Systematic climate model biases in the large-scale pattern of recent sea-surface temperature and sea-level pressure change

Jnglin Wills Robert C[1], Dong Yue[2], Proistosecu Cristian[3], Armour Kyle C[1], and Battisti David S[1]

[1]University of Washington
[2]Columbia University
[3]University of Illinois

November 16, 2022

Abstract

Observed surface temperature trends over recent decades are characterized by (i) intensified warming in the Indo-Pacific Warm Pool and slight cooling in the eastern equatorial Pacific, consistent with strengthening of the Walker circulation, and (ii) cooling in the Southern Ocean. In contrast, state-of-the-art coupled climate models generally project Walker circulation weakening, enhanced warming in the eastern equatorial Pacific, and warming in the Southern Ocean. Here we investigate the ability of 16 climate model large ensembles to reproduce observed sea-surface temperature and sea-level pressure trends over 1979-2020 through a combination of externally forced climate change and internal variability. We find large-scale differences between observed and modeled trends that are very unlikely (<5% probability) to occur due to internal variability as represented in models. Disparate trends are found even in regions with weak multi-decadal variability, suggesting that model biases in the transient response to anthropogenic forcing constitute part of the discrepancy.

Systematic climate model biases in the large-scale pattern 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

Key Points:

- The pattern of observed sea-surface temperature and sea-level pressure trends (1979–2020) differs significantly from climate model hindcasts
- The ratio of Indo-Pacific Warm Pool to tropical-mean warming is particularly anomalous in observations compared to models
- A signal-to-noise maximizing pattern analysis is used to isolate changes that occurred in observations that models do not reproduce

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

Abstract

Observed surface temperature trends over recent decades are characterized by (i) intensified warming in the Indo-Pacific Warm Pool and slight cooling in the eastern equatorial Pacific, consistent with strengthening of the Walker circulation, and (ii) cooling in the Southern Ocean. In contrast, state-of-the-art coupled climate models generally project Walker circulation weakening, enhanced warming in the eastern equatorial Pacific, and warming in the Southern Ocean. Here we investigate the ability of 16 climate model large ensembles to reproduce observed sea-surface temperature and sea-level pressure trends over 1979–2020 through a combination of externally forced climate change and internal variability. We find large-scale differences between observed and modeled trends that are very unlikely (<5% probability) to occur due to internal variability as represented in models. Disparate trends are found even in regions with weak multi-decadal variability, suggesting that model biases in the transient response to anthropogenic forcing constitute part of the discrepancy.

Plain Language Summary

Regional climate change depends not only on the magnitude of global warming, but also on the spatial pattern of warming. We show that the spatial pattern of observed temperature changes since 1979 is highly unusual, and many aspects of it cannot be reproduced in current climate models, even when accounting for the influence of natural variability. We find a particularly large discrepancy in the rate of warming within the western Pacific Ocean and eastern Indian Ocean, which suggests that models have systematic biases in the transient response of ocean temperature patterns to anthropogenic forcing, because the contribution of natural variability to multi-decadal trends is thought to be small in this region. Our work raises the possibility that the recent trends towards more La-Niña-like conditions may be partly a response to anthropogenic forcing, even though existing climate model and paleoclimate evidence suggest that trends will eventually reverse towards more El-Niño-like conditions, with an associated reversal in regional climate impacts.

1 Introduction

Earth's climatological pattern of sea-surface temperature (SST) plays a key role in shaping the large-scale atmospheric circulation and regional climate. In particular, the relative warmth of the Warm Pool in the western Indo-Pacific compared to the Cold Tongue in the eastern equatorial Pacific drives the Walker circulation in the tropical atmosphere, which through its impact on the upper tropospheric divergence in the Warm Pool generates large-scale atmospheric Rossby waves that propagate into higher latitudes and impact climate around the globe (Bjerknes, 1969; Sardeshmukh & Hoskins, 1988). This is part of a two-way coupling between the tropical atmosphere and ocean; the Walker circulation also helps shape the climatological SST pattern by driving upwelling of cold waters in the Cold Tongue and ocean heat-flux convergence in the Warm Pool (Bjerknes, 1969; Neelin et al., 1998).

In response to anthropogenic greenhouse gas forcing, climate models generally show a weakening of the Walker circulation (Vecchi et al., 2006) and enhanced warming in the eastern equatorial Pacific (Meehl & Washington, 1996). In contrast, SST observations show enhanced warming in the Indo-Pacific Warm Pool and weak cooling in the eastern equatorial Pacific over the 20th century (Cane et al., 1997; Solomon & Newman, 2012; Coats & Karnauskas, 2017) as well as a pronounced strengthening of the east-west SST gradient across the tropical Pacific since the mid 1970s (Wills et al., 2020; Watanabe et al., 2021). Sea-level pressure (SLP) observations show a weakening of the Walker circulation over the twentieth century (Vecchi et al., 2006; Tokinaga et al., 2012), however, the Walker circulation has strengthened since 1979 (L'Heureux et al., 2013; Kociuba & Power, 2015; Ma & Zhou, 2016; Chung et al., 2019; Zhao & Allen, 2019), in contrast to climate model hindcasts over this period (Fig. 1). This period has also been characterized by Southern

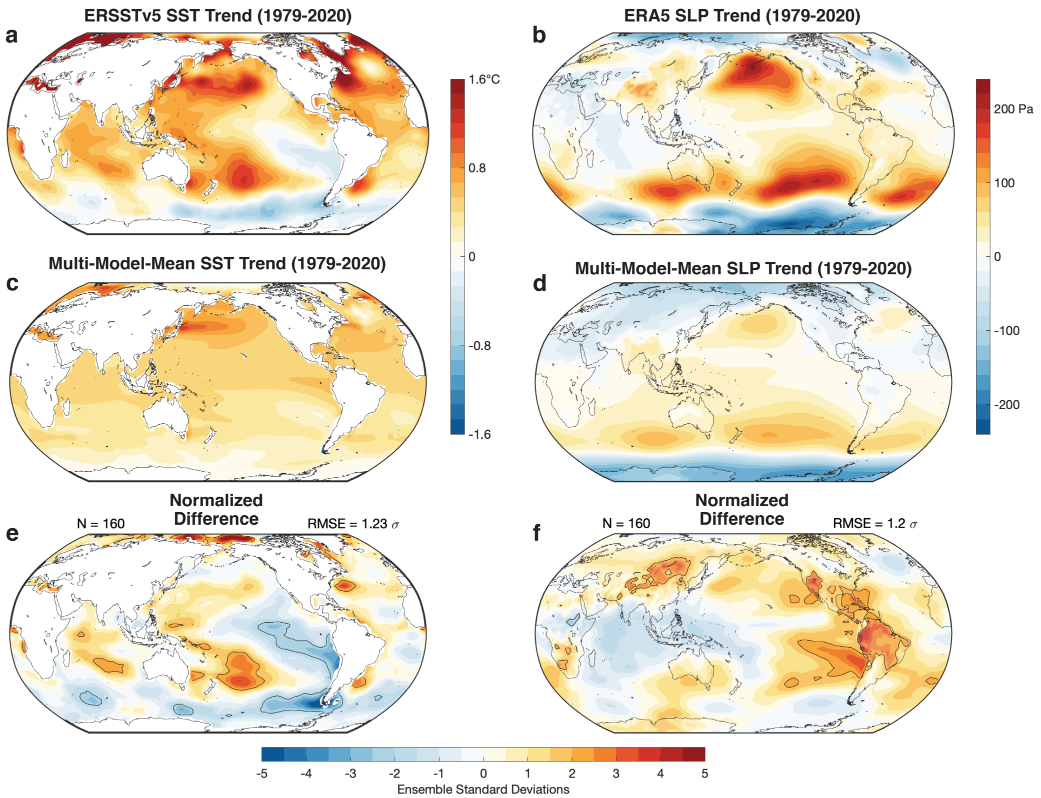

Figure 1. Observed trends (per 41 yr) in annual-mean (a) SST and (b) SLP over 1979–2020 from ERSSTv5 (Huang et al., 2017) and the ERA5 reanalysis (Hersbach et al., 2020), respectively. Modeled trends in (c) SST and (d) SLP over 1979–2020, from the multi-model mean of historical simulations with 16 climate model LEs (Table 1). The SST trends in each LE have been rescaled such that their global mean matches that in ERSSTv5. Observed trends in (e) SST and (f) SLP over 1979–2020, expressed in ensemble standard deviations away from the multi-model mean (i.e., the difference in trends between observations and the multi-model mean divided by the square root of the average variance in trends within LEs). Panels (c)-(f) are computed with the first 10 members of each large ensemble such that models are weighted equally. The ± 2 standard deviation contour is shown with a black line. The root mean square error (RMSE) of the maps in (e) and (f) are shown in the upper right.

