then use the same approach to test the minimum detectability threshold in the CSIRO model; we expect to find increased sensitivity to changes in event frequency, but not necessarily to amplitude. In fact, a 25% increase in warm event frequency and a 21% increase in cold event frequency are both detectable. The ENSO event amplitude detectability threshold in CSIRO is close to that found in the GISS model: 17% for warm events and 16% for cold events.

Lastly, we examine the CanESM model, which featured the largest and most frequent ENSO events, but also featured close agreement between ensemble members. The thresholds for detectability for the CanESM are increases of 16% for both warm and cold event frequency, 18% for an increase in warm event amplitude, and 15% for an increase in cold event amplitude. A priority in subsequent experiments would be to reduce the detectability thresholds. For the purposes of improving detectability, steady-state simulations that simulate 100 years or more of geoengineering in which a large CO_2 forcing is offset would likely be of most interest for additional detailed analysis, especially if a large number of simulations were available. This has been proposed for future GeoMIP experiments (Kravitz 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 apply our detection methodology and 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 in depiction of future climate between 4xCO2, G1 and the preindustrial control, this would likely be the comparison most likely to show a difference sufficient to exceed our detectability threshold.

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. We would also recommend comparing the extended GeoMIP simulation results to the late 20th century ENSO record. 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.

We now turn to another important issue that hinders detectability of changes in ENSO variability: the ability of current generation GCMs to accurately reproduce ENSO. 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.

Since the ENSO mechanism is so vulnerable to small perturbations, subtle differences in how forcing is applied could impact the ENSO mechanism in the models. Specifically, emissions are imposed differently in the historical runs from how they are imposed in RCP4.5. RCP4.5 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 RCP4.5 ENSO variability could seek to determine if the interannual variability in SSTs is muted enough by this smoothing in RCP4.5 to potentially alter the evolution of ENSO events.

While our results pertaining specifically to changes in ENSO frequency and amplitude reveal that detectability of changes is difficult, the overall trend in Niño3.4 is clear under each potential future scenario. The most important conclusion from analysis of the long-term trend in Niño3.4 is that the warming trend in RCP4.5 would be 62% stronger than in G4, the most realistic geoengineering scenario. Changes in SSTs in Niño3.4 are important because the mean climate of this region is an important factor in enhancing weather and climate trends on multiple time scales. For example, the 0.9 K warming in Niño3.4 over 40 years under RCP4.5 would greatly enhance the amount of precipitable water in the region. Meridional transport of water vapor out of the tropics occurs through relatively narrow regions in the atmosphere. These so-called atmospheric rivers advect moisture into the baroclinic zone and are responsible for many extreme precipitation events in North America and other places.

Superposing a 0.9 K warmer Niño3.4 mean onto a strong El Niño event could easily result in more extreme remote ENSO impacts, such as flooding. Additionally, increased SSTs results in warmer surface air temperatures and a steeper lapse rate. This promotes broad areas of enhanced convection over warm ocean areas, which induce poleward propagating wave motion in the atmosphere, which alter the general circulation and change weather patterns around the world. Would a 0.9 K warming of Niño3.4 cause generally enhanced convection over the Niño3.4 region and would this convection induce detectable changes in the general circulation? Would the warming of only 0.5 K under G4 result in different regional and global warming impacts than the RCP4.5 warming? Certainly it is worthwhile to study how long term trends in particular regions such as Niño3.4 may be closely related to other long term trends in remote places and how these relationships may differ under geoengineering as opposed to AGW.

Lastly, 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, 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.

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.

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

Caldeira, K., and Myhrvold, N. P.: Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. Environ. Res. Lett., 8, 034039, 2013.

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

Cobb, K. M., Westphal, N., Sayani, H. R., Watson, J. T., Di Lorenzo, E., Cheng, H., Edwards, R. L., and Charles, C. D.: Highly variable El Niño-Southern Oscillation throughout the Holocene, Science 339 (6115), 67-70, 2013.

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.

D'Agostino, R. B., Belanger A., and Ralph B.: A suggestion for using powerful and informative tests of normality. The American Statistician 44, 316–321, 1990.

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., and Yang Z.-L.: The Common Land Model (CLM), Bull. Am. Meteorol. Soc., 84, 1013–1023, 2003.

Decarlo, L.T.: On the meaning and use of kurtosis. Psychological Methods 2, 292-307, 1997.

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., and 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., and Haug, G.H.: Volcanoes and ENSO over the past millennium, J. Climate, 21(13), 3134–3148, 2007.

Gergis J. L., and Fowler, A.M.: How unusual was late twentieth century El Niño-Southern Oscillation? Adv. Geosci., 6, 173–79, 2006.

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. Adv. Modeling Earth Sys., 5, 572-597, 2013.

Guilyardi E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Oldenborgh, G. J. V., and 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.

Kaplan, A., Cane, M., Kushnir, Y., Clement, A., Blumenthal, M., and Rajagopalan, B.: Analyses of global sea surface temperature 1856-1991, J. Geophys. Res., 103, 18,567-18,589, 1998.

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.

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. Disc., 8, 4697-4736, doi:10.5194/gmdd-8-4697-2015, 2015.

