Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079764

Key Points:
- Internal variability explains 90% of the CMIP5 spread of historical ENSO SST changes and 80% of the spread under stronger forcing
- The large internal ENSO variability means that individual realizations can show very different changes compared to the true forced response
- Only with a large ensemble (more than 30 members) can internal variability in ENSO projections be quantified robustly

Supporting Information:
- Supporting Information S1

Correspondence to:
N. Maher,
nicola.maher@mpimet.mpg.de

Citation:
Maher, N., Matei, D., Milinski, S., & Marotzke, J. (2018). ENSO change in climate projections: Forced response or internal variability? *Geophysical Research Letters*, *45*, 11,390–11,398. https://doi.org/10.1029/2018GL079764

Received 27 APR 2018
Accepted 8 OCT 2018
Accepted article online 11 OCT 2018
Published online 27 OCT 2018

©2018. American Geophysical Union. All Rights Reserved.

ENSO Change in Climate Projections: Forced Response or Internal Variability?

N. Maher[1,2], D. Matei[1], S. Milinski[1,3], and J. Marotzke[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Alexander von Humboldt-Stiftung, Germany, [3]International Max Planck Research School on Earth System Modelling, Hamburg, Germany

Abstract Two large ensembles are used to quantify the extent to which internal variability can contribute to long-term changes in El Niño-Southern Oscillation (ENSO) characteristics. We diagnose changes that are externally forced and distinguish between multi-model simulation results that differ by chance and those that differ due to different model physics. The range of simulated ENSO amplitude changes in the large ensemble historical simulations encompasses 90% of the Coupled Model Intercomparison Project 5 historical simulations and 80% of moderate (RCP4.5) and strong (RCP8.5) warming scenarios. When considering projected ENSO pattern changes, model differences are also important. We find that ENSO has high internal variability and that single realizations of a model can produce very different results to the ensemble mean response. Due to this variability, 30–40 ensemble members of a single model are needed to robustly compute absolute ENSO variance to a 10% error when 30-year analysis periods are used.

Plain Language Summary The El Niño-Southern Oscillation (ENSO) is the dominant driver of interannual variability globally, with effects that are felt all over the world. As such it is important to understand whether ENSO might change in the future or has already changed in the recent past due to anthropogenic emissions. We show that ENSO strength is highly variable between simulations from a single model, independent of external forcing. This variability is known as internal variability and occurs due to the chaotic nature of the climate system. Such variability can cloud our projections of the future when we have limited model simulations. Here, we demonstrate that <30 simulations of the same model are needed to robustly estimate ENSO variability. Using the 100 possible futures simulated in the Max Planck Institute Grand Ensemble (MPI-GE) and 40 possible futures from the Community Earth System Large and Medium Ensemble Projects (CESM-LE/CESM-ME) we find that ENSO variability is large. Here, this strong variability will likely mask any possible observed changes, meaning that we are unlikely to be able attribute ENSO changes the near future to anthropogenic forcing.

1. Introduction

The El Niño-Southern Oscillation (ENSO) is the dominant driver of interannual variability globally and has effects which are felt in many remote regions of the world. As such it is vital to assess the potential future changes of ENSO; however, model simulations of future ENSO changes differ widely (Bellenger et al., 2014; Christensen et al., 2013; Stevenson, 2012). Here we use two large ensembles each with a single climate model, the Max Planck Institute Grand Ensemble (MPI-GE; Bittner et al., 2016; Stevens, 2015) and the Community Earth System Large and Medium Ensemble Projects (CESM-LE/CESM-ME; Kay et al., 2015; Sanderson et al., 2015), to quantify the extent to which internal variability can contribute to long-term changes in ENSO characteristics and to evaluate forced changes in ENSO projections in both models.

There is little consensus on how ENSO sea surface temperature (SST) may change in a future with increasing greenhouse gas emissions and underlying warming (Bellenger et al., 2014; Collins et al., 2010; Guilyardi et al., 2012; Ham & Kug, 2016), with large differences found between different Coupled Model Intercomparison Project 5 (CMIP5) model projections (Collins et al., 2010; Guilyardi et al., 2012). Many studies have addressed this question (e.g., Cai et al., 2014, 2015; Keupp et al., 2017; Kim et al., 2014; Power et al., 2013; Rashid et al., 2016; Takahashi et al., 2011; Timmermann, 1999), with robust results indicating a forced change in ENSO found

in multiple studies, which investigate the effects of precipitation indices rather than SST (Cai et al., 2014, 2015; Chung & Power, 2016; Power et al., 2017, 2013).

The range of projections of ENSO in the future could be due to differences in model physics, resulting in different projections from different models. However, the role of internal variability must also be considered, which occurs due to the chaotic nature of the climate system. Such internal variations can result in different projections from single ensemble members of the same climate model. Previous studies have highlighted the importance of internal variability in the climate system (Deser et al., 2012; Marotzke & Forster, 2015; Stevenson, 2012) and demonstrated that for a climate signal to emerge it must overcome the influence of the internal variations (Hawkins & Sutton, 2009, 2012).

ENSO itself is not only an interannual phenomenon and has been shown to have strong variability on interdecadal and decadal time scales (Burgman et al., 2008; Chowdary et al., 2012; Li et al., 2011; Ogata et al., 2013). For example, Wittenberg (2009) used a 2000-year run of the GFDL CM2.1 model to show highly irregular ENSO amplitude and period in different multidecadal epochs. This suggests that a large ensemble or a long equilibrated run of a single model in combination with projections from the same model are needed to investigate the question of how ENSO changes in the future (Stevenson, 2012; Wittenberg, 2009). In a large ensemble of a single model the forced changes can be cleanly separated from internal variability, unlike in large multimodel ensembles such as CMIP5 (Frankcombe et al., 2015).

