Supporting Information for "Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change"

Robert C. J. Wills[1], **Yue Dong**[2], **Cristian Proistosecu**[3], **Kyle C. Armour**[1,4],
David S. Battisti[1]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA

[2]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY

[3]Department of Atmospheric Sciences, University of Illinois, Urbana-Champaign, IL

[4]School of Oceanography, University of Washington, Seattle, WA

Contents of this file

1. The probability of getting an anomalous pattern as large as observed within the LEs

2. Figures S1 to S10

1 The probability of getting an anomalous pattern as large as observed within the LEs

To quantify how likely it is to find a pattern of differences within the multi-model large ensemble as large as that found between observations and the multi-model mean (Fig. 3), we repeat the signal-to-noise maximizing analysis described in the main text for three types of resampled difference ensembles constructed as follows:

- a random model simulation is taken as observations and compared to 10 members each from the other 15 LEs (inter-model differences),

- a random model simulation is taken as observations and compared to the other members of the same LE (intra-model differences),

- each member of the difference ensemble is composed of the difference between two random simulations (random sampling), meaning that the ensemble mean of this difference ensemble would be zero given sufficient ensemble size.

Corresponding author: Robert C. Jnglin Wills, `rcwills@uw.edu`

We generate 80 of each of these types of resampled difference ensembles, perform the same signal-to-noise maximizing pattern analysis on each resampled difference ensemble, and compare the resulting signal fractions, signal-to-noise-ratios, and trend magnitudes with those found when using the actual observations (Fig. S10). The signal-to-noise ratio (SNR) is defined as in the main text as the ratio of signal variance to noise variance, where signal is defined as the difference between observations and the multi-model mean on 5-year and longer timescales, and noise is defined as all other variance in the difference ensemble. The signal fraction s is related to the SNR by SNR $= s/(1-s)$.