Li, J., Xie, S.-P., Cook E. R., Morales, M., Christie, D., Johnson, N., Chen, F., D'Arrigo, R., Fowler, A., Gou, X., and Fang, K.: El Niño modulations over the past seven centuries. Nature Climate Change, 3, 822–826, 2013.

Maher, N., McGregor, S., England, M. H., and Gupta, A. S.: Effects of volcanism on tropical variability. Geophys. Res. Lett., 42(14):6024-6033, 2015.

Mann, M. E., Cane, M. A., Zebiak, S. E., and 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.

McGregor, S., Timmermann, A., England, M. H., Elison Timm, O., and Wittenberg, A. T.: Inferred changes in El Niño–Southern Oscillation variance over the past six centuries, Clim. Past, 9, 2269-2284, 2013.

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

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extension from 1765 to 2300, Climatic Change, 109, 213-241, doi:10.1007/s10584-011-0156-z, 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 1: Description and evaluation, Geoscientific Model Dev., 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 Dev., 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. Climate, 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., and 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., and 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., and Cane M. L.: A model El Niño/Southern Oscillation. Mon. Wea. Rev., 115, 2262–2278, 1987.

2208

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

*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)

b. Models in G2

*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

*BNU-ESM (1)
GISS-E2-R (3)
HadGEM2-ES (2)
IPSL-CM5A-LR (1)
MPI-ESM-LR (3)

d. Models in G4

*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)

Table 3. Difference in linear trend of Niño3.4 Index (5°S-5°N, 120°W-170°W) between each experiment G1-G4 and its control.

G1 minus 4xCO2	-0.18 K $decade^{-1}$
G2 minus +1% CO_2 yr^{-1}	-0.24 K $decade^{-1}$
G3 minus RCP4.5	-0.14 K $decade^{-1}$
G4 minus RCP4.5	-0.07 K $decade^{-1}$

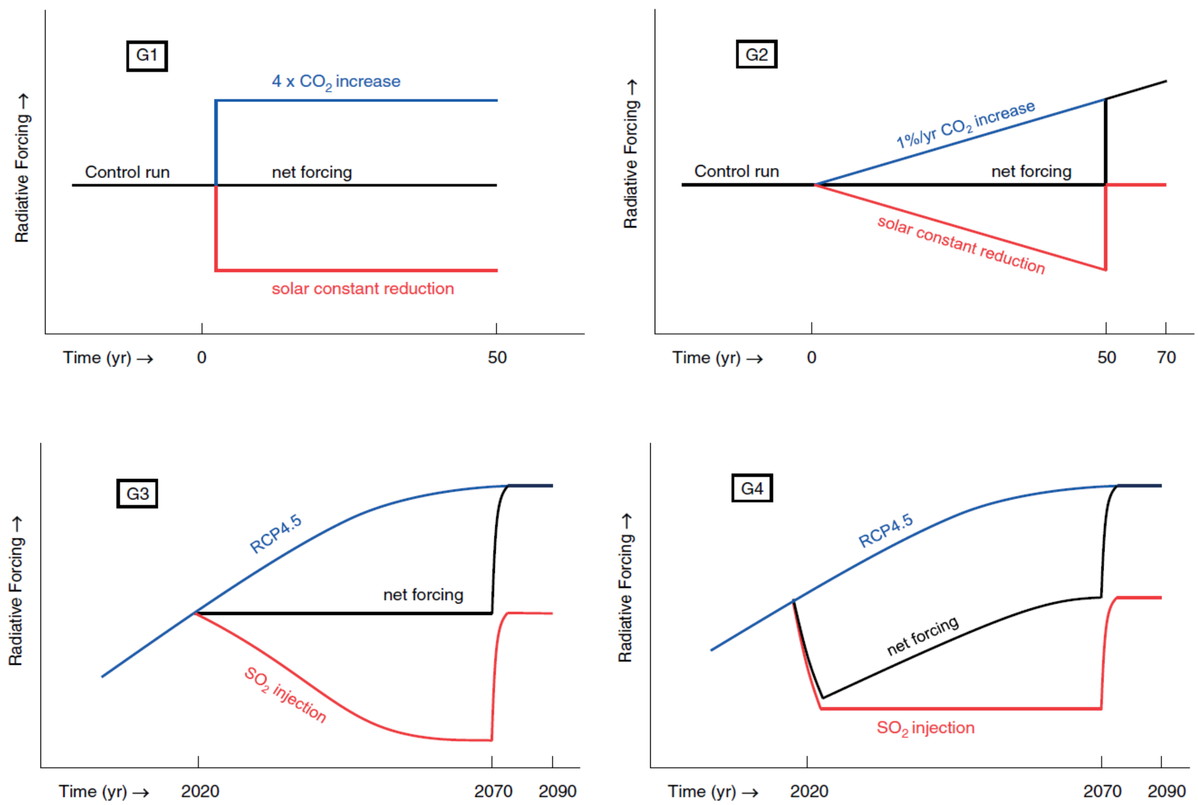
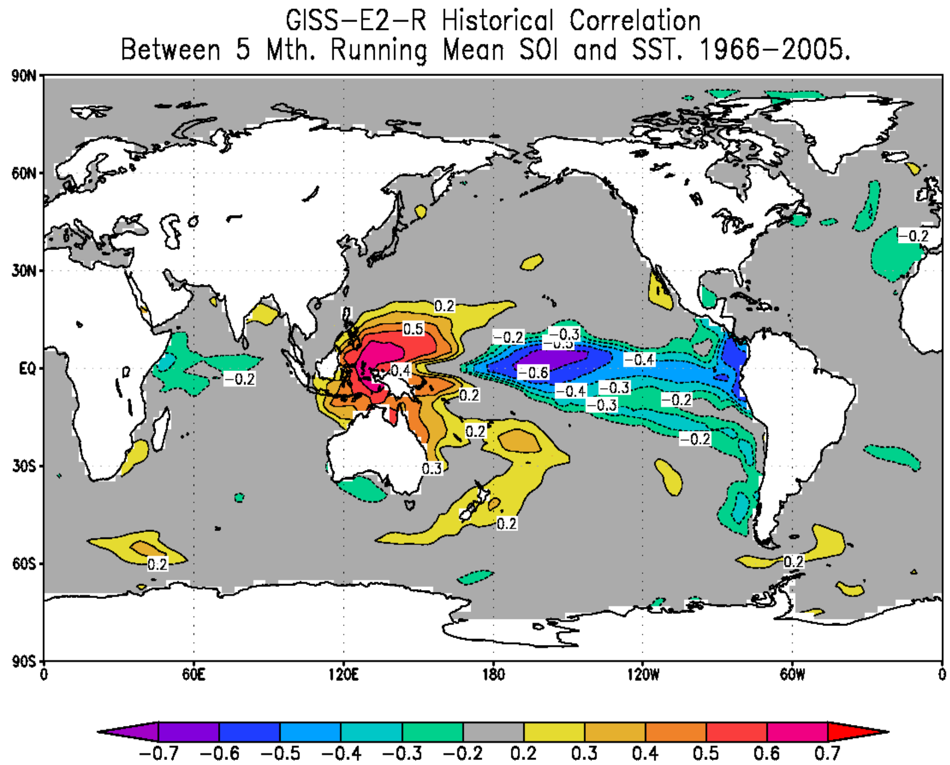