Internal variability in a changing climate manifests itself through the difference between single ensemble members of the same climate model. Changes in ENSO have previously been investigated by Zelle et al. (2005) in a 62-member ensemble of Community Climate System Model version 1.4. They found no discernible change in ENSO amplitude, period, or spatial patterns in the business-as-usual scenario. However, the authors emphasized flaws in the model representation of ENSO that they believed severely limited their results. Zheng et al. (2018) used the 40-member CESM-LE ensemble, in the RCP8.5 scenario. They showed an ensemble mean increase in ENSO amplitude from 1.21 (1950–1999) to 1.32° C (2046–2095), which is within the internal variability of the estimated amplitudes (1950–1999) in their study. While Zheng et al. (2018) found variability in CESM-LE similar to the variability in CMIP5, they recommended that changes in ENSO amplitude should also be investigated in other large ensembles due to the overestimation of mean ENSO amplitude in the CESM-LE ensemble.

In this study we compare ENSO amplitude projections from two large ensembles of single climate models (see section 2) to CMIP5 to assess whether the variations in future projections can be attributed to internal variability, model differences, or a combination of both factors. We then use the two ensembles to address whether ENSO will change in the future, beyond the internal variability that we can robustly quantify. Finally, we address the question of how large an ensemble is needed to capture ENSO variability.

2. Data and Methodology

In this study we use two large ensembles. The first is MPI-GE, which is currently the largest ensemble (100 members) of a state-of-the-art comprehensive climate model (Bittner et al., 2016; Stevens, 2015). The second large ensemble used in this study is the CESM-LE (Kay et al., 2015); we note that we also use the medium ensemble CESM-ME for the RCP4.5 scenario (Sanderson et al., 2015), but for ease of description will refer to the total ensemble of historical, RCP4.5, and RCP8.5 as CESM-LE from this point forward. Model details and an evaluation of how both models behave in the ENSO region can be found in the supporting information. From the ensembles, we define the following greenhouse gas forcing regimes and the simulations upon which they are based.

- *No warming*: Preindustrial control (2000 years MPI-GE, 1700 years CESM-LE, 1850 radiative conditions, 1 realization).
- *Weak warming*: Historical (1850–2005, observed radiative forcing, 100 realizations MPI-GE, 1920–2005, 42 realizations CESM-LE)
- *Moderate warming*: RCP4.5 (2006–2100, RCP4.5 radiative forcing, 100 realizations MPI-GE, 2006–2080, 15 realizations CESM-ME)
- *Strong warming*: RCP8.5 (2006–2100, RCP8.5 radiative forcing, 100 realizations MPI-GE, 40 realizations CESM-LE)

For comparison to the ensemble simulations we create surrogate ensembles from the preindustrial control simulations. Here nonoverlapping ensembles of length 30 years are created. 30-year analysis periods can be used given the large size of the ensembles, here the longer scale variability is captured by making computations across the ensemble-dimension. Due to length restrictions, only 67 nonoverlapping members can be created for MPI-GE. To create the additional 34 members we randomly resample the preindustrial control. The 30 years for each surrogate ensemble are taken in sequence from the preindustrial control to keep the correct autocorrelation in the time series. For MPI-GE, a 100 member ensemble is created, while for CESM-LE and CESM-ME, a 42 member ensemble is created.

An important advantage of using a large ensemble created from a single model is that variability patterns can be computed over the ensemble-dimension, rather than the traditional time-dimension, that is, computing an Empirical Orthogonal Function (EOF) across the individual members at each time step. When using the ensemble-dimension to compute EOFs (EOF-E), all time scales of variability are captured. This means that if the ensemble is large enough, changes in internal variability over time as captured by EOF-E can be interpreted as due to the forcing alone. This is not possible when using EOFs computed across the time-dimension (EOF-T), making EOF-E a powerful method of analysis. We note that when using EOF-E we are unable to distinguish at which time scales ENSO changes occur as all time scales are captured. This study is the first to demonstrate this method by computing EOFs for each year of the experiment.

The first two EOF's are computed first for each ensemble member (EOF-T) for comparison with observations, where the time-dimension must be used. We then compute the first two EOFs over the ensemble for each year (EOF-E) to quantify transient changes under multiple forcing regimes. In both cases the patterns are normalized by their own spatial standard deviation, rendering them unitless. The normalization means that pattern changes are not affected by the ENSO strength and that the principal components can be used to quantify changes in ENSO strength (standard deviation of the PC time series). Here, we are able to separate changes in ENSO pattern and strength using this methodology. The true expression of changes in ENSO in the SST fields can be determined by multiplying the pattern change by the strength change and then adding the two EOF components.

In this paper we interpret EOF1 and EOF2 (in both time and ensemble-dimension) in the following way, based on the work of Takahashi et al. (2011). We interpret EOF1, which explains 30–40% of the variability in the Pacific, as the base pattern for ENSO. EOF2 (7% of the variance) can then be added or subtracted to EOF1 as in Takahashi et al. (2011) to cause individual events to express themselves differently, contributing to ENSO diversity. Here, we assess the variability in the Pacific using EOFs computed over the region 30°S–30°N and 100°E–70°W. To complete this analysis the forced response and the seasonal cycle are removed by subtracting the ensemble mean from each ensemble member before the EOF is computed. Three observational SST data sets are also used in this study, namely HadISST (Rayner, 2003), COBE SST (provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, United States, from their Web site at https://www.esrl.noaa.gov/psd/), and ERSST.v4 (Smith et al., 2008). For the observational data a second-order polynomial fit is applied over the period 1891–2015 and removed as an approximation for the forced trend.