Ocean cooling and sea-ice expansion, in contrast to the anthropogenically forced changes in climate models (Turner & Overland, 2009; Fan et al., 2014).

It remains an open question whether the differences in recent multi-decadal trends between observations and models resulted from anomalous multi-decadal variability or from aspects of the forced climate response not captured by models. Some studies suggest that these difference in Pacific and Southern Ocean trends could have resulted from internal atmosphere-ocean variability (Zhao & Allen, 2019; Chung et al., 2019; L. Zhang et al., 2019; Olonscheck et al., 2020; Watanabe et al., 2021; Chung et al., 2022), while others suggest they result in part from model biases in the pattern of response to external forcing (Thompson et al., 2011; Bintanja et al., 2013; Kohyama et al., 2017; Coats & Karnauskas, 2018; D. P. Schneider & Deser, 2018; Kostov et al., 2018; Seager et al., 2019; Wills et al., 2020; Suarez-Gutierrez et al., 2021; Seager et al., 2022). It is critical to distinguish between these hypotheses in order to predict future SST trends and their impact on the atmospheric circulation.

2 Climate model large ensembles unable to reproduce observed trends

Here we leverage a recent proliferation of climate model data from initial-condition large ensembles (Deser, Lehner, et al., 2020) to evaluate the potential for internal multi-decadal variability to explain the mismatch between observed and modeled trends in recent decades, focusing on the well-observed period since 1979 during which the Walker circulation and Pacific SST gradient trends are particularly anomalous. In initial-condition large ensembles (LEs), the same model is run multiple times with the same forcing but small differences in the initial condition, such that each ensemble member shows a different realization of internal variability. The ensemble mean thus shows the forced climate response, while the ensemble spread shows the range of possible realizations due to internal variability. We analyze annual-mean SST and SLP in simulations from 16 climate models that have at least 10 ensemble members for the period 1979–2020 under historical and future forcing scenarios (Table 1). Historical simulations only extend to 2005 (CMIP5) or 2014 (CMIP6), and different ensembles use different scenarios afterwards, namely RCP8.5, SSP2-4.5, SSP3-7.0, and SSP5-8.5, but differences between these scenarios are small through the year 2020. We compare modeled trends against observational SST data from the Extended Reconstructed SST dataset v5 (ERSSTv5) (Huang et al., 2017), the COBE SST dataset (Ishii et al., 2005), and the Atmospheric Model Intecomparison Project SST boundary condition dataset (AMIPII) (Hurrell et al., 2008), and SLP data from the ERA5 (Hersbach et al., 2020) and JRA55 (Kobayashi et al., 2015) reanalyses. All model output and observational data are linearly interpolated to a common 1.5° analysis grid.

Table 1. CMIP5 and CMIP6 LEs that cover the period 1979-2020, the scenarios used, and the number of ensemble members (N, minimum of the two scenarios used). The experimental setups and forcing scenarios for the CMIP5 (top) and CMIP6 simulations (bottom) are described in (Taylor et al., 2012) and (Eyring et al., 2016), respectively.

Model	Scenarios	N	Reference
CESM1.1	Historical, RCP8.5	40	Kay et al. (2015)
CanESM2	Historical, RCP8.5	50	Kirchmeier-Young et al. (2017)
CSIRO-Mk3.6	Historical, RCP8.5	30	Jeffrey et al. (2013)
GFDL-CM3	Historical, RCP8.5	20	Sun et al. (2018)
GFDL-ESM2M	Historical, RCP8.5	30	Rodgers et al. (2015)
MPI-ESM	Historical, RCP8.5	100	Maher et al. (2019)
ACCESS-ESM1.5	Historical, SSP2-4.5	13	Ziehn et al. (2020)
CanESM5	Historical, SSP3-7.0	25	Swart et al. (2019)
CESM2.1	Historical, SSP3-7.0	99	Rodgers et al. (2021)
CNRM-CM6.1	Historical, SSP2-4.5	10	Voldoire et al. (2019)
EC-Earth3	Historical, SSP5-8.5	50	Wyser et al. (2021)
GISS-E2.1-G	Historical, SSP3-7.0	10	Kelley et al. (2020)
IPSL-CM6A-LR	Historical, SSP3-7.0	11	Boucher et al. (2020)
MIROC6	Historical, SSP5-8.5	50	Tatebe et al. (2019)
MIROC-ES2L	Historical, SSP2-4.5	30	Hajima et al. (2020)
NorCPM1	Historical, SSP2-4.5	30	Bethke et al. (2021)

The multi-model-mean SST trends over 1979–2020 (Fig. 1c) are relatively spatially uniform except for enhanced warming in the North Pacific and muted warming in the

North Atlantic warming hole and the Southern Ocean. Compared to the multi-model mean, observed SST trends (Fig. 1a) show much larger warming in the northwest Atlantic and southwest Pacific, cooling instead of warming in the Southern Ocean (Fig. 2c), and opposite trends in the zonal SST gradient in the tropical Pacific (Fig. 2a). Note that in this comparison of modeled and observed SST trends over 1979–2020, we have rescaled SST trends in each model such that the global-mean SST trend matches that in ERSSTv5 over that period, effectively removing differences in global-mean warming rate and focusing instead of differences in the pattern of SST trends.

The SLP trends over 1979–2020 in ERA5 reanalysis (Fig. 1b) and models (Fig. 1d) both show positive (anticyclonic) trends in the midlatitude oceans and negative (cyclonic) trends in the high latitudes, but the trends in the midlatitude oceans are much larger in observations than in models, and observations show a strengthening of the Walker circulation, as measured by the zonal SLP gradient across the equatorial Pacific, that is not seen in models (Fig. 2c). Global-mean SLP trends are retained in the analysis, because absolute surface pressure is one of many variables assimilated in ERA5. There is a global-mean SLP trend of 20.6 Pa (41 yr)$^{-1}$ in ERA5, compared to -0.3 Pa (41 yr)$^{-1}$ in the multi-model mean, potentially related to the lack of mass conservation in the reanalysis. Removing the global-mean SLP trend would serve to shift the observed trends towards more negative values, while preserving the range of values.

To analyze how internal variability could have contributed to the differences in trends between observations and models, we calculate where the observations lie within the distribution of trends simulated by the LEs (Fig. 1e,f). To do so, we divide the difference in trends (observations minus multi-model mean) by the multi-model ensemble standard deviation (i.e., the square root of the ensemble mean of the variance in trends within each LE). If the observations were consistent with the forced response and internal variability as represented in the models, and the distribution of anomalies due to internal variability is well-described by a Gaussian, then there would only be a $\sim$5% chance of observing a normalized difference more extreme than $\pm$2 ensemble standard deviations. However, observed trends in many regions lie well in the tails of what is possible in models, including the strong observed warming in the Indian Ocean, West Pacific, South Pacific Convergence Zone (SPCZ), and Gulf Stream, the observed cooling in the Southern Ocean and southeast Pacific, and the observed increase in SLP in the eastern Pacific, the Caribbean, South America, and the Mongolian Plateau (note however that SLP over topography is sensitive to the surface air temperature used in the adjustment to mean sea level). Differences in trends (from the multi-model mean) this extreme are very unlikely (<5% probability) to occur within the models.