Figure 1. GeoMIP G1-G4 experiment designs. Figures 1-4 from Kravitz et al. (2011).



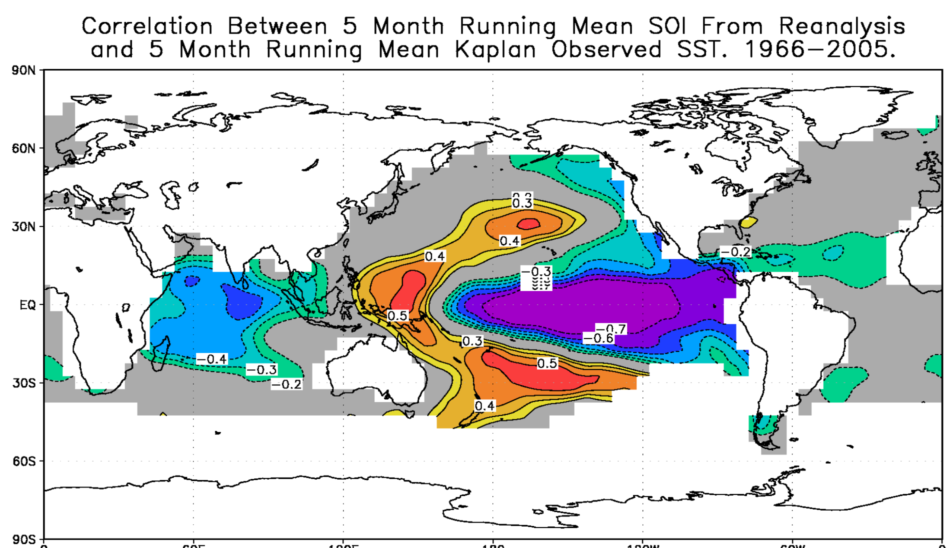


Figure 2. Top panel shows spatial correlation between GISS historical sea surface temperature (SST) and the Southern Oscillation Index (SOI). The area of strong negative correlation is confined to a small region in the central Pacific, relative to the broad area of strong negative correlation in the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