The standard Niño indices are also used in this analysis. Niño3 is defined as the average of the region 5°S–5°N and 150–90°W. Niño4 is defined as the average of the region 5°N–5°S and 160°E–150°W. ENSO amplitude is taken as the standard deviation of the Niño3 index. We compare the change in ENSO amplitude in the two large ensembles against the CMIP5 models (a list of the CMIP5 models used in this study can be found in Table S2) Additionally, we use a CMIP5 subset consisting of four smaller ensembles, EC-EARTH (13 members for both historical and RCP4.5 and 12 for RCP8.5), GISS-E2-R (18 historical members and 17 RCP4.5 members), CNRM-CM5 (10 historical members), and CSIRO-Mk3-6-0 (10 members for historical, RCP4.5 and RCP8.5).

3. Separating Internal Variability From Forced Changes

The two large ensembles, MPI-GE and CESM-LE, can be used to assess whether the diversity in projected ENSO amplitude found in CMIP5 is due to internal variability or differences in the model physics. In Figure 1a we see that the stars, which illustrate the largest positive and negative amplitude changes for each ensemble, are as large for both ensembles (red and blue) as for the CMIP5 5th and 95th percentiles (black diamonds). This strongly suggests that internal variability can explain 90% the CMIP5 diversity in the weak warming regime. When considering the moderate and strong warming scenarios (RCP4.5 and RCP8.5; Figures 1b and 1c), the ensembles encompass the 10th and 90th percentiles of the CMIP5 range in projections (black crosses). Here,

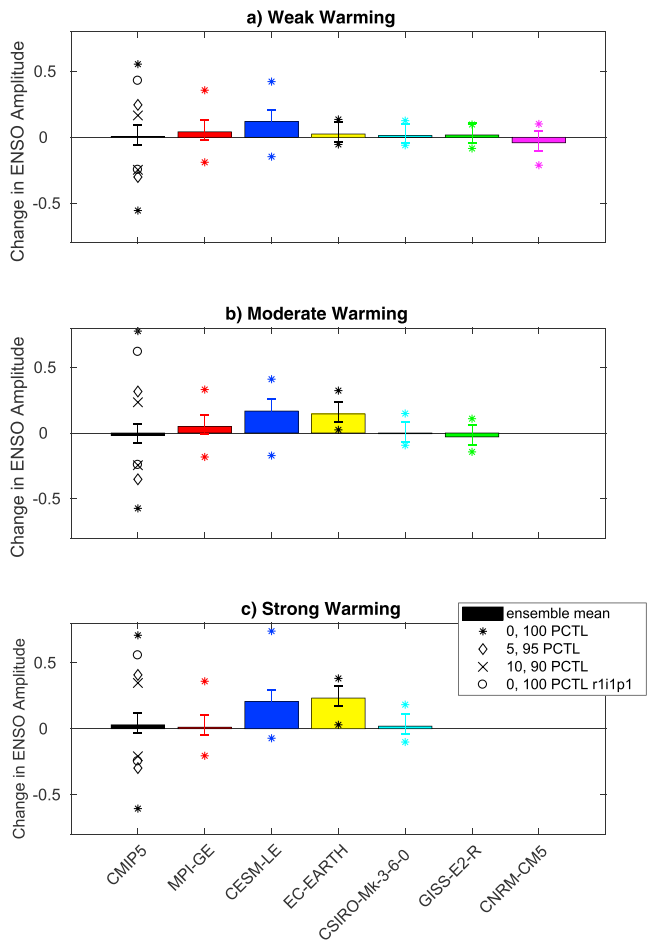

Figure 1. Ensemble mean (solid bars), 0th and 100th percentile (stars), 5th and 95th percentile (diamonds), 10th and 90th percentiles (crosses) of the change in Niño3 amplitude (standard deviation of sea surface temperature in the Niño3 region) compared to the preindustrial control amplitude. Shown for CMIP5 models (all ensemble members; black, 0th and 100th percentile for r1i1p1; black circles), MPI-GE (red), CESM-LE (blue), EC-EARTH (yellow), CSIRO-Mk-3-6-0 (pale blue), GISS-E2-R (green), and CNRM-CM5 (pink). Ensemble sizes can be found in the methods. Shown for (a) Historical (1975–2005), (b) RCP4.5 scenario (2050–2080), and (c) RCP8.5 scenario (2050–2080). The preindustrial control amplitude is estimated by taking the average standard deviation from the last eight independent 30-year periods of each model's control run and averaging these. We choose eight periods to correspond with the shortest preindustrial control simulation from CMIP5. The error bars show the variability in the preindustrial control. This variability on the mean estimate is found by using the 2000-year control run from the MPI-GE and estimating the possible changes in the mean estimate due to the control simulations being not long enough to capture the variability of ENSO. The error is the difference between using eight periods versus the whole control simulation (60 periods).

the ensembles can account for 80% of the CMIP5 diversity, with model physics responsible for the rest of the differences. Due to the lack of large ensemble simulations in CMIP5, we are unable to determine whether the different model physics results in a difference in the individual models internal variability or ensemble mean forced response.

The circles in Figure 1 demonstrate the range of CMIP5 results that can be found using just the first (r1i1p1) ensemble member from each model. The difference between this result and the full model spread clearly demonstrates the weakness of using a single member from each model of a multi-model ensemble. Our results differ from Zheng et al. (2018), who found that internal variability in CESM-LE could explain the CMIP5 model spread in the RCP8.5 scenario. This difference is due to the addition of nine extra CMIP5 models in our comparison and the use of all ensemble members available rather than just r1i1p1.

