

Supporting Information for: A two-year forecast for a 60-80% chance of La Niña in 2017-18

Pedro N. DiNezio¹, Clara Deser², Alicia Karspeck², Stephen Yeager², Yuko Okumura¹, Gokhan Danabasoglu², Nan Rosenbloom², Julie Caron², Gerald A. Meehl²

Submitted July 7, 2017

Revised September 20, 2017

Revised November 1, 2017

Corresponding author: Pedro N. Di Nezio, Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA. (pdn@ig.utexas.edu)

¹Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA.

²Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO, USA.

Contents of this file

1. Text S1
2. Figures S1 to S9
3. Table S1

Text S1**1. ENSO simulation in CESM1**

CESM1 simulates very realistic ENSO events in terms of their spatial patterns, amplitude, teleconnections, and particularly, the asymmetries in amplitude and duration between El Niño and La Niña. Sea-surface temperature anomalies (SSTA) computed from the CESM-CTL exhibit warming of central and eastern equatorial Pacific at the peak of El Niño events (Fig. S2a), albeit extending slightly westward relative to observed events (Fig. S2c). Simulated La Niña events show cooling of the central and eastern equatorial Pacific (Fig. S2b) also displaced slightly westward relative to observations (Fig. S2d). CESM1 simulates ENSO indices with amplitude consistent with the observed values. Niño-3.4 SST variability is slightly overestimated, as seen in the larger standard deviation ($\sim 10\%$ larger) of the Niño-3.4 SST index relative to observations (Table S1). The strong SST variability does not translate into increased subsurface variability, since the simulated \bar{Z}'_{TC} indices show about the same standard deviation as the reanalysis-derived variability (Table S1).

2. Decadal Prediction Large Ensemble

2.1. Drift correction

Each ensemble member evolves following a unique climate trajectory dictated by the influence of the initial conditions, variations in the external forcings, and each member's weather and internal climate variability. Over time, each ensemble member slowly drifts towards the climatology of the model, which will exhibit biases relative to the observed initial climate. We corrected this slowly evolving drift following the technique recommended by CLIVAR [*International CLIVAR Project Office*, 2011]. This is done computing mean climate fields from all ensembles at each lead time (months 1 to 122). The resulting time-evolving climatology captures the model's drifting climate. The 51 ensembles initialized over the 1964–2014 period were used because they allow the computation of a climatology for which the same-size ensemble ($51 \times 40 = 2040$) exists for each lead time. This time-evolving climatology is removed from each of the 2480 (62×40) members to obtain drift-corrected climate anomalies.

This approach removes the drift that is common to all ensembles, isolating the internally-generated and externally-forced climate anomalies. This methodology has been extensively used in seasonal and decadal climate forecasts [*Yeager et al.*, 2012; *Meehl and Teng*, 2012; *Hazeleger et al.*, 2013]. SST and rainfall anomalies are computed following this approach. For thermocline depth, we first compute the depth of the maximum the temperature gradient from the uncorrected subsurface temperature fields, and then apply the drift-removal technique to the full Z_{tc} fields to compute anomalies. This drift removal technique will be most effective when the model errors are independent of the forecast state (i.e., the system is linear in its error statistics), and if the error distribution is nor-

mal [Stockdale, 1997]. These are reasonable assumptions for ENSO, although interactions between forecast state and error are to be expected, diminishing forecast skill. In section 3 we show that the CESM-DP-LE is capable of predicting 2-year La Niña when initialized at the peak of strong El Niño events. This gives us confidence that the drift-removal technique does not cause a large reduction in forecast skill relative the skill shown in “perfect model” experiments [DiNezio *et al.*, 2017].

2.2. Forecast plumes

We computed the mean Niño-3.4g and \bar{Z}'_{TC} indices among all members of each ensemble to quantify the predictable component of each forecast. We define a 2nd year La Niña when the Niño-3.4g SST index is less than -0.60 K, during the November-December-January (NDJ) season, 2 years after the forecast start date (NDJ+2). Probability of 2-year La Niña are computed as the fraction of members showing Niño-3.4g SST index under the -0.60 K threshold during NDJ+2.