The same basic patterns of trend differences (in ensemble standard deviations) can be found by comparing observations to each LE separately (Figs. S1 and S2, where CESM2 and MPI-ESM show the smallest discrepancies from observations) or when using different observational products (Fig. S3). The pattern of SST trend differences can be found in both boreal winter and boreal summer (Fig. S4 and S5), though the SST trends in the South Pacific are more anomalous in austral winter. The pattern of SLP trend differences differs between boreal winter and boreal summer (Fig. S4 and S5), but both seasons show anomalous Walker circulation strengthening compared to the multi-model mean. Observed trends over a longer time period (1958-2021) are even more anomalous on average compared to the trends simulated by the LEs (Fig. S6), though the trends in the Southern Ocean SST and Walker circulation strength are more consistent with models over this time period.

The unusual nature of the observed trends compared to what is possible in coupled climate models is also apparent in a number of key climate indices including the Pacific SST gradient (Fig. 2a), Walker circulation (Fig. 2d), and Southern Ocean SST (Fig. 2c). The relative rate of Indo-Pacific Warm Pool warming (per degree of tropical-mean SST change), which plays a key role in global radiative feedbacks (Dong et al., 2019), is particularly anomalous (Fig. 2b), with most models showing trends of near 1 °C (°C)$^{-1}$ (i.e., Warm Pool

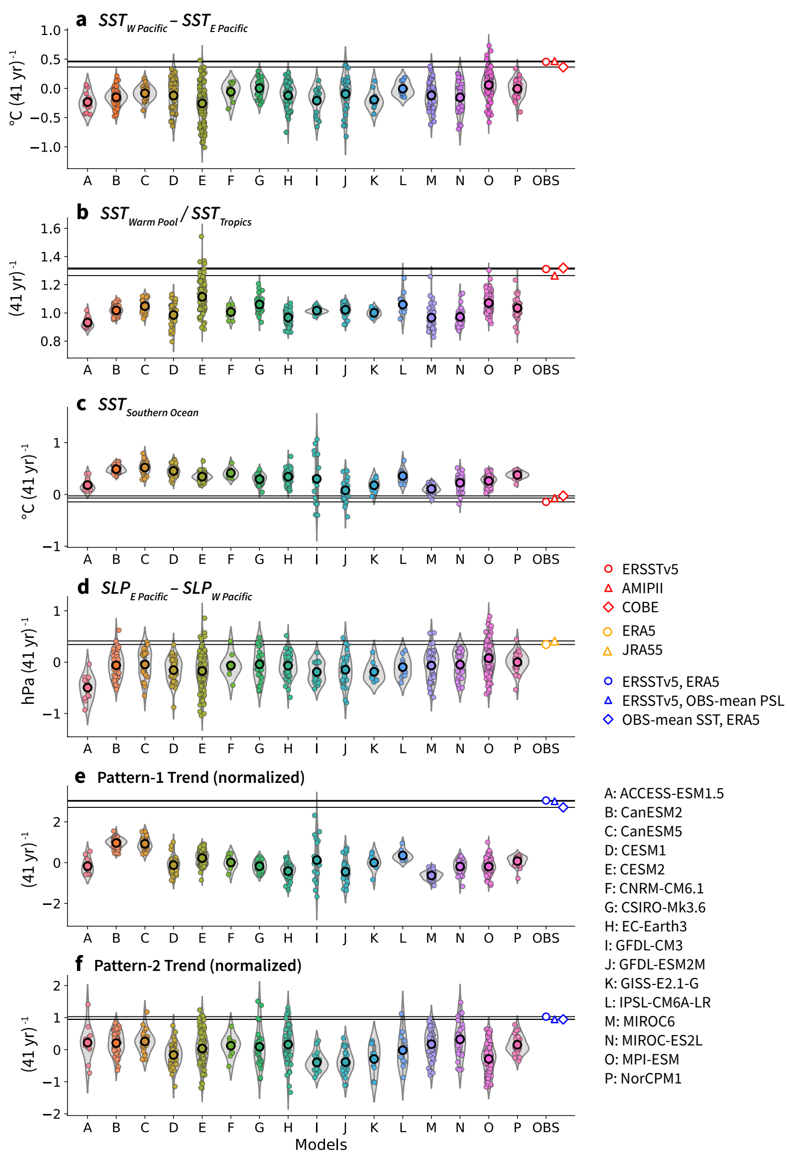

Figure 2. Comparison of observed trends (1979–2020) in key SST and SLP indices with those in all ensemble members from 16 LEs: (a) the Pacific SST gradient index used in Watanabe et al. (2021), defined as the difference between the western equatorial Pacific (5°S-5°N, 110°E-180°) and eastern equatorial Pacific (5°S-5°N, 180°-80°W); (b) the ratio of Indo-Pacific Warm Pool (30°S-30°N, 50°E-160°W) SST warming to tropical-mean (30°S-30°N) SST warming; (c) Southern Ocean SST (45°S-75°S); (d) Walker Circulation strength, defined as in Vecchi et al. (2006) as the difference in SLP between the eastern equatorial Pacific (5°S-5°N, 160°W-80°W) and western equatorial Pacific (5°S-5°N, 80°E-160°E); (e) and (f) the signal-to-noise maximizing pattern indices shown in Fig. 3. 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. See Fig. S7 for a map of the averaging regions.

warming rate equal to the tropical average), whereas observations show trends of around 1.3 °C $(°C)^{-1}$, which are only reproduced in a few ensemble members of one model (CESM2).

Previous studies have also reported anomalous observed trends in related metrics such as the warming in tropical convective regions (Fueglistaler & Silvers, 2021) or the tropical inter-basin warming contrast (L. Zhang & Karnauskas, 2017). There are also large discrepancies between observed and modeled SST trends in the southwest Pacific (Fig. S8), a discrepancy which has not been previously identified. The observed trends in this region (which has been referred to as the Southern Blob) have been linked to Southern Hemisphere SLP trends and an ongoing drought in Chile (Garreaud et al., 2019, 2021).