The results of this analysis (Fig. S10) show that the differences in low-frequency variability and change between observations and models encapsulated in pattern 1 (Fig. 3) are comparable in magnitude to patterns that could arise from a combination of inter-model differences and internal variability (i.e., comparing to the inter-model difference ensembles), but are larger than are likely to occur due to internal variability alone (i.e., signal fractions as high as found in observations occur in only 12.5% of the intra-model difference ensembles). This analysis shows that the magnitude of observations-model differences is slightly larger than the average inter-model differences, larger than typically occurs due to sampling of internal variability within a single LE, and much larger than could occur due to random sampling. Furthermore, the magnitude of trends found in patterns 1 and 2 are are at the upper end of those found in signal-to-noise maximizing pattern analysis of the resampled difference ensembles, showing that the model bias in multi-decadal trends compared to observations is about as large as can can be found by sub-sampling inter-model differences and internal variability in the multi-model large ensemble (Fig. S10c, d). As in Fig. 3, trends in Fig. S10c and d are shown for an index that is normalized to have unit standard deviation in the ensemble mean of the difference ensemble.
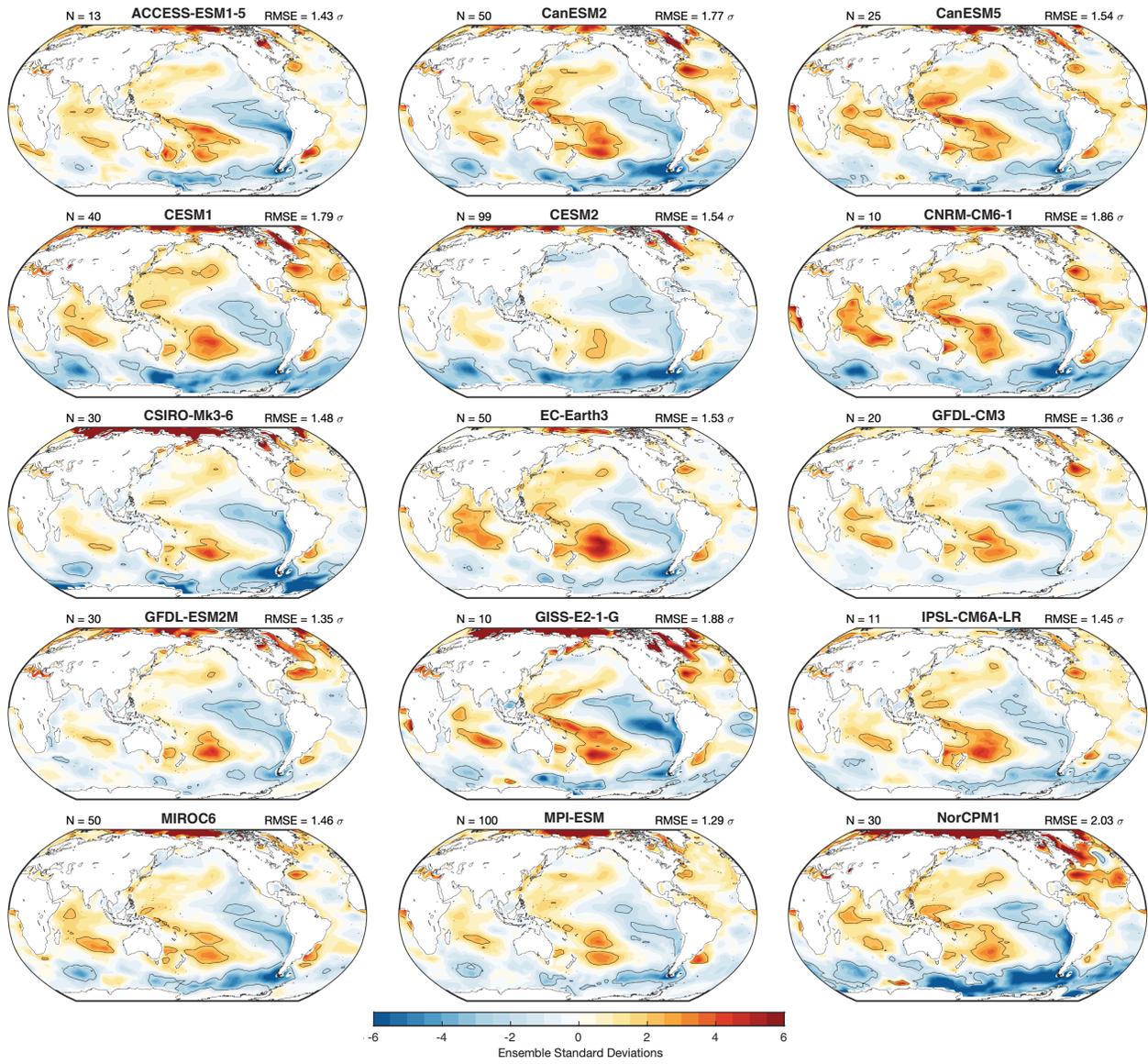
Figure S1. Same as Fig. 1e, but computed separately for each model. Unlike in Fig. 1, all ensemble members are used, and the number of ensemble members included is displayed in the upper left of each panel.
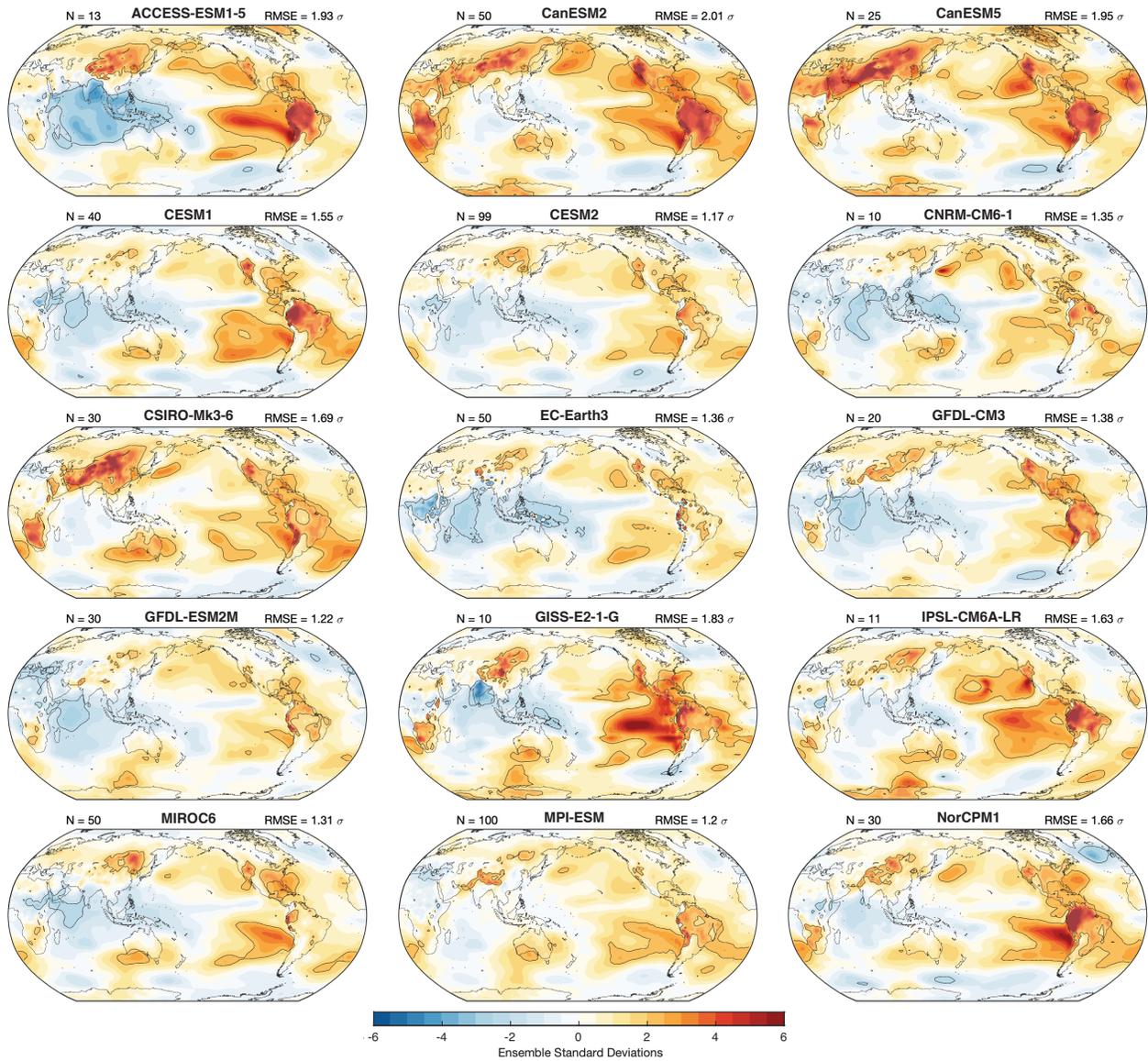
Figure S2. Same as Fig. 1f, but computed separately for each model. Unlike in Fig. 1, all ensemble members are used, and the number of ensemble members included is displayed in the upper left of each panel.
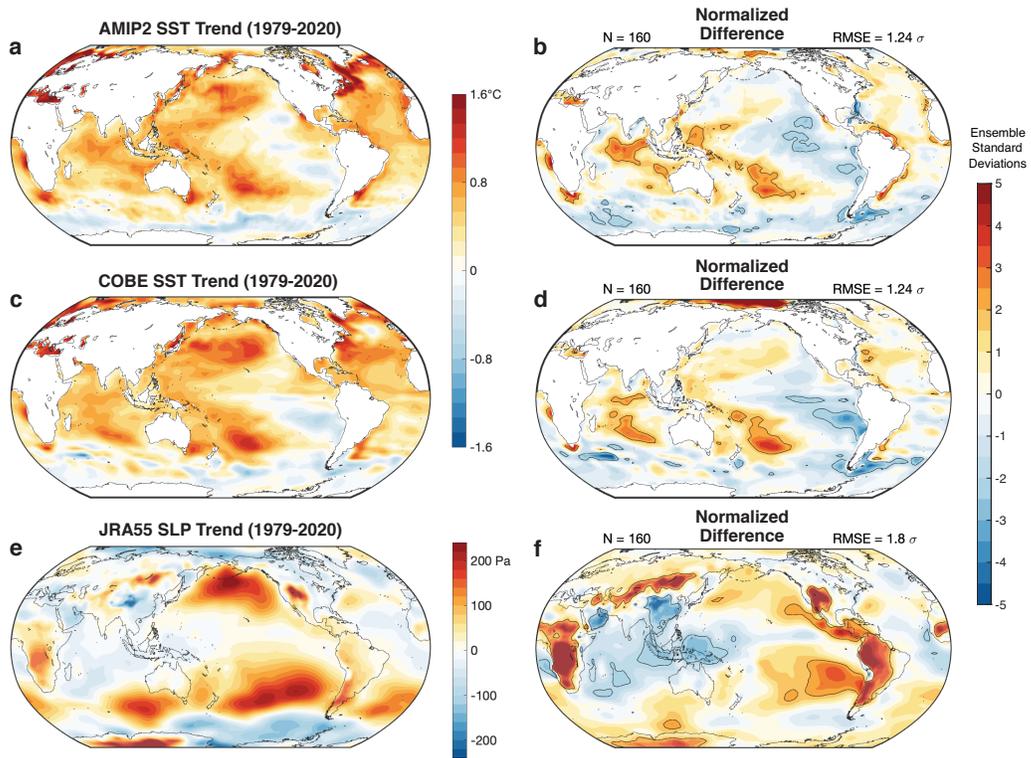
Figure S3. (a, b) Same as Fig. 1a and e, except for AMIPII instead of ERSSTv5. (c, d) Same as Fig. 1a and e, except for COBE instead of ERSSTv5. (e, f) Same as Fig. 1b and f, except for JRA55 instead of ERA5.