A composite of events starting from strong and moderate El Niño shows that on average the CESM-DP-LE hindcasts capture the evolution of the subsequent La Niña events in terms of timing and amplitude (Figs. S6a–b). These composites exclude members that show returning El Niño conditions during the NDJ season 1 year after the forecast start date (NDJ+1) (e.g. Figs. S6c–k, red curves). We identified the members that simulate a realistic transition from El Niño to La Niña during the first year by requiring that the Niño-3.4g SST index be less than 0.6 K (i.e. ENSO neutral or La Niña) during the May-June-July (MJJ) season during the first year of the hindcasts (MMJ+1). By the requiring neutral conditions during summer, this criterion imposes the condition that the

initial El Niño transitions to La Niña, however, it does not impose the 2-year persistence of La Niña. The fact that the composites for strong and moderate events agree with observations (Fig. S6a and Fig. S6b) indicates that this correction is effective at removing forecast bias. Note, that the CESM-DP-LE predicts a consistent transition to La Niña in three out of four of the strongest El Niño events on record. Consistent for 1972 (Fig. S6e), 1997 (Fig. S6i) and 2015 (now shown), with inconsistent transition for 1982 (Fig. S6f). The causes for the inconsistent transition in the 1982-initialized hindcast will be explored in a follow-up study. This approach is analogous to other techniques for bias correction, such as amplitude corrections, which are commonly applied to correct biases in climate forecasts. We emphasize that this correction is applied systematically to all ensembles, without seeking to increase the skill of individual ensembles. This correction is less needed for most forecasts initialized from strong El Niño events. For instance, we removed only one member from the 2015-initialized forecasts used to predict the current La Niña.

3. Retrospective verification of 2-year La Niña forecasts

Retrospective verification is required to demonstrate the skill of a forecast system prior to its use for actual predictions [*Kirtman et al.*, 2001; *Tang et al.*, 2005; *Barnston et al.*, 2009; *Wang et al.*, 2010; *Tippett et al.*, 2011; *Ham et al.*, 2014]. This section presents an assessment of the quality of the ENSO forecasts produced by the CESM-DP-LE. We assessed quality in terms of bias and RMS error of the Niño-3.4g SST index relative to observations focusing on the first 27 months (9 seasons) of the hindcasts.

3.1. Verifying data

We evaluate the CESM-DP-LE against the observed evolution of ENSO events based on the Niño-3.4g SST index. ERSST3b [Smith *et al.*, 2008] data are used to compute the observed Niño-3.4g SST index. Observed SSTA are computed relative to the 1964–2014 monthly-mean climatology, as in the CESM1-DP-LE. Neither the observed nor the CESM-DP-LE Niño-3.4g SST indices are detrended since we expect that forced signals will have a minimal effect on our gradient-based definition.

3.2. Retrospective verification

We assess the quality of the CESM1-DP-LE predictions for different initial ENSO states. We cluster the 61 ensembles according to whether they were initialized from an El Niño or a La Niña year. We use the definition of El Niño and La Niña years given by NOAA’s CPC (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). Each cluster contains 20 and 19 ensembles initialized from El Niño and La Niña respectively. We evaluate the quality of the CESM1-DP-LE forecasts in both cases to assess the skill at 2 and 1 year lead times. The 2015-initialized forecast is excluded from the verification.

3.3. Hindcast bias

We explore biases in the CESM-DP-LE retrospective predictions by comparing composites of the hindcast and observed Niño-3.4g SST indices for each cluster. The temporal evolution of the hindcast and observed composites suggest substantial differences, particularly for hindcasts initialized from El Niño conditions (Fig. S7) These differences reveal pronounced biases during the first year of the El Niño-initialized hindcasts. On average,

the CESM-DP-LE predicts positive Niño-3.4g values peaking one year after initialization and negative values for the following year (Fig. S7a, gray line). In contrast, the observed composite shows negative Niño-3.4g SST values from the first to the second year after initialization consistent with the tendency of La Niña events to last 2 years (Fig. S7a, black line). This difference arises because some CESM-DP-LE hindcasts show an excessive tendency for predicting returning El Niño conditions for the first year, as discussed in section 2.2.

The composite of El Niño-initialized members with neutral conditions during the first MJJ season (MJJ+1) shows Niño-3.4 SST index with striking similarity to the observed composite. This confirms that hindcasts that do not transition into 2-year El Niño have very low bias (Fig. S7b). Therefore we can exclude the unrealistic hindcasts in our evaluation of predictive skill. In real-time predictions, we would not know if year 1 turns out to be El Niño or La Niña. However, the low bias found after this procedure suggests that excluding these members for moderate and strong El Niño will lead to higher quality forecasts. Last, the La Niña-initialized ensembles show that, on average, the CESM-DP-LE predicts negative Niño-3.4g SST index for the following boreal winter (Fig. S7c, grey line). This suggest that CESM-DP-LE could be skillful at predicting the return of La Niña one year after the first peak.

3.4. RMS error

We use an autoregressive model of order 1 (AR1) derived from the observed Niño-3.4g SST index to generate “forced persistence” hindcasts initialized in each November from 1954 to 2014. For each initial November we generate an ensemble of 1000 AR1 hindcasts.