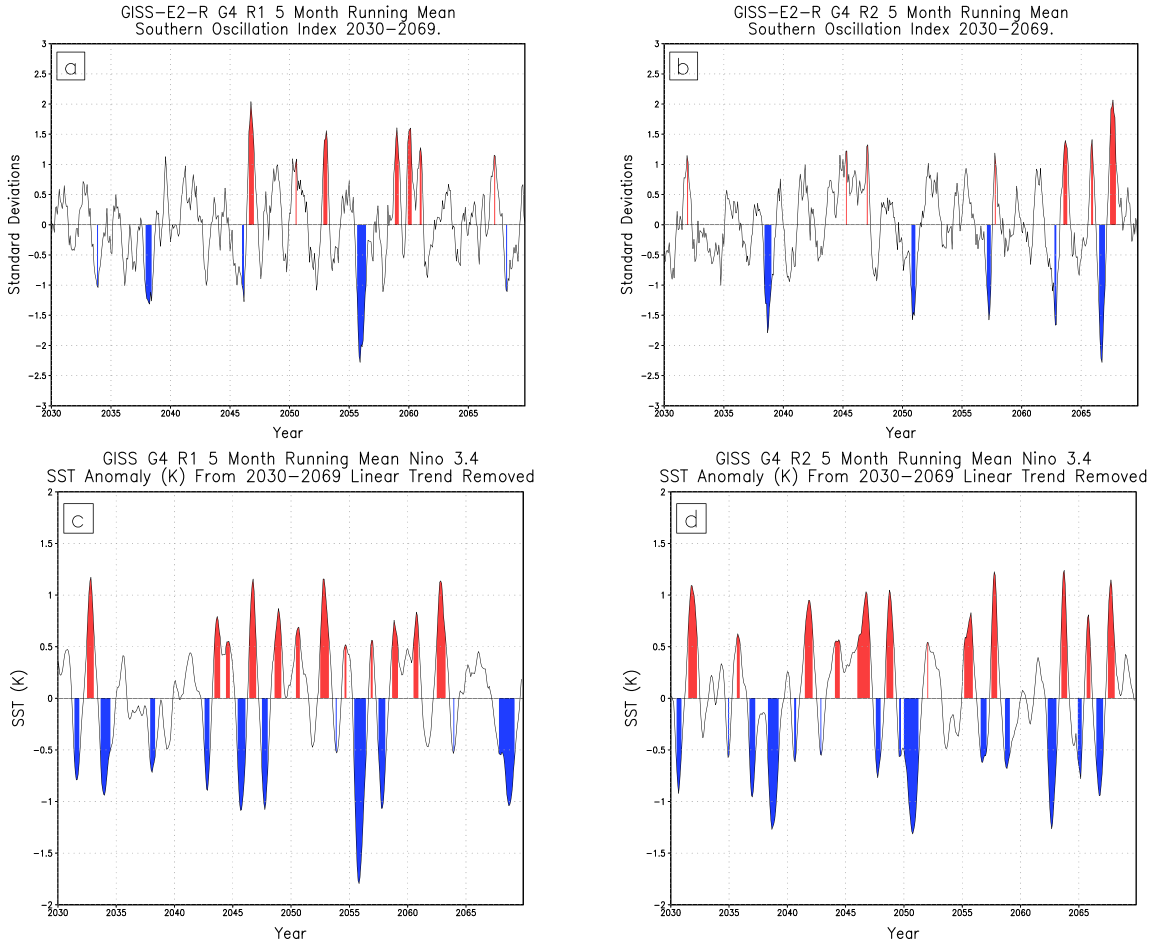


Figure 3. Time series of normalized Southern Oscillation Index (SOI) for (a) GISS G4 Run 1 and (b) GISS G4 Run 2. In the context of SOI, ENSO events are defined as departures of 0.5 standard deviations from zero. SOI warm events are highlighted in red, while cold events are highlighted in blue. No highlight is applied during an ENSO neutral phase. Time series of SST in Niño3.4 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.