The forced change in ENSO amplitude can be identified by computing the ensemble mean of the ENSO amplitude changes (Figure 1; solid bars). For CMIP5, MPI-GE, and four smaller ensembles the change in ensemble mean ENSO amplitude in the weak warming scenario is within the variability estimated from the MPI-GE preindustrial control (shown as errorbars), however, CESM-LE shows an increase in amplitude (Figure 1a). In the moderate and strong warming scenarios both CESM-LE and EC-EARTH show positive changes, with an increase in the variability of CESM-LE seen only in the strong warming scenario, while the other ensembles show no significant change (Figures 1b and 1c). Here, the best estimate from the ensembles available is that the forced response of ENSO amplitude in the Niño3 region to warming is likely zero or positive, but not negative.

4. ENSO Change Under Different Forcing Regimes

We now investigate changes in the ENSO SST patterns under differing greenhouse gas forcing regimes. We find that in the weak-forcing regime, there are no significant changes in either EOF1-E or EOF2-E patterns in MPI-GE (Figures 2e, 2g, 2m, and 2o). In CESM-LE a weak pattern change in EOF1-E occurs at the end of the weak-forcing regime (Figure 2f), however while this change is significant (Figure 2n) it is not meaningful as it dips back well below the significance threshold after this point. This may be caused by the ensemble being too small or influences of the Pinatubo eruption, which has been shown to impact ENSO (e.g., Maher et al., 2015). Changes in EOF2-E are also seen (Figure 2h), however, the lack of trend in the pattern correlation for EOF2-E suggests that these changes are also meaningless, although at some time points the correlation fluctuates outside the significance bands (Figure 2p).

The strong forcing scenario shows significant pattern changes in EOF1-E in both large ensembles (MPI-GE, Figures 2i and 2m; CESM-LE, Figures 2j and 2n); however, the changes oppose each other, demonstrating a disagreement of the two models in projecting this future change. EOF2-E shows a large significant change in both models (MPI-GE, Figures 2k and 2o; CESM-LE, Figures 2l and 2p). While the changes in EOF2-E in the two models exhibit spatial differences, particularly in the equatorial central Pacific, both overall oppose the EOF2-E mean pattern. This suggests that in a strong warming scenario the canonical and central Pacific ENSO patterns could become more similar, demonstrating a decrease in ENSO diversity. We note that while these pattern changes are significant, they will not necessarily occur in a single realization.

To determine whether ENSO itself changes, both the change in EOF pattern and strength (standard deviation of PC time series) must be quantified. First, we investigate whether observations of ENSO strength, expressed

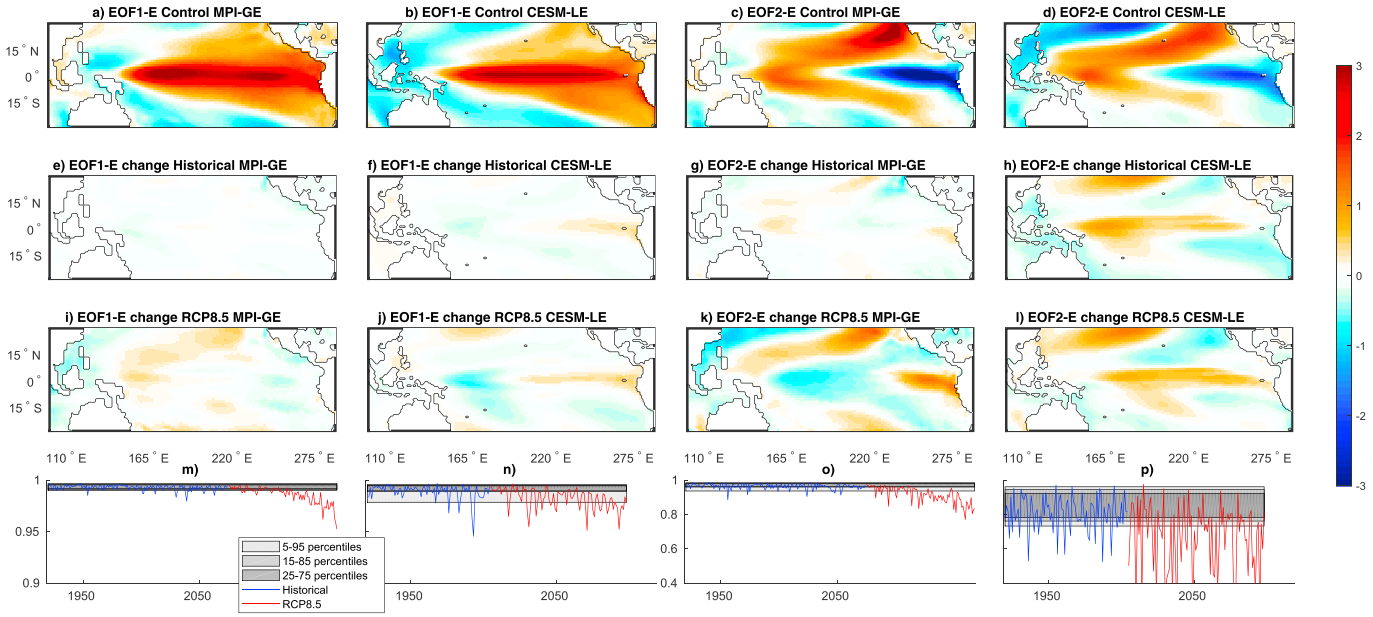

Figure 2. EOF-E pattern for the surrogate preindustrial control ensemble for (a) EOF1 MPI-GE, (b) EOF1 CESM-LE, (c) EOF2 MPI-GE, and (d) EOF2 CESM-LE. Difference in EOF-E patterns and the surrogate preindustrial control pattern for the historical regime (1975–2005) for (e) EOF1 MPI-GE, (f) EOF1 CESM-LE, (g) EOF2 MPI-GE, and (h) EOF2 CESM-LE and for RCP8.5 (2070–2100) for (i) EOF1 MPI-GE, (j) EOF1 CESM-LE, (k) EOF2 MPI-GE, and (l) EOF2 CESM-LE. Note that the patterns are normalized by the overall pattern standard deviation, rendering them unitless. A pattern correlation between the mean of the surrogate preindustrial control ensemble and the EOF-E pattern at each time point is computed to determine significance for the pattern changes. This correlation is shown for (m) EOF1 MPI-GE, (n) EOF1 CESM-LE, (o) EOF2 MPI-GE, and (p) EOF2 CESM-LE. The gray bars show the percentile range when each of the surrogate preindustrial control member is compared to the mean of the surrogate preindustrial control ensemble. This gives a range of significance levels for the pattern changes.