We compare the RMS error of the Niño-3.4g SST index predicted by these AR1 ensembles vs. the CESM-DP-LE. The RMS error is computed for each realization of the AR1 and CESM-DP-LE member over the first 9 seasons (27 months). This procedure yields distributions of RMS errors for the AR1 model and the CESM-DP-LE. Before comparing these distributions, we separated the hindcasts into the same clusters discussed in the previous section. The distribution of RMS error from the CESM-DP-LE is broader than the AR1 model (Fig. S8a). CESM-DP-LE appears to be more skillful than the AR1 model under certain conditions, because of its ability to predict 2-year La Niña; but also less skillful, because of the excessive 2-year El Niño. The distribution of the RMS error improves dramatically when we apply the neutral MJJ+1 condition (as in the composite analysis) to the hindcasts initialized from strong and moderate El Niño (Fig. S8b). The CESM-DP-LE shows less RMS error than the AR1 model. This supports the notion that a realistic transition to La Niña is critical for more skillful CESM-DP-LE forecasts. Lastly, forecasts initialized from La Niña conditions are also more skillful than the AR1 model (Fig. S8c) supporting our previous point.

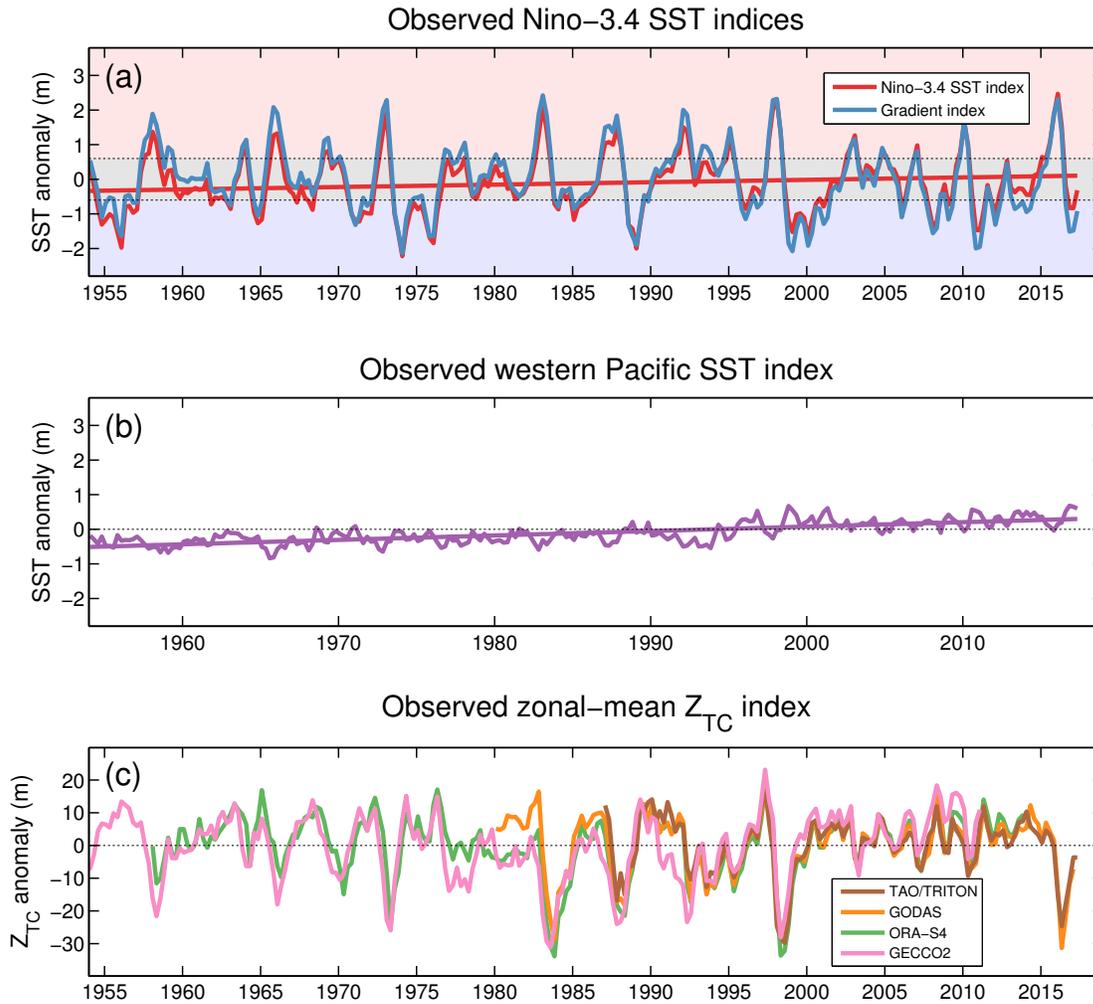


Figure S1. Observed ENSO indices. (a) Niño-3.4 sea-surface temperature (SST) indices and (b) zonally averaged thermocline depth index (\bar{Z}'_{TC}). The linear trend line corresponds to the Niño-3.4 SST index. SST indices are computed using ERSST3b [Smith *et al.*, 2008] data. The \bar{Z}'_{TC} index is computed using potential temperature data from different ocean reanalyses (ORAS-4 [Balmaseda *et al.*, 2013], GECCO2 [Khl and Stammer, 2008], and GODAS [Huang *et al.*, 2010]) and from in-situ observations from the TAO/TRITON moored array.