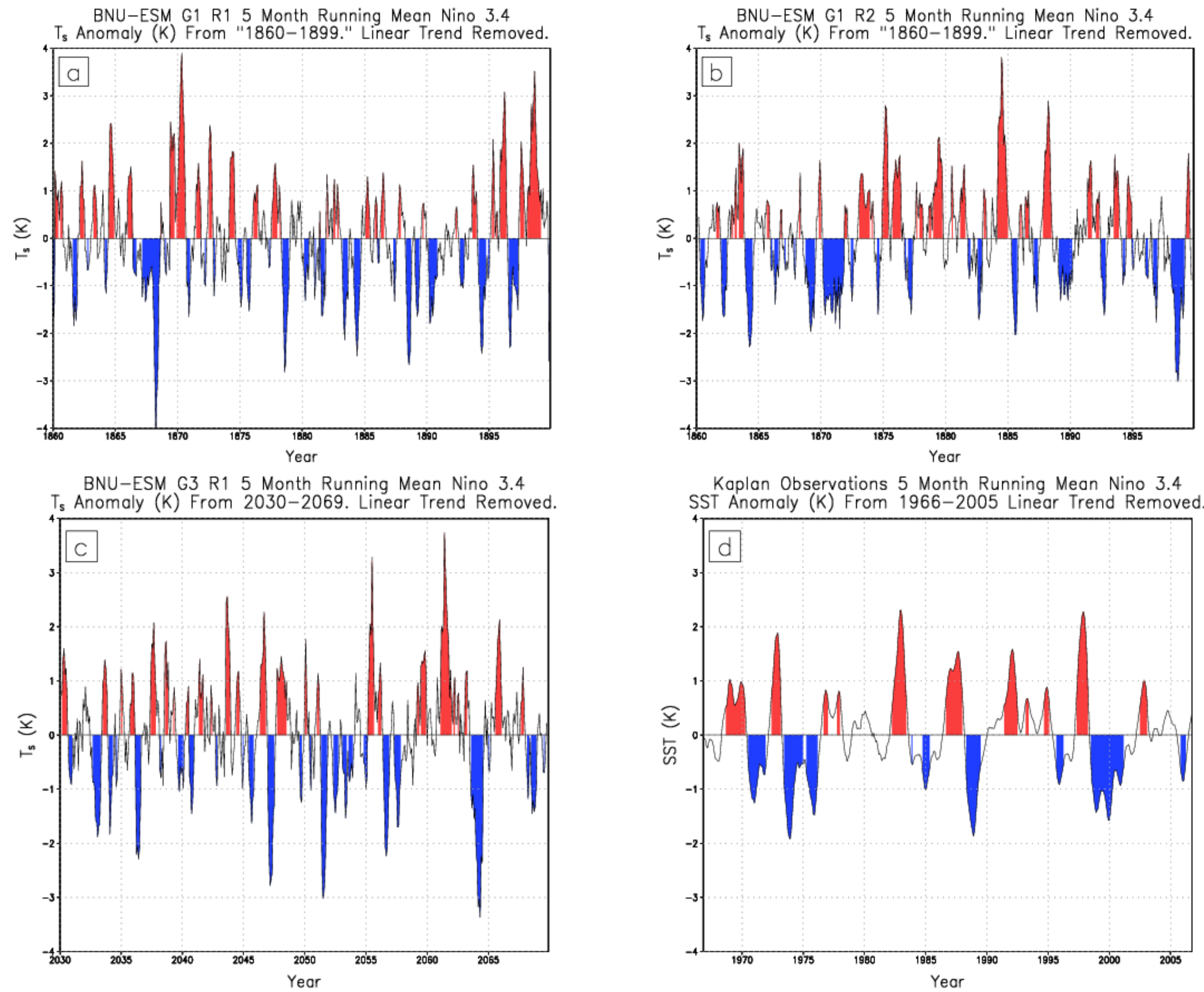


Figure 4. Time series of Niño3.4 anomalies from three experimental runs, (a) G1 run 1, (b) G1 run 2, and (c) G3 run 1, of the BNU-ESM model compared to observations (d). Red coloring indicates ENSO warm events, while blue shading indicates ENSO cold events. The model is excluded due to unrealistic variability and amplitude.