3 Isolating the observed pattern of change not reproduced in models

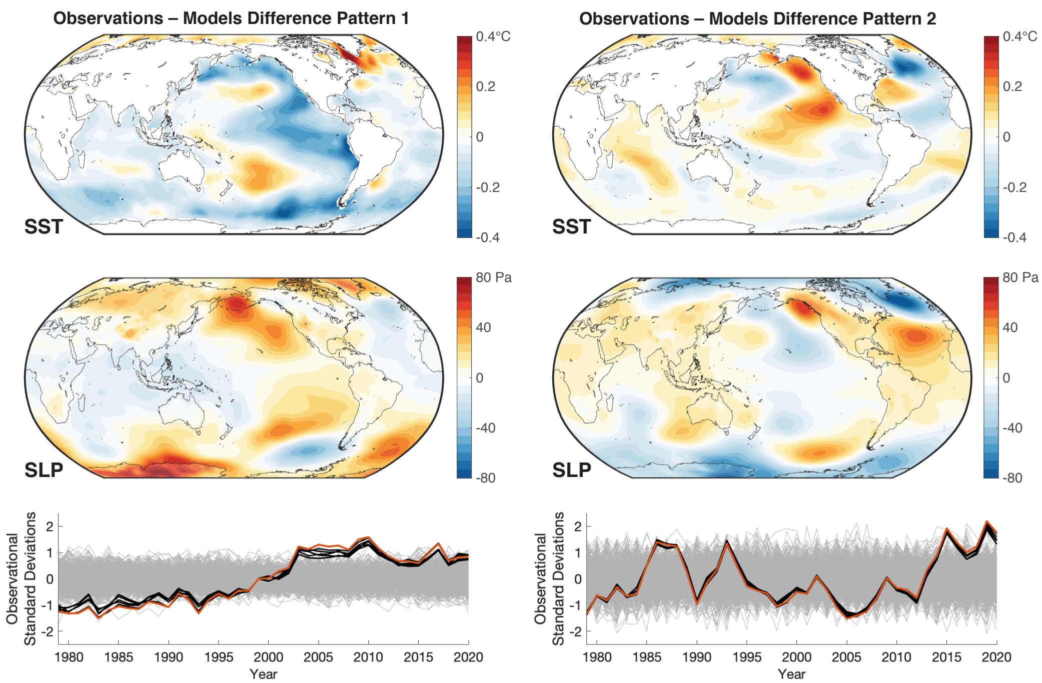

Figure 3. First and second multi-field (SST and SLP) signal-to-noise maximizing patterns, from an analysis that maximizes the ratio of signal to noise in the leading patterns, where signal is defined as the difference between observations and the multi-model mean on 5-year and longer timescales and noise consists of inter-model differences, inter-ensemble-member differences, and sub-5-year variability, with 20 EOFs included in the analysis (see Fig. S9 for the sensitivity to the number of EOFs included). The orange (black) lines show the amplitude of anomalies in these patterns in ERSSTv5/ERA5 (and other combinations of observational products) relative to the multi-model mean. The grey lines show the amplitude of these patterns in each member of the multi-model large ensemble. Normalization is such that the orange line has unit standard deviation and the SST/SLP pattern shows the anomalies associated with a 1-standard-deviation anomaly in the associated index.

To isolate the time varying SST and SLP anomalies contributing to the discrepancy between modeled and observed trends, we use a signal-to-noise maximizing pattern analysis (Déqué, 1988; T. Schneider & Griffies, 1999; Ting et al., 2009; Wills et al., 2020). Our goal is to identify the aspects of observed SST and SLP variability and trends over 1979-2020 that are not captured by any of the simulations in the multi-model large ensemble. In this way, we can highlight aspects of the observed trends that are least consistent with the variability and change simulated by models.

To do so we generate a difference ensemble, where each member is composed of the difference between observations and an individual member of one of the LEs, with 10 ensemble members used from each model (160 members total). The ensemble mean of the difference ensemble is thus the difference between observations and the multi-model mean, and the ensemble variance arises from inter-model and inter-ensemble-member differences within the multi-model large ensemble. We then solve for patterns with the maximum signal-to-noise ratio (SNR), 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 (i.e., due to inter-model differences, inter-ensemble-member differences, and sub-5-year variability). We use a 5-year Lanczos lowpass filter in defining the signal to focus on low-frequency differences between observations and models that could contribute to the multi-decadal trends. Unlike in the analysis for Fig. 1, we do not use an SST rescaling to account for differences in global-mean warming rate between models, because this would also modify the amplitude of internal variability. SST and SLP are normalized by their total variance such that they are weighted equally in the analysis.

The leading SST/SLP pattern shows less warming than models in a triangular region in the eastern tropical and subtropical Pacific, the Pacific sector of the Southern Ocean, and the subpolar North Pacific; more warming than models in the Labrador and Irminger Seas and the southwest Pacific; strengthening of the Walker Circulation and weakening of the Aleutian Low compared to models; and anomalies associated with the Pacific-South American pattern (Fig. 3). The patterns of anomalies are similar to those in Fig. 1e,f but are expressed in absolute units instead of being normalized by the ensemble standard deviation of trends. They therefore show the actual magnitude of anomalies (compared to the multi-model mean) that occurred in the observational data. This SST/SLP pattern increased monotonically from 1979 through 2003, then has shown relatively little change since 2003 (Fig. 3). Despite the lack of changes in this pattern since 2003, its trend over the full time-period is still highly anomalous compared to trends in this pattern in models (Fig. 2e), with none of the 598 ensemble members reproducing the observed trend in this pattern. Only one ensemble member of one model (GFDL-CM3) comes anywhere close to the observed trends, owing to the large amplitude of Southern Ocean multi-decadal variability in versions of the GFDL model (L. Zhang et al., 2017; Wills et al., 2021).

The second pattern of SST anomalies is focused in the North Pacific and North Atlantic, potentially showing differences in the patterns of SST variability in these regions between observations and models (Fig. 3). There are also large SLP anomalies in the North Atlantic, indicating a long-term trend towards positive anomalies in the North Atlantic Oscillation (NAO) in observations compared to models, which is part of an anomalous trend in the NAO since the 1960s. Together, the SST and SLP anomaly patterns are consistent with the forcing of upper-ocean temperature anomalies by the atmosphere (Battisti et al., 2019). The time evolution of this SST/SLP pattern shows decadal variability, a trend over the full time-period that only a handful of ensemble members reproduce (Fig. 2f), and a particularly anomalous trend since 2005. This pattern thus shows how observed decadal variability differs from that in models, as well as including part of the observations-models difference in the monotonic trend over 1979-2020. Interestingly, the model that is closest to reproducing trends in pattern 1 (GFDL-CM3) is furthest from reproducing trends in pattern 2 and trends in the relative Warm Pool warming rate (Fig. 2).

The signal-to-noise maximizing analysis used here is designed to identify patterns that are highly anomalous in observations compared to models, so a question naturally arises regarding the extent to which it is guaranteed to find something even if the models were capable of reproducing observations given enough realizations. To address this, we repeat this analysis on difference ensembles sampling inter-model differences and internal variability within the multi-model large ensemble (Supporting Information). We find that anomaly patterns with as high a SNR as that for pattern 1 in Fig. 3 commonly occur due to a combination of inter-model differences and internal variability within the LEs, but are unlikely

(<12.5% chance) to occur due to internal variability alone (Fig. S10). Even when patterns this anomalous do occur, they are different patterns than found for the difference between models and observations, and they rarely have such large trends over the full time period (<4% chance for internal variability alone).

4 Summary and Discussion

We have shown that observed SST and SLP trends over 1979–2020 are highly anomalous in several regions (Fig. 1) and indices (Fig. 2) compared to those simulated through a combination of forced response and internal variability within a multi-model large ensemble. Our results illustrate that there are systematic climate model biases in large-scale multi-decadal SST and SLP trends, despite the fact that climate models can reproduce observed SST trends over shorter time periods or when considering long-term trends in smaller-scale tropical climate indices with large amounts of internal variability (Coats and Karnauskas (2017); Chung et al. (2019); Olonscheck et al. (2020); Watanabe et al. (2021); Fig. 2a,d; see also Seager et al. (2019, 2022)). The anomalous trends can be encapsulated in large-scale SST and SLP patterns that are undoubtedly outside the range of what can be reproduced in climate models (Figs. 3, 2e, 2f). However, with only a single realization of the real climate system, it remains difficult to robustly identify the forced response in observations, meaning that these trend differences could result either from systematic model biases in the transient response to historical forcing (e.g., Seager et al. (2019)) or from model biases in the amplitude or pattern of multi-decadal variability (e.g., Laepple and Huybers (2014)).

4.1 Possible Interpretations

Many previous studies have invoked negative trends in the Pacific Decadal Oscillation (PDO) as an explanation for the anomalous pattern of trends in observations (e.g., Trenberth and Fasullo (2013); Chung et al. (2019)). However, with a return towards positive PDO conditions between 2013 and 2020, trends are no longer as anomalous in the North Pacific, while they remain anomalous in the South Pacific (Fig. 1e), suggesting that the PDO is not the primary driver of the trend discrepancy for the full 1979–2020 period.