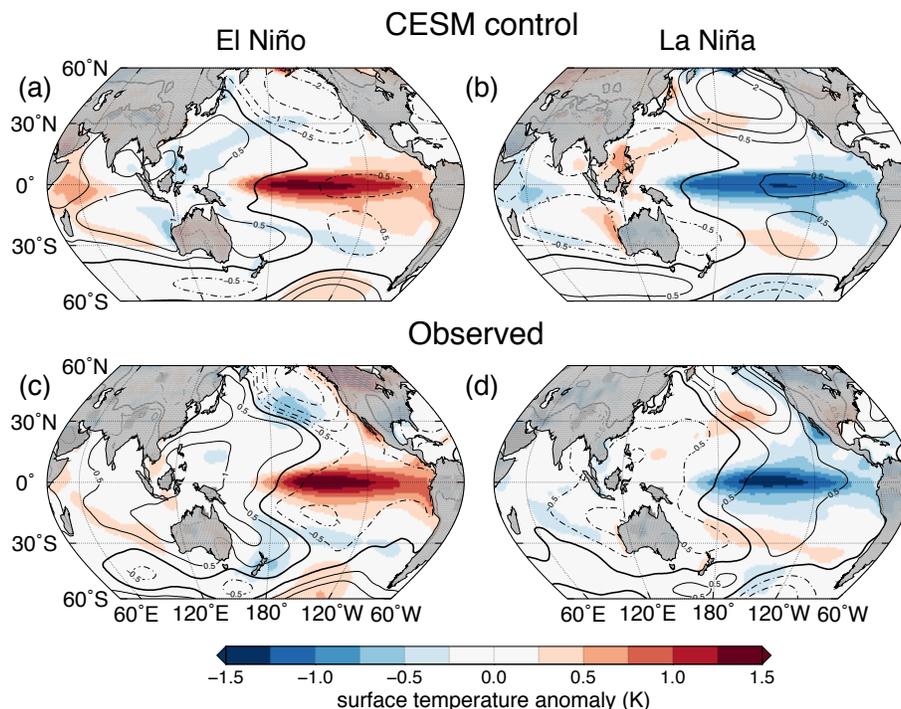


Figure S2. Observed and simulated climate anomalies associated with El Niño and La Niña. Composite anomalies of sea-surface temperature (SST, shading) and sea level pressure (SLP, contours) during December-January-February (DJF) for simulated (top) and observed (bottom) El Niño (left) and La Niña (right) events. The composite El Niño and La Niña events are based on observations (years 1901–2012) and the CESM-CTL (years 401–2200). Observed SST and SLP data are from ERSST3b [Smith *et al.*, 2008] and the 20th Century Reanalysis [Compo *et al.*, 2011]. Solid (dashed) contours show positive (negative) SLP anomalies at 0.5 hPa intervals. The simulated composite anomalies are scaled so that the amplitude of the Niño-3.4 SST index matches observations.