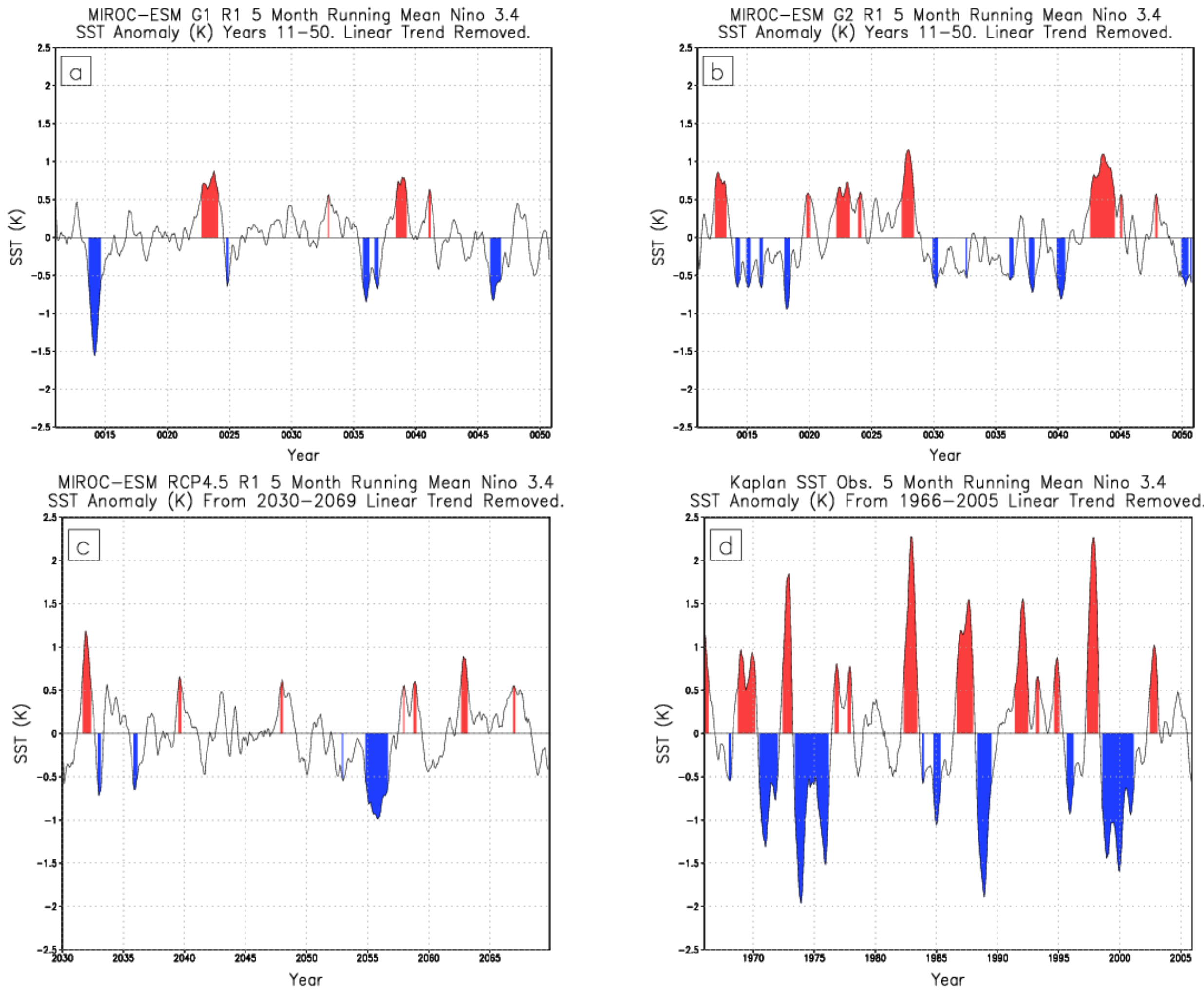
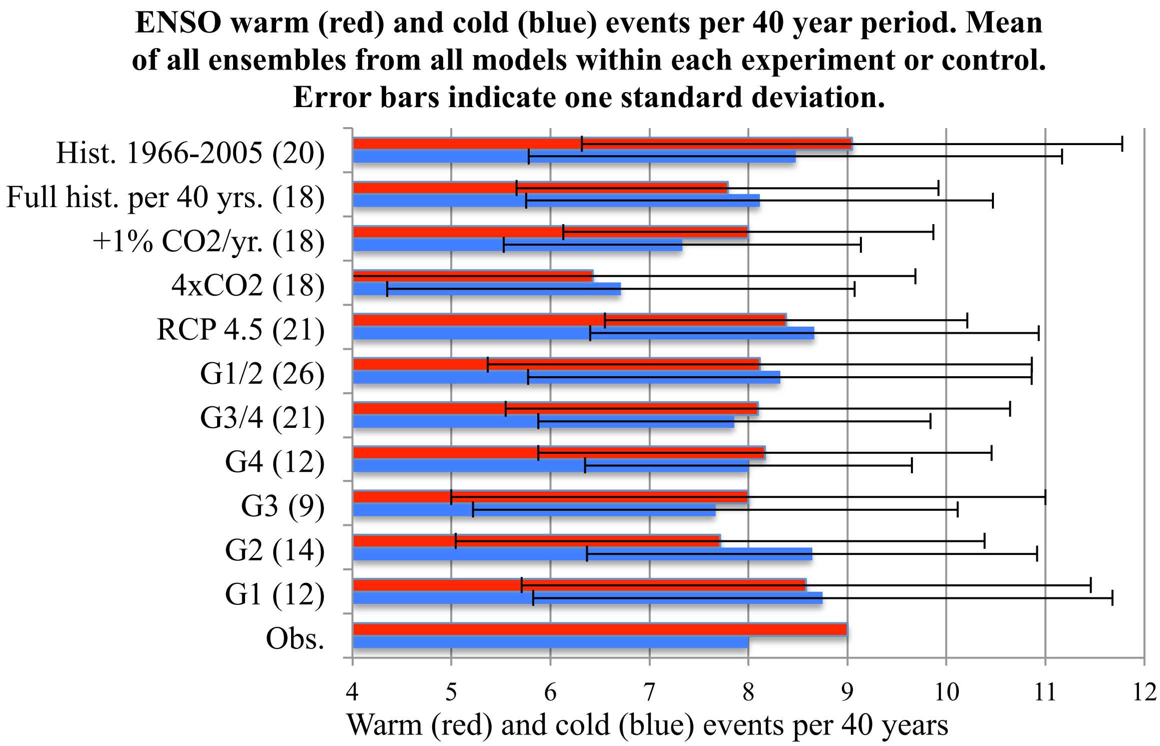
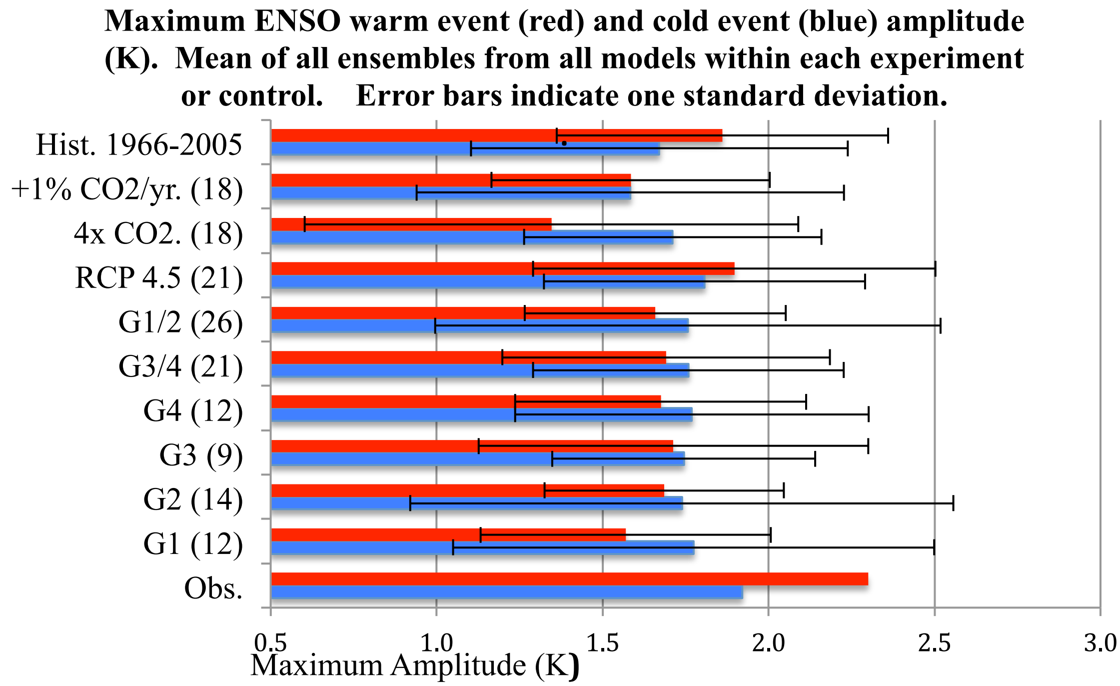