It has also been suggested that observed trends in the Southern Ocean could result from an anomalous phase of Southern Ocean multi-decadal variability (SOMV, e.g., L. Zhang et al. (2019)). This remains plausible, though its relevance for lower latitude SST trends depends on an active body of work to quantify the teleconnections from Southern Ocean SST changes (Hwang et al., 2017; Kang et al., 2019; X. Zhang et al., 2021; Dong et al., 2022). Furthermore, there are several mechanisms for how recent Southern Ocean cooling and sea ice expansion could result from anthropogenic forcing, including wind shifts due to a combination of greenhouse gas and ozone forcing (Thompson et al., 2011; Kostov et al., 2018) and increased surface stratification resulting from precipitation changes and/or ice-sheet melt (Bintanja et al., 2013; De Lavergne et al., 2014; Pauling et al., 2016; Purich et al., 2018), the latter of which is not included in CMIP models. Specifying observed winds or adding meltwater forcing to a climate model both shift the SST trend pattern closer to observations (Dong et al., 2022), but discrepancies in winds or meltwater forcing could result from a biased/missing forced response or from multi-decadal variability, so the ultimate cause of the observed Southern Ocean cooling trend remains an open question.

The large difference in the relative Warm Pool warming rate between models and observations (Fig. 2b) is particularly hard to explain with internal variability. CMIP models show little multi-decadal variability in Warm Pool SST, because the strong damping feedbacks in response to surface warming in this region precludes persistent SST anomalies without either large energy budget anomalies or compensating feedbacks in other regions (Wills et al., 2021). A bias in the transient response of the tropical Pacific to greenhouse gas forcing could result from an ocean dynamical thermostat mechanism (Clement et al., 1996) in the eastern equatorial Pacific that is too weak in models (Seager et al., 2019, 2022), model biases in the response to geographic changes in aerosol optical depth over this time period (Smith et al., 2016; Deser, Phillips, et al., 2020; Heede & Fedorov, 2021; Shi et al., 2022), or remote influences of biases in the North Atlantic (McGregor et al., 2018) or Southern Ocean (as discussed in the previous paragraph). Another possibility is that multi-decadal variability of tropical and subtropical SSTs is much too weak in models, as suggested by a comparison to paleoclimate proxies (Laepple & Huybers, 2014). Further, we hypothesize that the damping feedbacks in response to Warm Pool warming could be too efficient in models (e.g., Keil et al. (2021)), which would reduce both the modeled warming rate and the modeled amplitude of multi-decadal variability in this region.

Building on methods to isolate the forced response in observations (Wills et al., 2020), our analysis in Fig. 3 identifies the patterns that distinguish models and observations on long timescales, in an attempt to detect the difference in forced response between models and observations from amongst the noise of internal multi-decadal variability. There presumably remains an unquantifiable contribution of multi-decadal variability to these anomaly patterns. However, the large magnitude of multi-decadal trends in these patterns compared to what is found in models, together with the projection of pattern 1 onto the ratio of Indo-Pacific Warm Pool to tropical-mean warming, which shows little multi-decadal variability in CMIP models, lead us to conclude that it is extremely unlikely that this pattern of trend discrepancies results entirely from internal variability. Our analysis provides a starting point for more detailed mechanistic analysis to understand where model biases in the forced response are contributing.

4.2 Implications for Future Trends

Regardless of whether the differences in observed and modeled trends results from internal variability or biases in the transient response to forcing, modeling and paleoclimate studies (Fedorov et al., 2015; Armour et al., 2016; Tierney et al., 2019; Heede & Fedorov, 2021) suggest the East Pacific and Southern Oceans will eventually warm. Eventual warming in these regions favors more positive radiative feedbacks (Andrews et al., 2015; Dong et al., 2020), leading to an increase in effective climate sensitivity. That observations have shown much less warming in these regions than almost any model suggests that this so-called pattern effect on climate sensitivity could be even larger in the real world than in models, potentially leading to a shift towards much higher effective climate sensitivity at some unknown point in the future. Similarly, a future shift towards a more El-Niño-like warming pattern, with more warming in the eastern equatorial Pacific, would lead to major changes in the Walker circulation and shifts in the associated extratropical circulation and precipitation patterns. Without understanding why large-scale SST and SLP trends are so anomalous over the recent observational period or when and by how much delayed warming regions in the East Pacific and Southern Oceans will warm, we are left with a huge source of uncertainty in multi-decadal projections of regional and global climate.

Data Availability Statement

CMIP5 LE data were obtained from the U.S. CLIVAR Multi-Model Large Ensemble Archive, which can be downloaded following instructions at `https://www.cesm.ucar.edu/projects/community-projects/MMLEA/`. CMIP6 LE data were obtained from the CMIP6 next-generation archive at ETH Zurich (Brunner et al., 2020). CESM2 LE data were obtained from the National Center for Atmospheric Research following instructions at `https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html`.

All observational data are publicly available. ERSSTv5 and COBE SST data were obtained from NOAA/OAR/ESRL PSL (https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html; https://psl.noaa.gov/data/gridded/data.cobe.html). ERA5 data were obtained from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysisera5-single-levels-monthly-means). JRA55 data were obtained from the NCAR Research Data Archive (https://rda.ucar.edu/datasets/ds628.1/).

The code used for the analysis in this paper is available at `https://github.com/rcjwills/pattern-biases/` and will be transferred to Zenodo upon acceptance of the paper.

Acknowledgments

R.C.J.W and D.S.B. were supported by the National Science Foundation grants AGS-1929775 and AGS-2128409. R.C.J.W, C.P., and K.C.A. were supported by National Oceanic and Atmospheric Administration MAPP Program Award NA20OAR4310391. Y.D. was supported by the NOAA Climate and Global Change Postdoctoral Fellowship Program, administered by UCAR's Cooperative Programs for the Advancement of Earth System Science (CPAESS) under award NA21OAR4310383. Y.D. and K.C.A. were supported by National Science Foundation Grant AGS-1752796. K.C.A. was supported by an Alfred P. Sloan Research Fellowship (grant number FG-2020-13568).

References

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, *28*(4), 1630–1648.

Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A., & Newsom, E. R. (2016). Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geoscience*, *9*(7), 549–554.

Battisti, D. S., Vimont, D. J., & Kirtman, B. P. (2019). 100 years of progress in understanding the dynamics of coupled atmosphere–ocean variability. *Meteorological Monographs*, *59*, 8.1–8.57.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., ... others (2021). NorCPM1 and its contribution to CMIP6 DCPP. *Geoscientific Model Development*, *14*(11), 7073–7116.

Bintanja, R., van Oldenborgh, G. J., Drijfhout, S., Wouters, B., & Katsman, C. (2013). Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion. *Nature Geoscience*, *6*(5), 376–379.

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly weather review*, *97*(3), 163–172.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... others (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.*, *12*(7), e2019MS002010.

Brunner, L., Hauser, M., Lorenz, R., & Beyerle, U. (2020). The ETH Zurich CMIP6 next generation archive: technical documentation. *ETH Zürich, Zürich*, *10*. doi: 10.5281/zenodo.3734128

Cane, M. A., Clement, A. C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., ... Murtugudde, R. (1997). Twentieth-century sea surface temperature trends. *Science*, *275*(5302), 957–960.

Chung, E.-S., Kim, S.-J., Timmermann, A., Ha, K.-J., Lee, S.-K., Stuecker, M. F., ... Huang, L. (2022). Antarctic sea-ice expansion and Southern Ocean cooling linked to tropical variability. *Nature Climate Change*, 1–8.

Chung, E.-S., Timmermann, A., Soden, B. J., Ha, K.-J., Shi, L., & John, V. O. (2019). Reconciling opposing Walker circulation trends in observations and model projections. *Nature Climate Change*, *9*(5), 405–412.