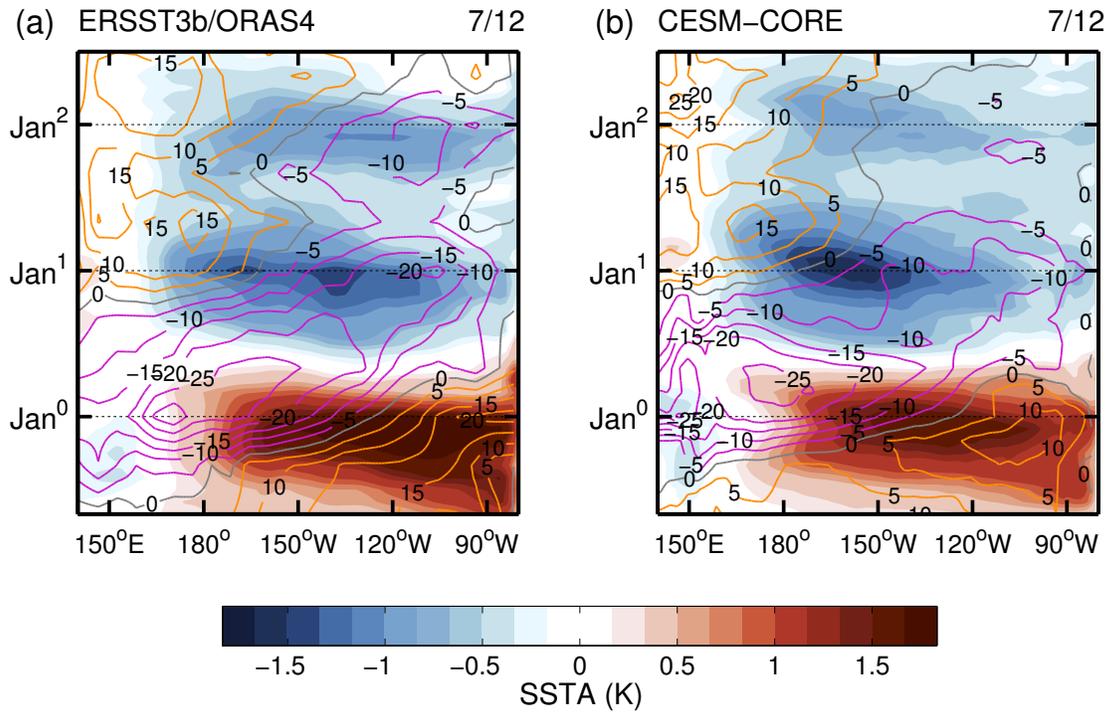


Figure S3. 2-year La Niña in observational data and the CESM-CORE simulation.

Composite sea-surface temperature anomalies (SSTA, shading) and thermocline depth anomalies (Z'_{tc} , contours) for 2-year La Niña events in (a) observational data and the (b) CESM-CORE simulation used to initialize predictions. The SSTA and Z'_{tc} are averaged over the 5°S - 5°N band. Orange (purple) contours show positive (negative) Z'_{tc} on 5 m intervals. Observational anomalies (right) are computed from detrended ERSST3b [Smith *et al.*, 2008] and ORAS4 [Balmaseda *et al.*, 2013] data. The CESM-CORE simulation is described in section ???. Both composite La Niña events were formed using anomalies from seven 2-year events observed over the 1958-2014 period when the two datasets overlap.

Predicted and observed 2-year La Niña (first peak)