as EOF-T and ENSO amplitude from the Niño3 and 4 regions fall within the range of internal variations of the historical simulation for both models, to determine whether there are any forced changes in ENSO seen in observations. The large range of variability that can be found in a single model is shown in Figures 3a and 3b. Here, it is clear that the range of model internal variability on 30-year time scales is larger than changes seen in the observations and that different observation data sets themselves show a range of results. We conclude that observed changes could be due to internal variability alone and certainly cannot be attributed as forced changes.

We next investigate ENSO amplitude and strength changes in the weak and strong forcing scenarios (Figures 3c and 3d). The large range of changes seen in single ensemble members demonstrates the large internal variability of ENSO and highlights the danger of using single realizations to make ENSO projections. An advantage of using large ensembles is that the ensemble mean changes can be used to quantify the true forced changes in ENSO strength (shown in the bold symbols). These changes are small in MPI-GE, however CESM-LE shows an increase in the strength of both PC1-E and PC2-E in both warming scenarios, with a greater increase in the strong warming scenario (Figure 3c). Given the accurate quantification of variability, made possible by the large ensembles, this result can be attributed to model disagreement in the forced response and calls for further study into why these differences are found.

The combination of the EOF-E pattern and PC-E strength changes is expressed in the Niño amplitudes (amplitudes: Figure 3d, maps: Figure S3). Here, the ensemble mean shows limited changes in MPI-GE, consistent with the limited pattern changes, with a small increase in the ENSO amplitude in the Niño4 region and a small decrease in Niño3 region due to the pattern change in the strong warming scenario. We note that north of these regions MPI-GE shows larger changes. CESM-LE shows an increase in both Niño regions in both the weak and strong warming scenarios. This is due to an increase in PC-E strength combined with the pattern changes

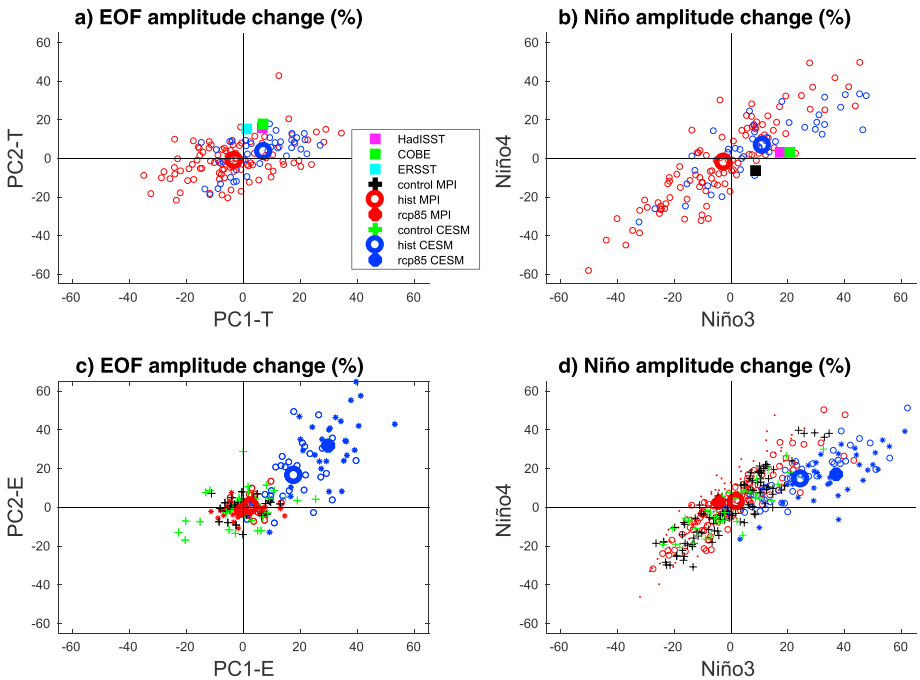

Figure 3. (a) ENSO strength (standard deviation of the PC time series) and b) Niño amplitude (standard deviation of the Niño indices) difference expressed as a percentage of the mean of the period 1945–1975, between the period 1975–2005 and 1945–1975 for the historical scenario and three observational data sets (HadISST, COBE, and ERSST). These periods were chosen due to the large discrepancies between observational data sets pre-1940. (c) ENSO strength and (d) Niño amplitude difference expressed as a percentage of the mean of the control between the last 30 years of each forcing regime (historical 1975–2005, RCP8.5 2070–2100) and the preindustrial control ensemble. The ensemble means are shown in the large bold symbols. Results are shown for both MPI-GE and CESM-LE.

(Figure 2 consistent with result shown in Figure 1). These results demonstrate how the Niño responses can be attributed to pattern changes, strength changes, or a combination of the two using the EOF-E approach.