Clement, A. C., Seager, R., Cane, M. A., & Zebiak, S. E. (1996). An ocean dynamical thermostat. *J. Climate*, *9*(9), 2190–2196.

Coats, S., & Karnauskas, K. (2017). Are simulated and observed twentieth century tropical Pacific sea surface temperature trends significant relative to internal variability?

Geophys. Res. Lett., *44*(19), 9928–9937.
Coats, S., & Karnauskas, K. (2018). A role for the equatorial undercurrent in the ocean dynamical thermostat. *J. Climate*, *31*(16), 6245–6261.
De Lavergne, C., Palter, J. B., Galbraith, E. D., Bernardello, R., & Marinov, I. (2014). Cessation of deep convection in the open southern ocean under anthropogenic climate change. *Nature Climate Change*, *4*(4), 278–282.
Déqué, M. (1988). 10-day predictability of the Northern Hemisphere winter 500-mb height by the ECMWF operational model. *Tellus A*, *40*(1), 26–36.
Deser, C., Lehner, F., Rodgers, K., Ault, T., Delworth, T., DiNezio, P., ... M, T. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 1–10.
Deser, C., Phillips, A. S., Simpson, I. R., Rosenbloom, N., Coleman, D., Lehner, F., ... Stevenson, S. (2020). Isolating the evolving contributions of anthropogenic aerosols and greenhouse gases: a new CESM1 large ensemble community resource. *J. Climate*, *33*(18), 7835–7858.
Dong, Y., Armour, K. C., Battisti, D. S., & Blanchard-Wrigglesworth, E. (2022). Two-way teleconnections between the Southern Ocean and the tropical Pacific via a dynamic feedback. *Journal of Climate, in press*.
Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., & Andrews, T. (2020). Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate*, *33*(18), 7755–7775.
Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's function approach: The preeminence of the Western Pacific. *J. Climate*, *32*(17), 5471–5491.
Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016
Fan, T., Deser, C., & Schneider, D. P. (2014). Recent Antarctic sea ice trends in the context of Southern Ocean surface climate variations since 1950. *Geophysical Research Letters*, *41*(7), 2419–2426.
Fedorov, A. V., Burls, N. J., Lawrence, K. T., & Peterson, L. C. (2015). Tightly linked zonal and meridional sea surface temperature gradients over the past five million years. *Nature Geoscience*, *8*(12), 975–980.
Fueglistaler, S., & Silvers, L. (2021). The peculiar trajectory of global warming. *Journal of Geophysical Research: Atmospheres*, *126*(4), e2020JD033629.
Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., & Veloso-Aguila, D. (2019). The central Chile mega drought (2010–2018): a climate dynamics perspective. *International Journal of Climatology*, *40*(1), 421–439.
Garreaud, R. D., Clem, K., & Veloso, J. V. (2021). The South Pacific pressure trend dipole and the southern blob. *J. Climate*, *34*(18), 7661–7676.
Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., ... others (2020). Development of the MIROC-ES2L earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244.
Heede, U. K., & Fedorov, A. V. (2021). Eastern equatorial Pacific warming delayed by aerosols and thermostat response to CO2 increase. *Nature Climate Change*, *11*(8), 696–703.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049.
Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... Zhang, H.-M. (2017). Extended reconstructed sea surface temperature, version

5 (ersstv5): upgrades, validations, and intercomparisons. *J. Climate*, *30*(20), 8179–8205.

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, *21*(19), 5145–5153.

Hwang, Y.-T., Xie, S.-P., Deser, C., & Kang, S. M. (2017). Connecting tropical climate change with Southern Ocean heat uptake. *Geophys. Res. Lett.*, *44*(18), 9449–9457.

Ishii, M., Shouji, A., Sugimoto, S., & Matsumoto, T. (2005). Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using icoads and the kobe collection. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *25*(7), 865–879.

Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., ... Syktus, J. (2013). Australia's CMIP5 submission using the CSIRO Mk3. 6 model. *Aust. Meteor. Oceanogr. J*, *63*, 1–13.

Kang, S. M., Hawcroft, M., Xiang, B., Hwang, Y.-T., Cazes, G., Codron, F., ... others (2019). Extratropical–tropical interaction model intercomparison project (ETIN-MIP): Protocol and initial results. *Bulletin of the American Meteorological Society*, *100*(12), 2589–2606.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., ... M, V. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteorol. Soc.*, *96*(8), 1333–1349.

Keil, P., Schmidt, H., Stevens, B., & Bao, J. (2021). Variations of tropical lapse rates in climate models and their implications for upper-tropospheric warming. *J. Climate*, *34*(24), 9747–9761.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., ... others (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS002025.

Kirchmeier-Young, M. C., Zwiers, F. W., & Gillett, N. P. (2017). Attribution of extreme events in Arctic sea ice extent. *Journal of Climate*, *30*(2), 553–571.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., ... others (2015). The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48.

Kociuba, G., & Power, S. B. (2015). Inability of CMIP5 models to simulate recent strengthening of the Walker circulation: Implications for projections. *J. Climate*, *28*(1), 20–35.

Kohyama, T., Hartmann, D. L., & Battisti, D. S. (2017). La Niña–like mean-state response to global warming and potential oceanic roles. *J. Climate*, *30*(11), 4207–4225.

Kostov, Y., Ferreira, D., Armour, K. C., & Marshall, J. (2018). Contributions of greenhouse gas forcing and the Southern Annular Mode to historical Southern Ocean surface temperature trends. *Geophysical Research Letters*, *45*(2), 1086–1097.

Laepple, T., & Huybers, P. (2014). Ocean surface temperature variability: Large model–data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences*, *111*(47), 16682–16687.

L'Heureux, M. L., Lee, S., & Lyon, B. (2013). Recent multidecadal strengthening of the Walker circulation across the tropical Pacific. *Nature Climate Change*, *3*(6), 571–576.

Ma, S., & Zhou, T. (2016). Robust strengthening and westward shift of the tropical Pacific Walker circulation during 1979–2012: A comparison of 7 sets of reanalysis data and 26 CMIP5 models. *J. Climate*, *29*(9), 3097–3118.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., ... others (2019). The Max Planck Institute grand ensemble: Enabling the exploration of climate system variability. *J. Adv. Model. Earth Sy.*.

McGregor, S., Stuecker, M. F., Kajtar, J. B., England, M. H., & Collins, M. (2018). Model tropical Atlantic biases underpin diminished Pacific decadal variability. *Nature Climate Change*, *8*(6), 493–498.