ERSST5 DJF SST Trend (1979-2020)

a

1.6°C
0.8
0
-0.8
-1.6

ERA5 DJF SLP Trend (1979-2020)

b

600 Pa
400
200
0
-200
-400
-600

Multi-Model-Mean DJF SST Trend (1979-2020)

c

1.6°C
0.8
0
-0.8
-1.6

Multi-Model-Mean DJF SLP Trend (1979-2020)

d

400 Pa
200
0
-200
-400

Normalized Difference

N = 160

RMSE = 1.11 σ

e

4
2
0
-2
-4

Normalized Difference

N = 160

RMSE = 1.1 σ

f

4
2
0
-2
-4
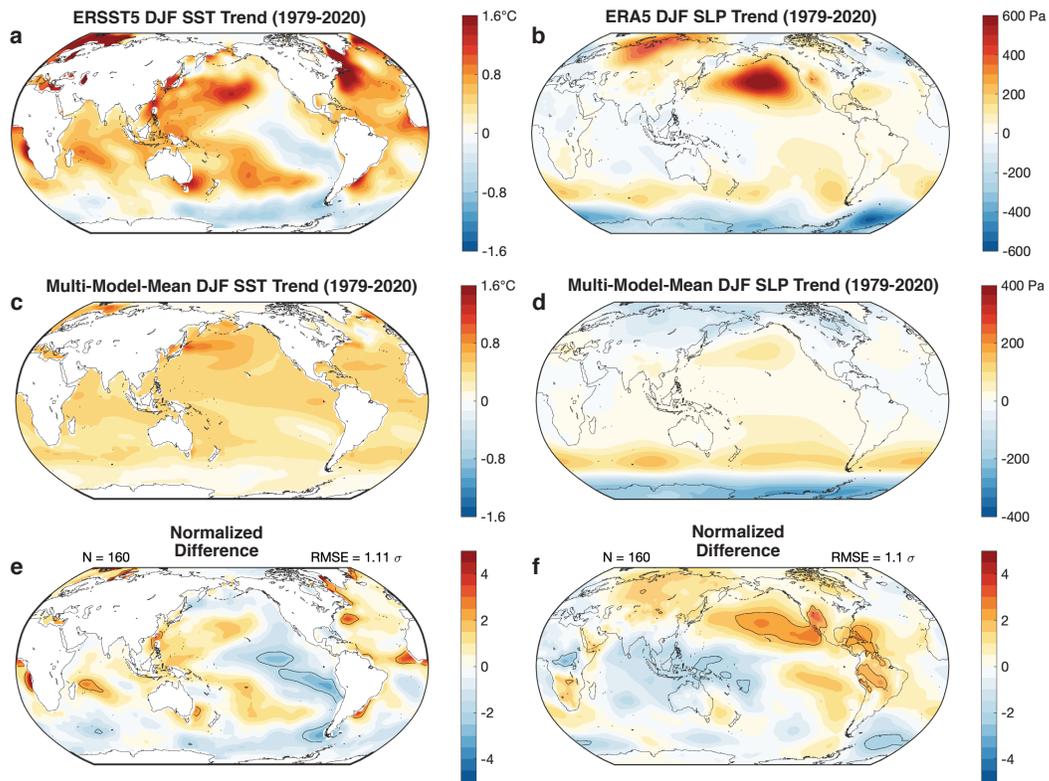
Figure S4. Same as Fig. 1, but for December-January-February (DJF) instead of annual mean.