5. How Many Ensemble Members are Enough?

Given that internal variability can explain a large portion of the diversity in CMIP5 projections, we investigate how many ensemble members are needed to accurately capture ENSO SST variability. To complete this analysis we can only use MPI-GE because CESM-LE is not large enough (see Figure S4 for a qualitative comparison with CESM-LE at low ensemble numbers). The errors in the spectral peak should not affect this calculation as we choose analysis periods of greater than 10 years so that the analysis period is greater than the ENSO period in the model. We use the variance in the Niño3 and Niño4 indices as proxies for ENSO variability. As there is little change in the ensemble variance of the indices over the historical period we assume constant variability throughout the historical period. Figures 4 and 4b show the maximum and minimum estimates of variability as a function of ensemble size and analysis period. As expected, the error in the variability estimate decreases both with length of the time period and ensemble size. This result is in agreement with Wittenberg (2009) who stated that longer periods can be used to compensate for low numbers of ensemble members. We note, however, that caution needs to be taken using such an approach if the internal variability itself is changing over time. If the variability itself changes over time, then longer periods cannot be used to compensate for low numbers of ensemble members, in this case shorter averaging periods and more members must be used.

Figures 4c and 4d show the number of ensemble members needed to be sure that the variability estimate is within 5%,10%, and 20% of the true variability (here taken as the estimate using 155 years and 100 ensemble members; this is possible because there is no change in variability in MPI-GE over the historical period). We see that for a study that wishes to estimate the ENSO variability with a 10% error, 30–40 members are sufficient for most analysis periods. To get within 5% error, however, much longer analysis periods or larger ensemble size is needed. Eighty members are needed for the 30-year analysis period, while 30 members are sufficient if 50-year windows are used. This demonstrates that both the methodology and ensemble size are important

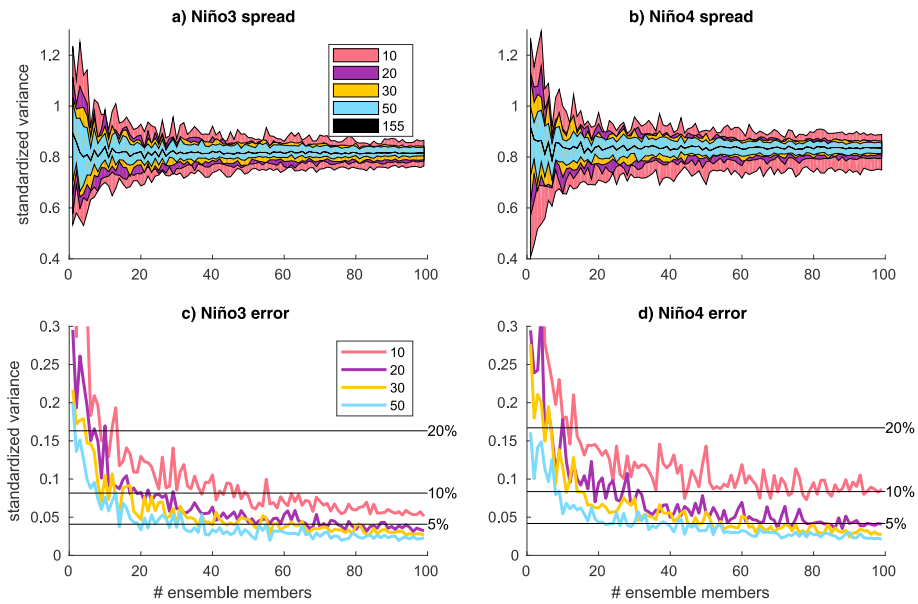

Figure 4. Standardized variance in the Niño3 and Niño4 indices calculated for varying ensemble size and analysis period (colors) for Max Planck Institute Grand Ensemble. The variance is calculated for the detrended historical simulation. We then average over the ensemble members. Panels (a) and (b) show the variance found by randomly sampling the historical ensemble in both the ensemble- and time-dimensions 100 times each (note that resampling is not used). Panels (c) and (d) show the maximum difference between the real variance and the estimated variance using the random sampling method. The real variance is taken as the estimate that uses the maximum averaging length and ensemble size (155 years and 100 members, respectively) to give the closest possible result to the model truth. The black solid lines show when the estimated variance is within 5%, 10%, and 20% of the real variance.

to obtain a robust estimate of ENSO variability in a model. These numbers are slightly higher for Niño4 than Niño3. This is likely due to the cold tongue bias that results in higher variability in the Niño4 region and hence a greater range of possible variance in this region. We note that these results are robust to all forcing scenarios, however they could be model dependent.

While previous studies have investigated the number of ensemble members needed to detect changes in ENSO amplitude in a specific scenario (Zheng et al., 2018) we evaluate the number of members needed to estimate ENSO variance itself. In Figures 4c and 4d we clearly show that minimum of 30–40 members are needed to obtain a realistic estimate of ENSO variability. This result agrees well with the variability captured by the six ensembles (Figure 1a), where only the two large ensembles (>40 members) are able to capture 90% of variability seen in the CMIP5 models in the historical simulation. This also explains why Sun et al. (2018) find that the CESM Last Millennium Ensemble Project 10 member ensemble can only account for 50% of the CMIP5 historical spread in Niño3 amplitude. In this case their ensemble is too small to fully capture ENSO variance, although the CMIP5 spread in this study is likely also overestimated due to a lack of detrending before the standard deviation is calculated.

6. Conclusions

In the historical simulation the 5th and 95th percentiles of the ENSO amplitude changes simulated by CMIP5 are encompassed by the range of changes simulated in MPI-GE and CESM-LE. This means that internal variability, as simulated by the two models, can explain 90% the CMIP5 range. In the moderate and strong warming scenarios, internal variability can account for 80% of this range, with differences in model physics likely responsible for the additional differences. While previous studies have suggested that model differences matter, they have not been able to accurately estimate internal variability. Here, we can truly demonstrate the role for both internal variability and model differences.