Meehl, G. A., & Washington, W. M. (1996). El Niño-like climate change in a model with increased atmospheric CO2 concentrations. *Nature*, *382*(6586), 56–60.
Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., & Zebiak, S. E. (1998). ENSO theory. *Journal of Geophysical Research: Oceans*, *103*(C7), 14261–14290.
Olonscheck, D., Rugenstein, M., & Marotzke, J. (2020). Broad consistency between observed and simulated trends in sea surface temperature patterns. *Geophys. Res. Lett.*, *47*(10), e2019GL086773.
Pauling, A. G., Bitz, C. M., Smith, I. J., & Langhorne, P. J. (2016). The response of the Southern Ocean and Antarctic sea ice to freshwater from ice shelves in an earth system model. *J. Climate*, *29*(5), 1655–1672.
Purich, A., England, M. H., Cai, W., Sullivan, A., & Durack, P. J. (2018). Impacts of broad-scale surface freshening of the Southern Ocean in a coupled climate model. *J. Climate*, *31*(7), 2613–2632.
Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... Yeager, S. G. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics Discussions*, *2021*, 1–22.
Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an earth system model. *Biogeosciences*, *12*(11), 3301–3320.
Sardeshmukh, P. D., & Hoskins, B. J. (1988). The generation of global rotational flow by steady idealized tropical divergence. *Journal of the Atmospheric Sciences*, *45*(7), 1228–1251.
Schneider, D. P., & Deser, C. (2018). Tropically driven and externally forced patterns of Antarctic sea ice change: Reconciling observed and modeled trends. *Climate Dynamics*, *50*(11), 4599–4618.
Schneider, T., & Griffies, S. M. (1999). A conceptual framework for predictability studies. *J. Climate*, *12*(10), 3133–3155.
Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019). Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, *9*(7), 517–522.
Seager, R., Henderson, N., & Cane, M. (2022). Persistent discrepancies between observed and modeled trends in the tropical Pacific Ocean. *J. Climate*, 1–41.
Shi, J.-R., Kwon, Y.-O., & Wijffels, S. E. (2022). Two distinct modes of climate responses to the anthropogenic aerosol forcing changes. *J. Climate*, *35*(11), 3445–3457.
Smith, D. M., Booth, B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., ... Thompson, V. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*(10), 936–940.
Solomon, A., & Newman, M. (2012). Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nature Climate Change*, *2*(9), 691–699.
Suarez-Gutierrez, L., Milinski, S., & Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. *Climate Dynamics*, *57*(9), 2557–2580.
Sun, L., Alexander, M., & Deser, C. (2018). Evolution of the global coupled climate response to arctic sea ice loss during 1990–2090 and its contribution to climate change. *J. Climate*, *31*(19), 7823–7843.
Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... others (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.
Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... others (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765.
Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485 - 498. doi: 10.1175/BAMS-D-11-00094.1

Thompson, D. W., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., & Karoly, D. J. (2011). Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature geoscience*, *4*(11), 741–749.
Tierney, J. E., Haywood, A. M., Feng, R., Bhattacharya, T., & Otto-Bliesner, B. L. (2019). Pliocene warmth consistent with greenhouse gas forcing. *Geophysical Research Letters*, *46*(15), 9136–9144.
Ting, M., Kushnir, Y., Seager, R., & Li, C. (2009). Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, *22*(6), 1469–1481.
Tokinaga, H., Xie, S.-P., Deser, C., Kosaka, Y., & Okumura, Y. M. (2012). Slowdown of the Walker circulation driven by tropical Indo-Pacific warming. *Nature*, *491*(7424), 439–443.
Trenberth, K. E., & Fasullo, J. T. (2013). An apparent hiatus in global warming? *Earth's Future*, *1*(1), 19–32.
Turner, J., & Overland, J. (2009). Contrasting climate change in the two polar regions. *Polar Research*, *28*(2), 146–164.
Vecchi, G. A., Soden, B. J., Wittenberg, A. T., Held, I. M., Leetmaa, A., & Harrison, M. J. (2006). Weakening of tropical pacific atmospheric circulation due to anthropogenic forcing. *Nature*, *441*(7089), 73–76.
Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., ... others (2019). Evaluation of CMIP6 deck experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, *11*(7), 2177–2213.
Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. doi: 10.21105/joss.03021
Watanabe, M., Dufresne, J.-L., Kosaka, Y., Mauritsen, T., & Tatebe, H. (2021). Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nature Climate Change*, *11*(1), 33–37.
Wills, R. C. J., Armour, K. C., Battisti, D. S., Proistosescu, C., & Parsons, L. A. (2021). Slow modes of global temperature variability and their impact on climate sensitivity estimates. *Journal of Climate*, *34*(21), 8717–8738.
Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T., & Deser, C. (2020). Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *J. Climate*, *33*(20), 8693–8719.
Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P., & Kruschke, T. (2021). The SMHI Large Ensemble (SMHI-LENS) with EC-Earth3.3.1. *Geoscientific Model Development*, *14*(7), 4781–4796.
Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2019). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, *9*(1), 59–65.
Zhang, L., Delworth, T. L., & Jia, L. (2017). Diagnosis of decadal predictability of Southern Ocean sea surface temperature in the GFDL CM2.1 model. *Journal of Climate*, *30*(16), 6309–6328.
Zhang, L., & Karnauskas, K. B. (2017). The role of tropical interbasin SST gradients in forcing Walker circulation trends. *J. Climate*, *30*(2), 499–508.
Zhang, X., Deser, C., & Sun, L. (2021). Is there a tropical response to recent observed Southern Ocean cooling? *Geophys. Res. Lett.*, *48*(5), e2020GL091235.
Zhao, X., & Allen, R. J. (2019). Strengthening of the Walker circulation in recent decades and the role of natural sea surface temperature variability. *Environmental Research Communications*, *1*(2), 021003.
Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., ... Srbinovsky, J. (2020). The Australian earth system model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214.

Supporting Information for "Systematic climate model biases in the large-scale pattern 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

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, 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 subsampling 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.

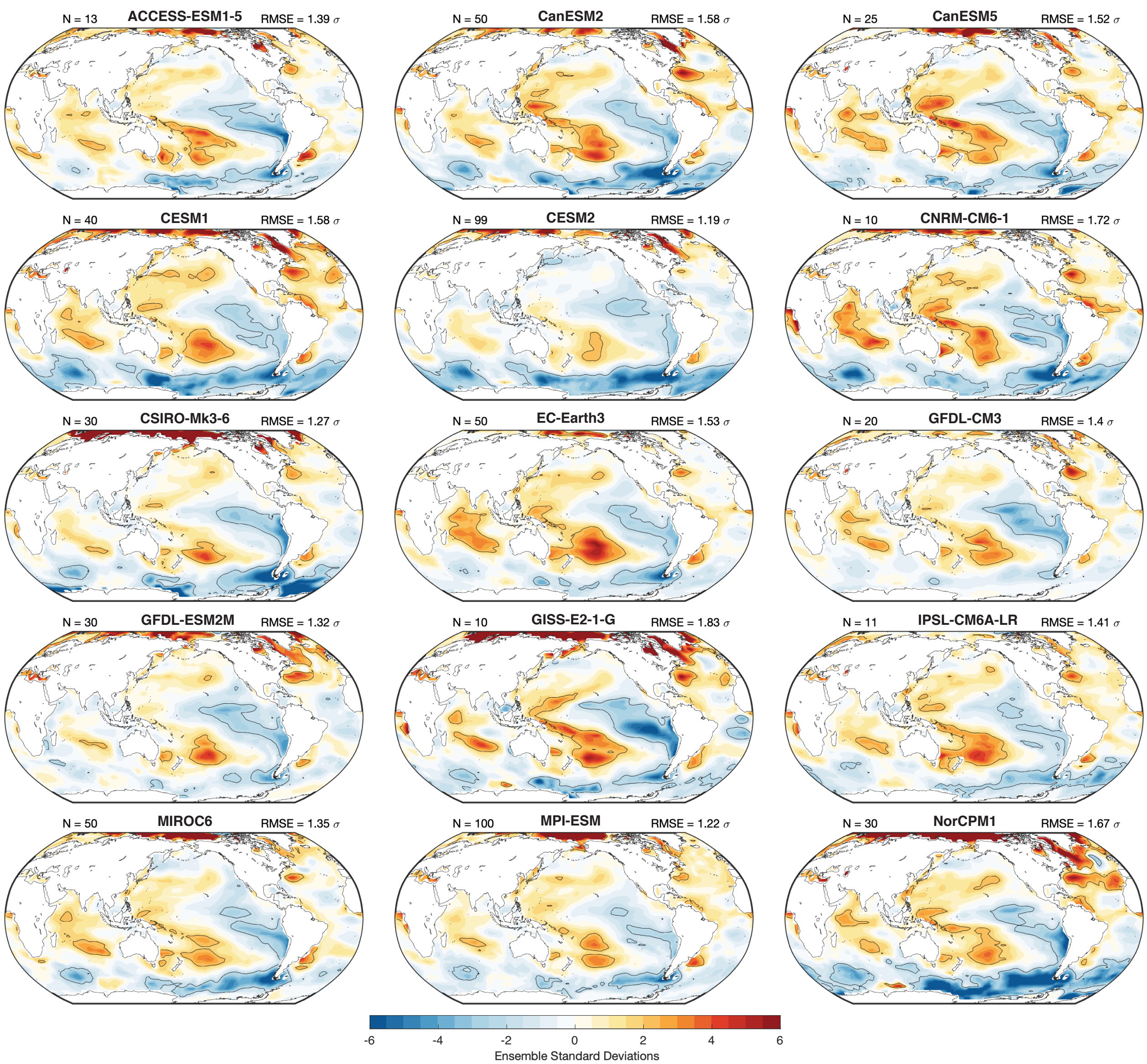

Figure S1. Same as Fig. 1e, but computed separately for each model. Unlike in Fig. 1, all ensemble members are used, and the number of ensemble members included is displayed in the upper left of each panel.