Figure 5. Niño3.4 anomalies for MIROC-ESM. Time series of (a) G1, (b) G2 and (c) RCP4.5 from each model all show significantly muted variability and amplitude compared to (d) observations, with few, if any, warm events exceeding a 1 K anomaly. All other MIROC family experiments showed the same muted 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.



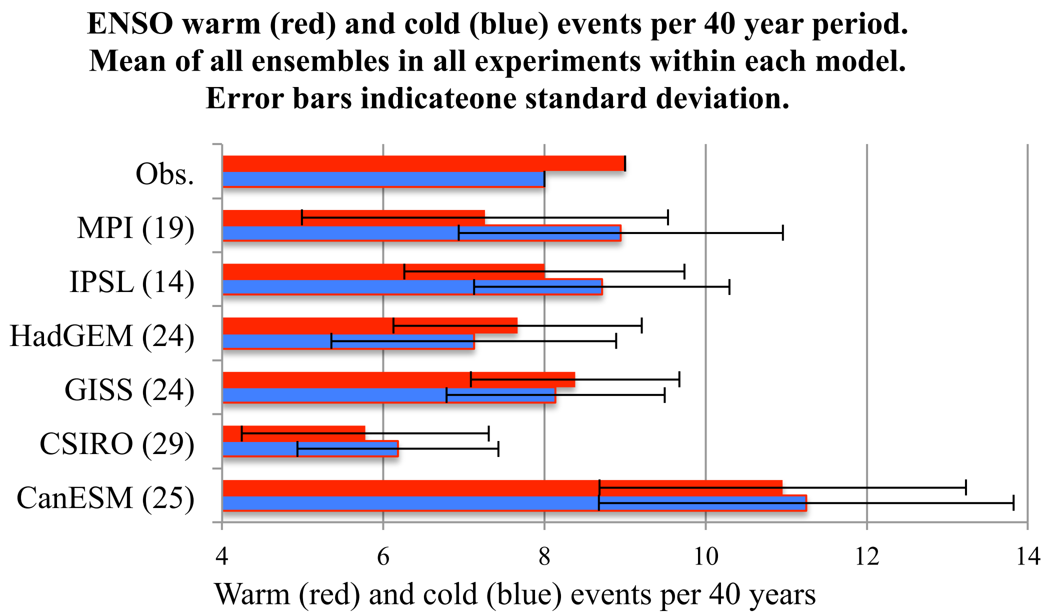
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2 /yr. (18)	1850-2005 /40 yr. (18)	1966-2005 (18)
Warm events	9.00	8.58	7.71	8.00	8.17	8.10	8.12	8.38	6.43	8.00	7.79	9.05
Cold events	8.00	8.75	8.64	7.67	8.00	7.86	8.32	8.67	6.71	7.33	8.11	8.48

Figure 6. Number of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G3, G4, RCP4.5, Years 11-50 for G1, G2, +1% CO_2 yr^{-1} 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.