By using large ensembles, ENSO variability can be robustly quantified. The two large ensembles show limited consensus in changes in ENSO strength under warming, with an increase in ENSO strength in CESM-LE in both weak and strong warming scenarios. By contrast, MPI-GE shows limited changes in strength in both scenarios. The pattern changes also show some disparity, with opposing changes in EOF1-E patterns in the

strong warming scenario. There is some agreement in the EOF2-E pattern change under strong warming, where both models simulate a weakening of the EOF2-E pattern suggesting a decrease in ENSO diversity; however, changes in the central Pacific are of opposite sign. The combined pattern and strength changes express themselves by showing an increase in both Niño3 and 4 amplitudes in CESM-LE and limited changes in MPI-GE. In combination with CMIP5 results, this suggests that the forced change in Niño3 and 4 amplitude under warming is likely to be zero or positive, but not negative, although a large variety of results can occur in a single realization.

We find that ENSO varies in a single realization from another due to internal variability and conclude that small ensembles and single realizations cannot be used to investigate ENSO changes. Finally, we find that approximately 30–40 ensemble members of a single model are needed to robustly sample ENSO variance using a 30-year analysis period (<10% error), with larger ensembles or longer analysis periods needed to decrease the error to below 5%. Here, multi-model ensembles are not the best tool for investigating future changes in ENSO, because they confuse internal variability and model differences. Large single-model ensembles must be used instead.

Acknowledgments
We thank the Alexander von Humboldt Foundation for funding the first author's postdoctoral fellowship. We also thank those involved with setting up and running the Grand Ensemble simulations: L. Kornblueh for the historical simulations and the 1% CO2 scenarios, J. Kröger and M. Botzet for producing the RCP4.5 scenario simulations, and Yohei Takano for producing the RCP8.5 scenario simulations. We are indebted to T. Schulthess and the Swiss National Computing Centre (CSCS) for providing the computational resources for the historical simulations and the 1% CO2 scenarios. The RCP4.5 and 8.5 scenario simulations were performed with the facilities at the German Climate Computing Centre (DKRZ). This study was also supported by the BMBF project CLIMPRE InterDec (FKZ:01LP1609A; DM). We are indebted to Urs Beyerle for his effort to download and organize the immense CMIP5 data set. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We would like to acknowledge Dian Putrasahan for her input on spectral calculations and conducting an internal review and Thibault Tabarin for his input on pattern analysis. Finally, we thank Clara Deser and one anonymous reviewer for their comments on this paper. A detailed model evaluation for the ENSO regions can be found in the supporting information (Bellenger et al., 2014; Bittner et al., 2016; Bjerknes, 1969; Guilyardi et al., 2009; Jungclaus et al., 2013; Kay et al., 2015; Sanderson et al., 2015; Stevens, 2015; Zebiak & Cane, 1987). The data used are listed in the references, tables, and information on the publication of the Max Planck Institute Grand Ensemble (MPI-GE) can be found on our website at https://www.mpimet.mpg.de/en/grand-ensemble/.

References

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., & Vialard, J. (2014). ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics*, *42*(7-8), 1999–2018. https://doi.org/10.1007/s00382-013-1783-z

Bittner, M., Schmidt, H., Timmreck, C., & Sienz, F. (2016). Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. *Geophysical Research Letters*, *43*, 9324–9332. https://doi.org/10.1002/2016GL070587

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, *97*, 163–172.

Burgman, R. J., Schopf, P. S., & Kirtman, B. P. (2008). Decadal modulation of ENSO in a hybrid coupled model. *Journal of Climate*, *21*(21), 5482–5500. https://doi.org/10.1175/2008JCLI1933.1

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., et al. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, *4*(2), 111–116. https://doi.org/10.1038/nclimate2100

Cai, W., Santoso, A., Wang, G., Yeh, S.-W., An, S.-I., Cobb, K. M., et al. (2015). ENSO and greenhouse warming. *Nature Climate Change*, *5*(9), 849–859. https://doi.org/10.1038/nclimate2743

Chowdary, J. S., Xie, S.-P., Tokinaga, H., Okumura, Y. M., Kubota, H., Johnson, N., & Zheng, X.-T. (2012). Interdecadal variations in ENSO teleconnection to the Indo–western pacific for 1870–2007. *Journal of Climate*, *25*(5), 1722–1744. https://doi.org/10.1175/JCLI-D-11-00070.1

Christensen, J. H., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I. F. A., de Castro, M., et al. (2013). Climate Phenomena and their Relevance for Future Regional Climate Change. In T. F. Stocker, et al. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY: Cambridge University Press.

Chung, C. T. Y., & Power, S. B. (2016). Modelled impact of global warming on ENSO-driven precipitation changes in the tropical Pacific. *Climate Dynamics*, *47*(3-4), 1303–1323. https://doi.org/10.1007/s00382-015-2902-9

Collins, M., An, S.-I., Cai, W., Ganachaud, A., Guilyardi, E., Jin, F.-F., et al. (2010). The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geoscience*, *3*(6), 391–397. https://doi.org/10.1038/ngeo868

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3-4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Frankcombe, L. M., England, M. H., Mann, M. E., & Steinman, B. A. (2015). Separating internal variability from the externally forced climate response. *Journal of Climate*, *28*(20), 8184–8202. https://doi.org/10.1175/JCLI-D-15-0069.1

Guilyardi, E., Bellenger, H., Collins, M., Ferrett, S., Cai, W., & Wittenberg, A. (2012). A first look at ENSO in CMIP5. *CLIVAR Exchanges*, *17*(58), 29–32.