ERSST5 JJA SST Trend (1979-2020)

a

ERA5 JJA SLP Trend (1979-2020)

b

Multi-Model-Mean JJA SST Trend (1979-2020)

c

Multi-Model-Mean JJA SLP Trend (1979-2020)

d

Normalized Difference

e N = 160   RMSE = 1.35 σ

Normalized Difference

f N = 160   RMSE = 1.12 σ
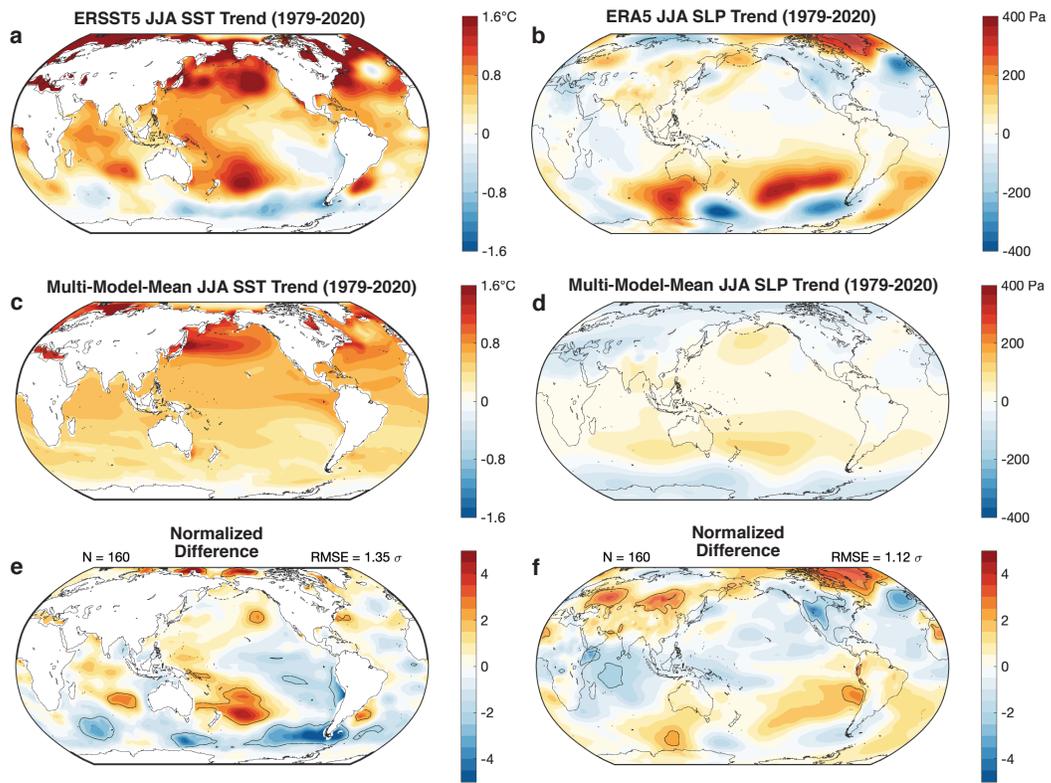
Figure S5. Same as Fig. 1, but for June-July-August (JJA) instead of annual mean.
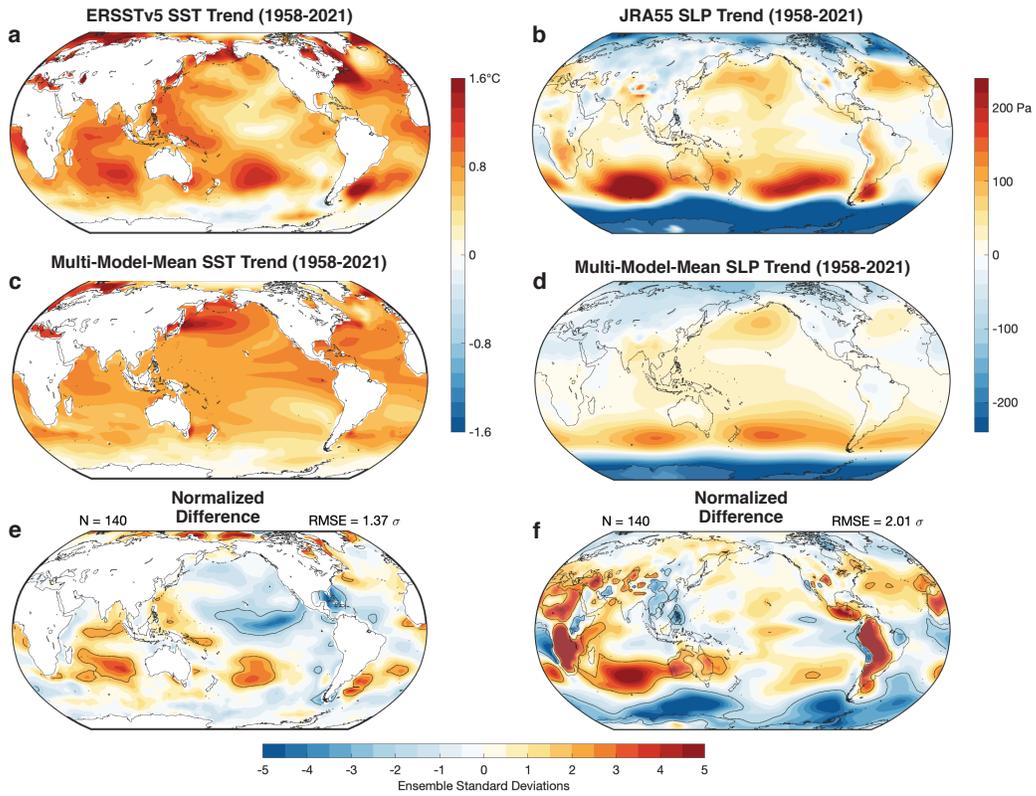
Figure S6. Same as Fig. 1, but for longer-term trends (1958-2021) and using JRA-55 for SLP instead of ERA5.
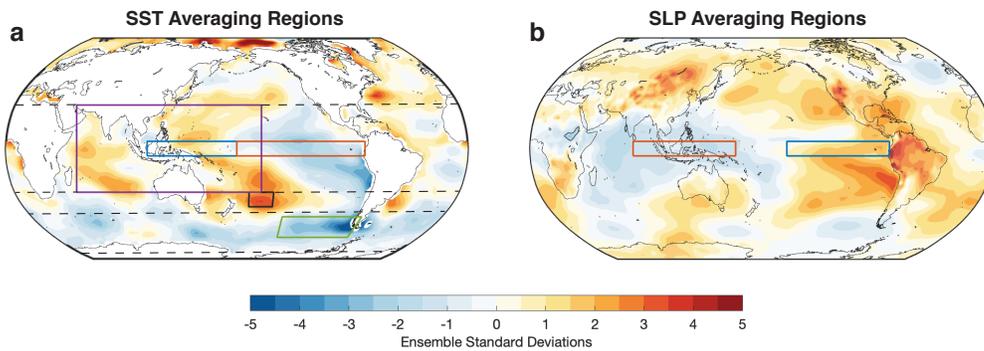
Figure S7. Same as Fig. 1e and f, but with the addition of the averaging regions used in Figs. 2 and S8. Dashed lines show the latitudes of 30°N, 30°S, 45°S, and 75°S.
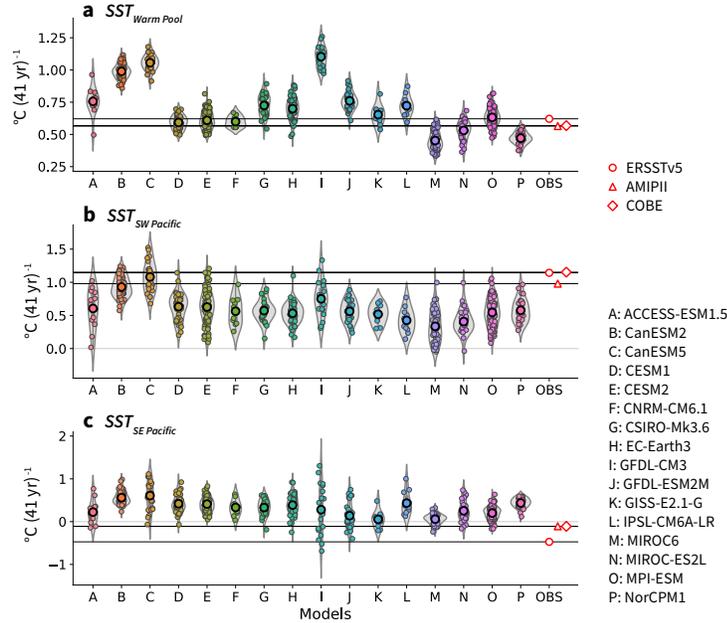
Figure S8. Comparison of observed trends (1979–2020) in key SST indices with those in all ensemble members from 16 LEs: SST in the (a) Warm Pool (30°S-30°N, 50°E-160°W); (b) southwest Pacific (30°S-40°S, 