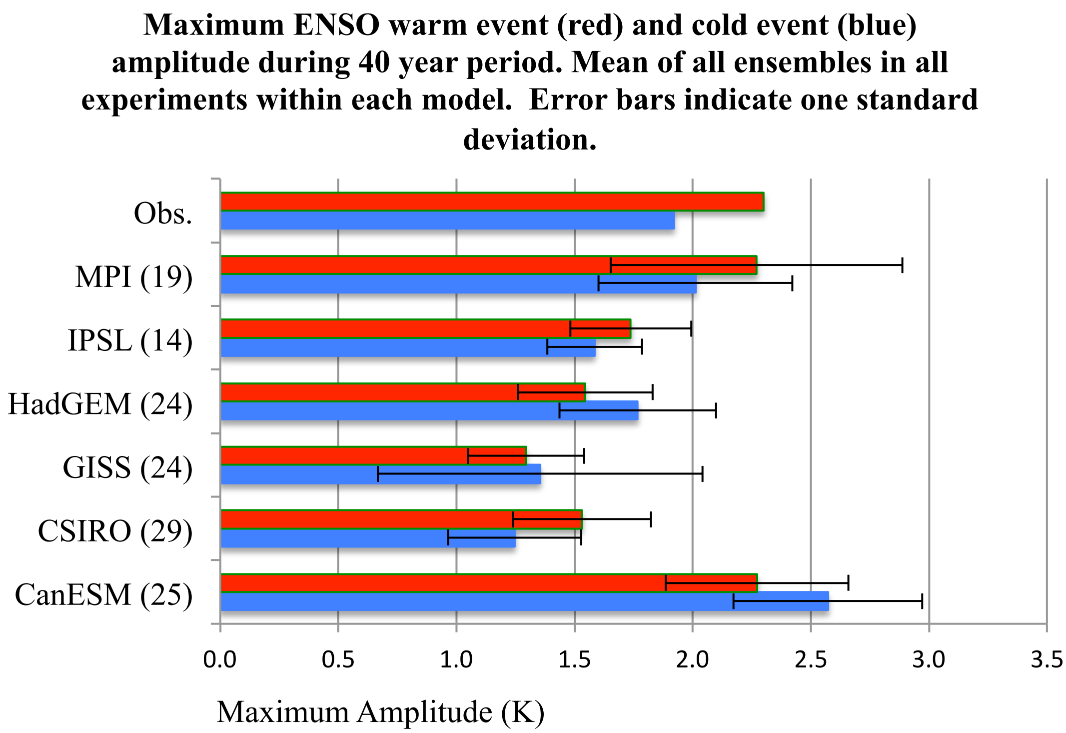
	Obs.	G1 (12)	G2 (14)	G3 (9)	G4 (12)	G3/4 (21)	G1/2 (26)	RCP4.5 (21)	4XCO2 (18)	+1% CO_2/yr. (18)	1966-2005 (18)
Warm amp.	2.30	1.57	1.69	1.71	1.68	1.69	1.66	1.90	1.35	1.58	1.86
Cold amp.	1.92	1.77	1.74	1.74	1.77	1.76	1.76	1.81	1.71	1.58	1.67

Figure 7. Maximum amplitude (K) of ENSO warm (red) or cold (blue) events simulated or observed between 2030-2069 for G3, G4, RCP4.5, Years 11-50 for G1, G2, +1% CO_2 yr^{-1} 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.

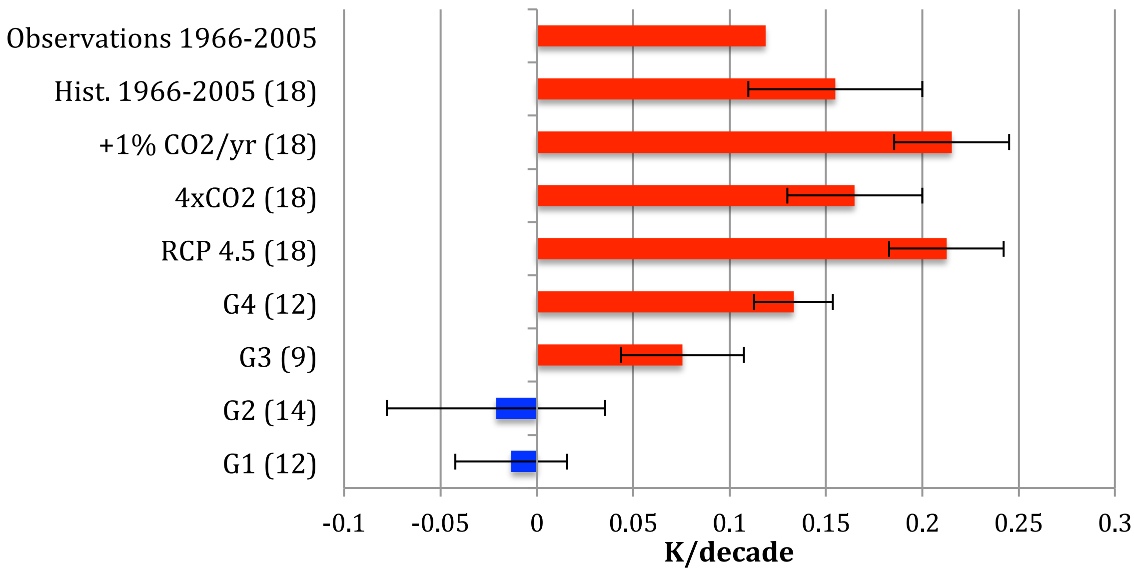


Figure 10. Linear trend in Niño3.4 Index (5°S-5°N, 120°W-170°W) during the applicable 40-year comparison periods. The applicable comparison periods are 1966-2005 for historical simulations; years 11-50 for G1, G2, +1% CO_2 yr^{-1} and 4xCO2; 2030-2069 for G3, G4 and RCP4.5; and 1966-2005 for historical simulations and observations. The values in parenthesis are the number of ensemble members for each experiment. Red bars indicate a warming trend. Blue bars indicate a cooling trend. Error bars indicate one standard deviation.