Guilyardi, E., Braconnot, P., Jin, F.-F., Kim, S. T., Kolasinski, M., Li, T., & Musat, I. (2009). Atmosphere feedbacks during ENSO in a coupled GCM with a modified atmospheric convection scheme. *Journal of Climate*, *22*(21), 5698–5718. https://doi.org/10.1175/2009JCLI2815.1

Ham, Y.-G., & Kug, J.-S. (2016). ENSO amplitude changes due to greenhouse warming in CMIP5: Role of mean tropical precipitation in the twentieth century. *Geophysical Research Letters*, *43*, 422–430. https://doi.org/10.1002/2015GL066864

Hawkins, E., & Sutton, R. (2009). Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *Journal of Climate*, *22*(14), 3960–3978. https://doi.org/10.1175/2009JCLI2720.1

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011GL050087

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advancesd Modelling of the Earth System*, *5*, 422–446. https://doi.org/10.1002/jame.20023

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Keupp, L., Pollinger, F., & Paeth, H. (2017). Assessment of future ENSO changes in a CMIP3/CMIP5 multi-model and multi-index framework. *International Journal of Climatology*, *37*(8), 3439–3451. https://doi.org/10.1002/joc.4928

Kim, S. T., Cai, W., Jin, F.-F., Santoso, A., Wu, L., Guilyardi, E., & An, S.-I. (2014). Response of El Niño sea surface temperature variability to greenhouse warming. *Nature Climate Change*, *4*(9), 786–790. https://doi.org/10.1038/nclimate2326

Li, J., Xie, S.-P., Cook, E. R., Huang, G., D'Arrigo, R., Liu, F., et al. (2011). Interdecadal modulation of El Niño amplitude during the past millennium. *Nature Climate Change*, *1*(2), 114–118. https://doi.org/10.1038/nclimate1086

Maher, N., McGregor, S., England, M. H., & Alexander, S. (2015). Effects of volcanism on tropical variability. *Geophysical Research Letters*, *42*, 6024–6033. https://doi.org/10.1002/2015GL064751

Marotzke, J., & Forster, P. M. (2015). Forcing, feedback and internal variability in global temperature trends. *Nature*, *517*(7536), 565–570. https://doi.org/10.1038/nature14117

Ogata, T., Xie, S.-P., Wittenberg, A., & Sun, D.-Z. (2013). Interdecadal amplitude modulation of El Niño-Southern Oscillation and its impact on tropical Pacific decadal variability. *Journal of Climate*, *26*(18), 7280–7297. https://doi.org/10.1175/JCLI-D-12-00415.1

Power, S., Delage, F., Chung, C., Kociuba, G., & Keay, K. (2013). Robust twenty-first-century projections of El Niño and related precipitation variability. *Nature*, *502*(7472), 541–545. https://doi.org/10.1038/nature12580

Power, S. B., Delage, F. P. D., Chung, C. T. Y., Ye, H., & Murphy, B. F. (2017). Humans have already increased the risk of major disruptions to Pacific rainfall. *Nature Communications*, *8*(14), 368. https://doi.org/10.1038/ncomms14368

Rashid, H. A., Hirst, A. C., & Marsland, S. J. (2016). An atmospheric mechanism for ENSO amplitude changes under an abrupt quadrupling of CO2 concentration in CMIP5 models. *Geophysical Research Letters*, *43*, 1687–1694. https://doi.org/10.1002/2015GL066768

Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Sanderson, B. M., Oleson, K. W., Strand, W. G., Lehner, F., & O'Neill, B. C. (2015). A new ensemble of GCM simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change*, *146*(3), 303–318. https://doi.org/10.1007/s10584-015-1567-z

Smith, T. M., Reynolds, R. W., Peterson, T. C., & Lawrimore, J. (2008). Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *Journal of Climate*, *21*(10), 2283–2296. https://doi.org/10.1175/2007JCLI2100.1

Stevens, B. (2015). Rethinking the lower bound on aerosol radiative forcing. *Journal of Climate*, *28*(12), 4794–4819. https://doi.org/10.1175/JCLI-D-14-00656.1

Stevenson, S. L. (2012). Significant changes to ENSO strength and impacts in the twenty-first century: Results from CMIP5. *Geophysical Research Letters*, *39*, L17703. https://doi.org/10.1029/2012GL052759

Sun, C., Liu, L., Li, L.-J., Wang, B., Zhang, C., Liu, Q., & Li, R.-Z. (2018). Uncertainties in simulated El Niño-Southern Oscillation arising from internal climate variability. *Atmospheric Science Letters*, *19*(3), e805. https://doi.org/10.1002/asl.805

Takahashi, K., Montecinos, A., Goubanova, K., & Dewitte, B. (2011). ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters*, *38*, L10704. https://doi.org/10.1029/2011GL047364

Timmermann, A. (1999). Detecting the nonstationary response of ENSO to greenhouse warming. *Journal of Atmospheric Science*, *56*(14), 2313–2325. https://doi.org/10.1175/1520-0469(1999)056<2313:DTNROE>2.0.CO;2

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*, L12702. https://doi.org/10.1029/2009GL038710

Zebiak, S., & Cane, M. (1987). A model El Niño-Southern Oscillation. *Monthly Weather Review*, *97*(3), 163–172.

Zelle, H., Jan van Oldenborgh, G., Burgers, G., & Dijkstra, H. (2005). El Niño and greenhouse warming: Results from ensemble simulations with the NCAR CCSM. *Journal of Climate*, *18*(22), 4669–4683. https://doi.org/10.1175/JCLI3574.1

Zheng, X.-T., Hui, C., & Yeh, S.-W. (2018). Response of ENSO amplitude to global warming in CESM large ensemble: Uncertainty due to internal variability. *Climate Dynamics*, *50*, 4019. https://doi.org/10.1007/s00382-017-3859-7