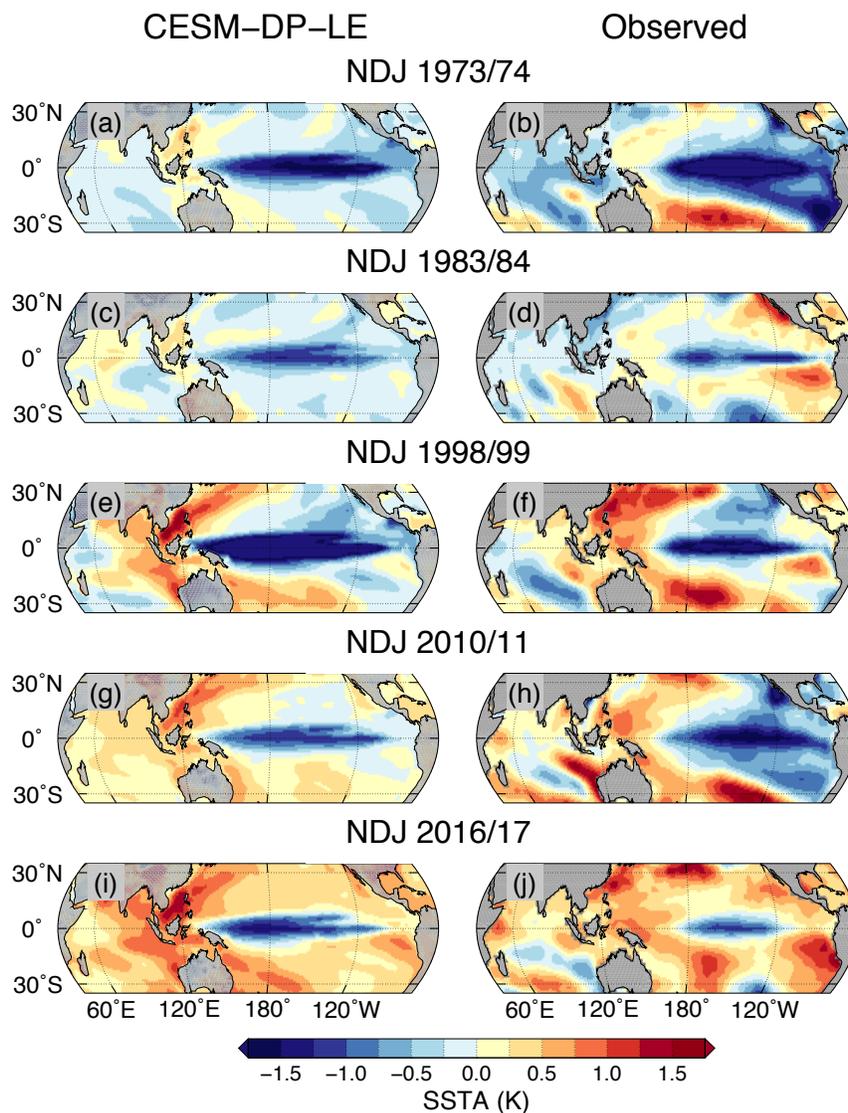


Figure S4. Predicted vs. observed sea-surface temperature anomalies for the first peak of 2-year La Niña. Ensemble-mean (left) and observed (right) sea-surface temperature anomaly (SSTA) during the first November-December-January (NDJ) peak of selected 2-year La Niña events. The ensemble-mean SSTA are from CESM-DP-LE forecasts initialized at the peak of the El Niño events of 1972, 1982, 1997, 2009, and 2015. Members that evolve into 2-year El Niño are not included in the ensemble mean calculation as described in section 2.2. Observed SSTA are from ERSST3b [Smith *et al.*, 2008].

Predicted and observed 2-year La Nina (second peak)

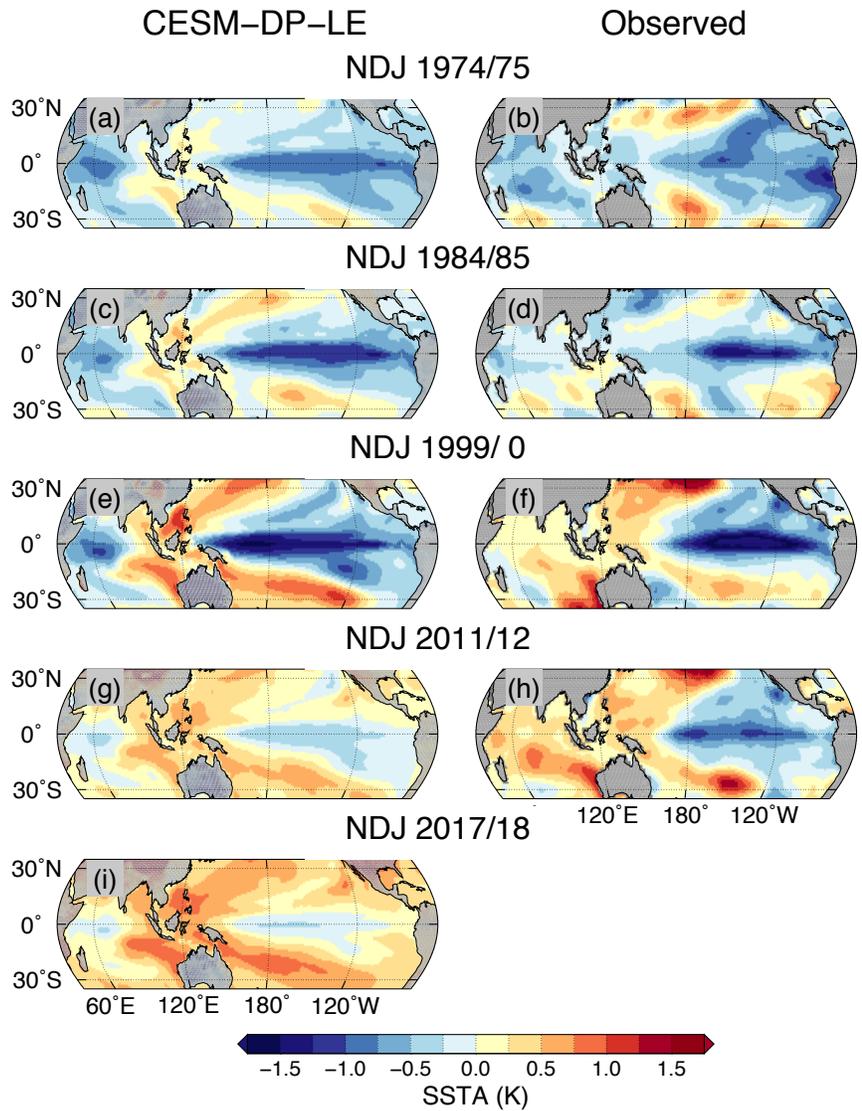


Figure S5. Predicted vs. observed sea-surface temperature anomalies for the second peak of 2-year La Niña. As Fig S4, but for the second NDJ peak of 2-year La Niña.