170°W-150°W); and (c) southeast Pacific (47°S-62°S, 140°W-70°W). The southwest and southeast Pacific are regions of highly anomalous observed trends (Fig. 1) . Violin plots (Waskom 2021) for each model can be compared with multiple observational products, shown on the right-hand side. Ensemble average trends for each index and model are shown with black circles.
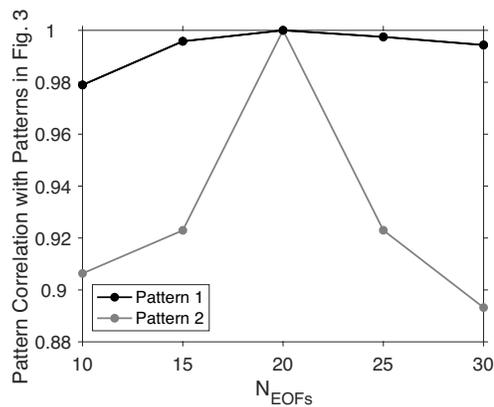
Figure S9. Robustness of the analysis shown in Fig. 3 to the number of EOFs included. For each choice of the number of EOFs (N_{EOFs}), the pattern correlation with the patterns in Fig. 3 are shown for the pattern with the maximum pattern correlation. Only the absolute value of the pattern correlation is considered.
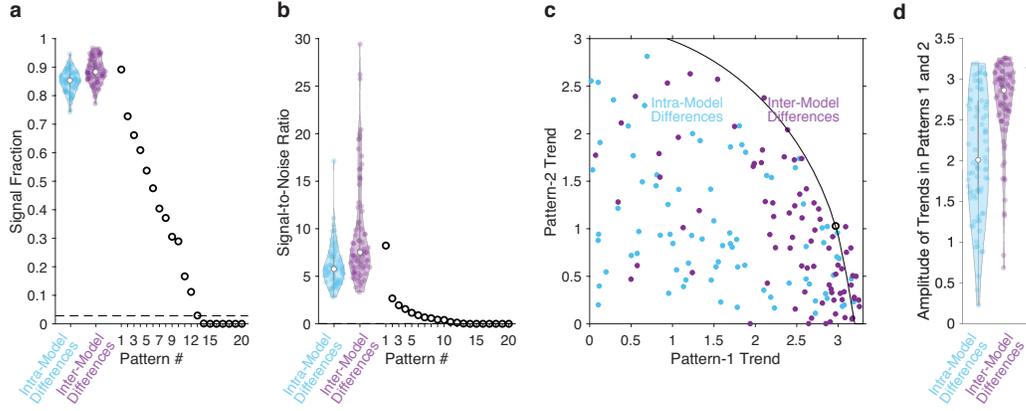
Figure S10. Eigenvalue spectrum of the signal-to-noise maximizing pattern analysis shown in Fig. 3, shown both in terms of (a) signal fraction s and (b) signal-to-noise ratio SNR $= s/(1 - s)$ (black circles). Values that have a 5% chance of occurring due to random sampling of differences within the multi-model ensemble are shown with a black dashed line. The range of pattern-1 values that could occur due to inter-model differences are shown with a purple violin plot. The range of pattern-1 values that could occur due to internal variability within individual LEs are shown with a cyan violin plot. (c) The range of trend magnitudes (per 41 yr) in signal-to-noise maximizing patterns 1 and 2 in difference ensembles sampling inter-model and intra-model differences, compared to the trends in patterns 1 and 2 for the analysis of observations compared to models shown in Fig. 3 (black circle). (d) The amplitude of trends in patterns 1 and 2 (i.e., the radial distance in panel (c)) in observations (black circle) and the range of amplitudes that could occur due to internal variability and inter-model differences (cyan and purple violin plots, respectively). Resampling procedures are described in the text of the Supporting Information.