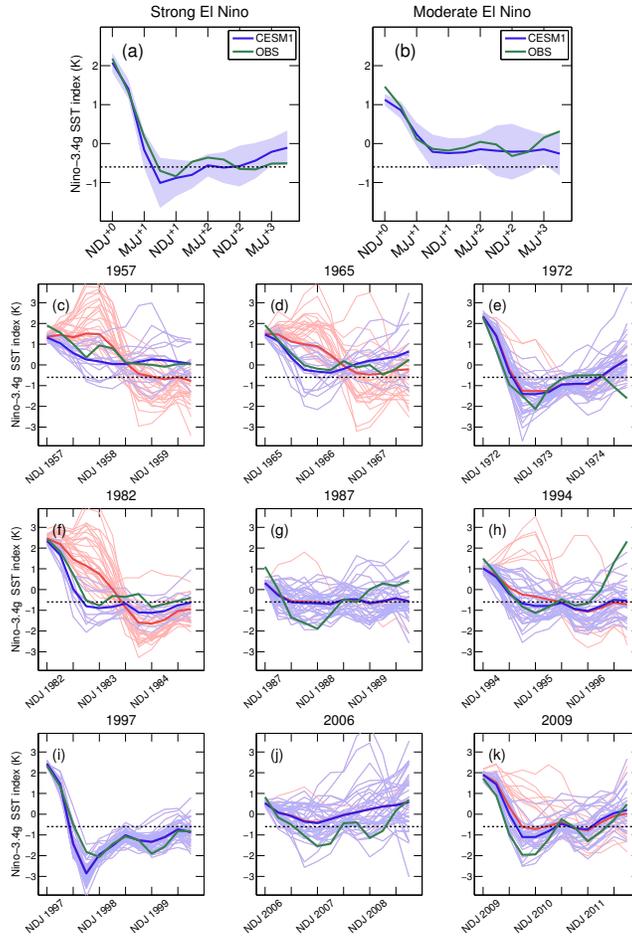


Figure S6. Niño-3.4g SST index from CESM-DP-LE hindcasts initialized at the peak of strong and moderate El Niño events. Observed and ensemble-mean predicted Niño-3.4g SST index for a composite of historical La Niña events preceded by (a) strong and (b) moderate El Niño conditions. Shading indicates the standard deviation among the ensembles in each composite. (c-k) Niño-3.4g SST plumes for the individual ensembles. Green curves correspond to the observed Niño-3.4g SST index. Blue curves identify the members (thin curves) and ensemble-mean (solid curve) that show realistic initial transition from El Niño to La Niña. Thin red curves identify the members that are excluded because of their unrealistic evolution into 2-year El Niño. Solid red curve shows the ensemble-mean of all members regardless of their initial transition into La Niña.

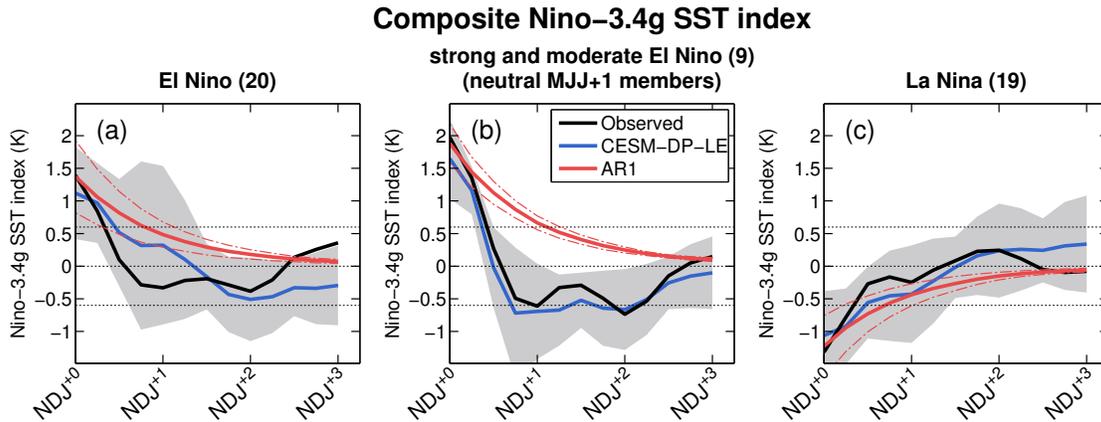


Figure S7. Composite evolution of Niño-3.4 SST index from El Niño- and La Niña-initialized hindcasts. Composite ensemble-mean Niño-3.4g SST index in hindcasts initialized from initial conditions characterized by (a) El Niño, (b) strong and moderate El Niño followed by neutral May-June-July (MJJ), and (c) La Niña. Composites include observed events shown in black, CESM-DP-LE hindcasts shown in blue, and forced persistence hindcasts generated with an autoregressive process of order 1 (AR1), shown in red. Gray shading indicates the spread among the ensemble-mean indices. Dash-dotted red curves indicate the spread among the ensemble-mean AR1 hindcasts. Dashed horizontal lines indicate the 0.6 K and -0.6 K thresholds used to define El Niño or La Niña conditions.

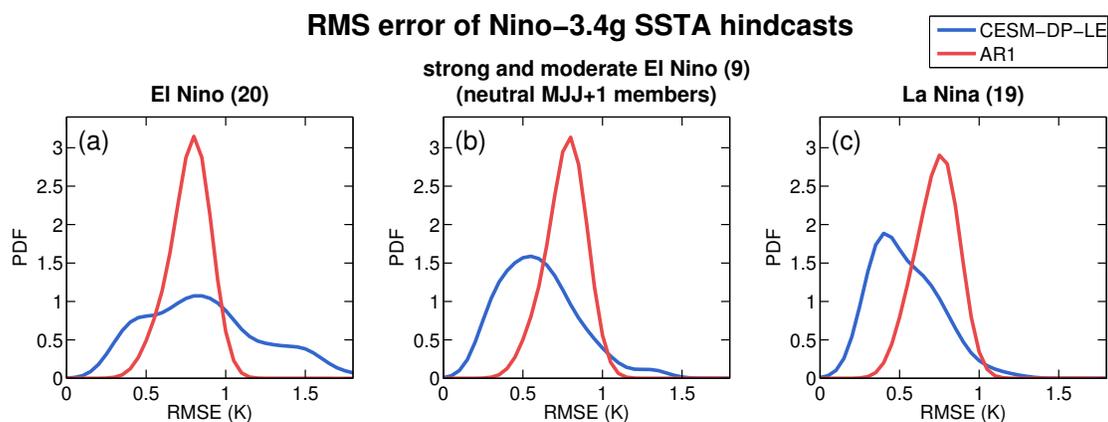


Figure S8. RMS error of Niño-3.4g SST index from El Niño- and La Niña-initialized hindcasts. Probability density function (PDF) of the root-mean-square (RMS) error of the Niño-3.4g index in CESM-DP-LE hindcasts (blue) initialized from conditions characterized by (a) El Niño, (b) strong and moderate El Niño followed by neutral May-June-July (MJJ), and (c) La Niña. RMS error PDF from hindcasts performed with an autoregressive process of order 1 (AR1) are also shown (red). The RMS error of the Niño-3.4g SST index is computed relative to the observed values over the first nine seasons (27 months) of the CESM-DP-LE and AR1 hindcasts.

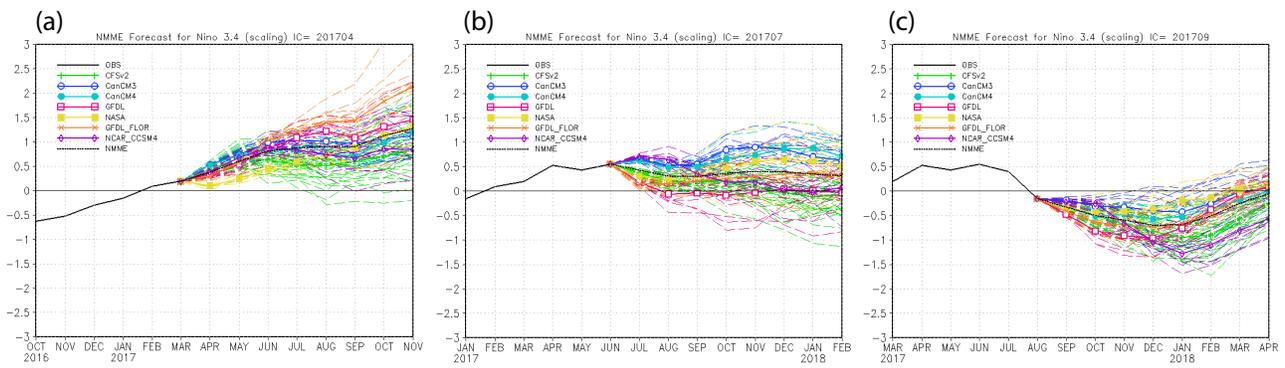


Figure S9. NMME ENSO plumes. ENSO predictions produced by models participating in the NMME initialized on (a) April, (b) July, and (c) September of 2017. Obtained from <http://www.cpc.ncep.noaa.gov/products/NMME/>.

Table S1. Standard deviation of ENSO indices computed from different observational datasets and from the CESM control (CESM-CTL).

| Dataset | Standard deviation |
|--|---------------------------|
| <hr/> Zonal-mean thermocline depth index <hr/> | |
| CESM-CTL | 9.7 m |
| GODAS | 9.8 m |
| TAO/TRITON | 9.4 m |
| ORA-S4 | 10.5 m |
| <hr/> Niño-3.4 SST index <hr/> | |
| CESM1-CTL | 0.91 K |
| ERSST3b | 0.80 K |
| ERSST4 | 0.80 K |
| HadISST1.1 | 0.76 K |
| <hr/> Niño-3.4g SST index <hr/> | |
| CESM-CTL | 1.12 K |
| ERSST3b | 0.93 K |
| ERSST4 | 1.01 K |
| HadISST1.1 | 0.89 K |