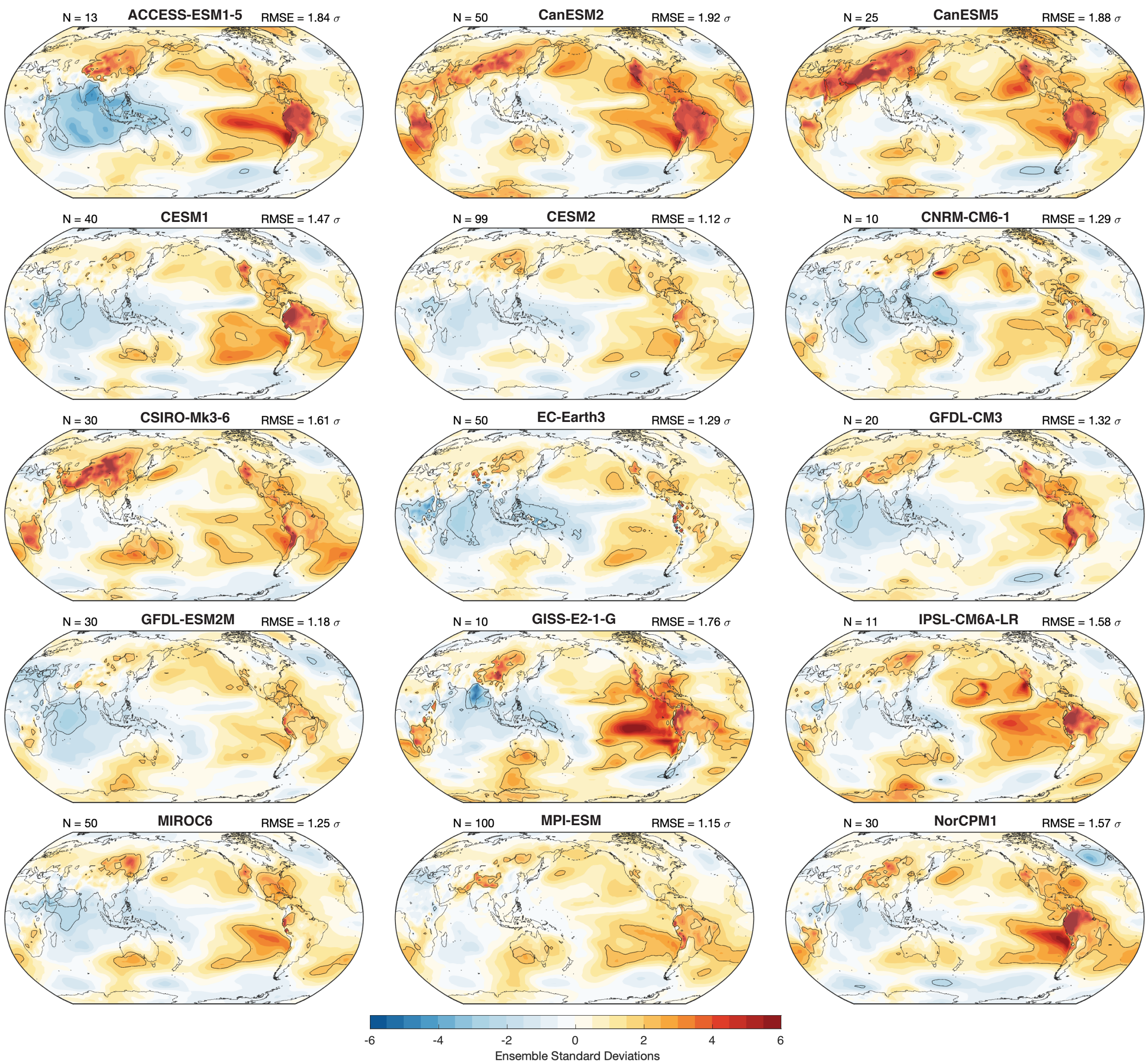

Figure S2. Same as Fig. 1f, but computed separately for each model. Unlike in Fig. 1, all ensemble members are used, and the number of ensemble members included is displayed in the upper left of each panel.

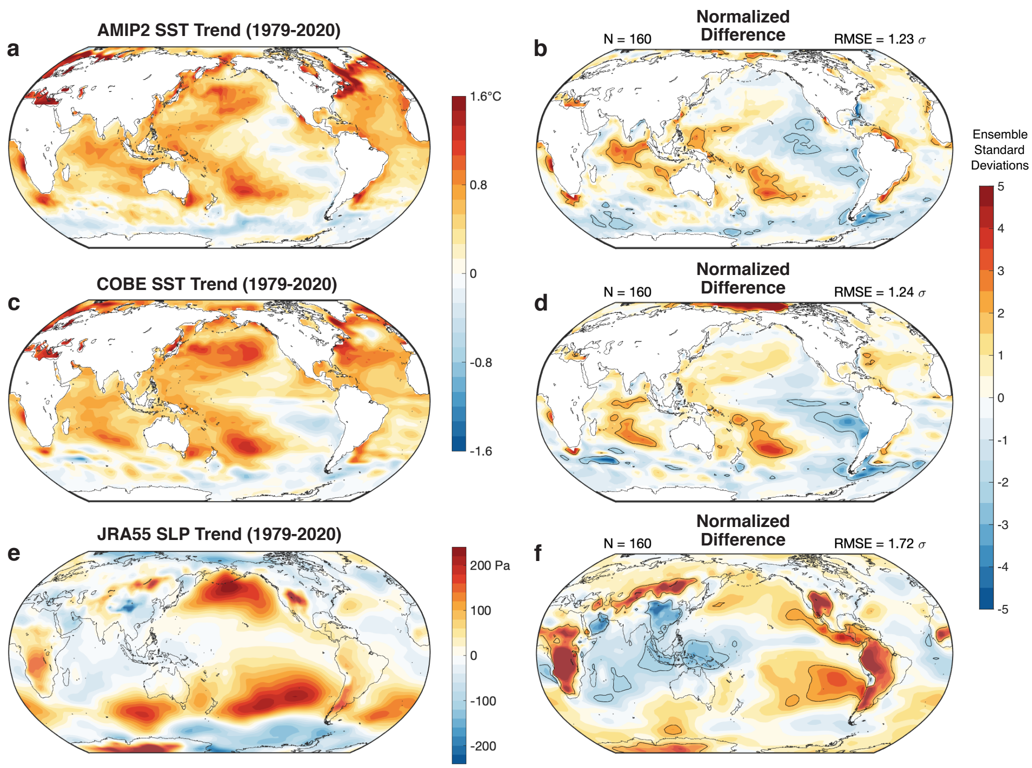

Figure S3. (a, b) Same as Fig. 1a and e, except for AMIPII instead of ERSSTv5. (c, d) Same as Fig. 1a and e, except for COBE instead of ERSSTv5. (e, f) Same as Fig. 1b and f, except for JRA55 instead of ERA5.

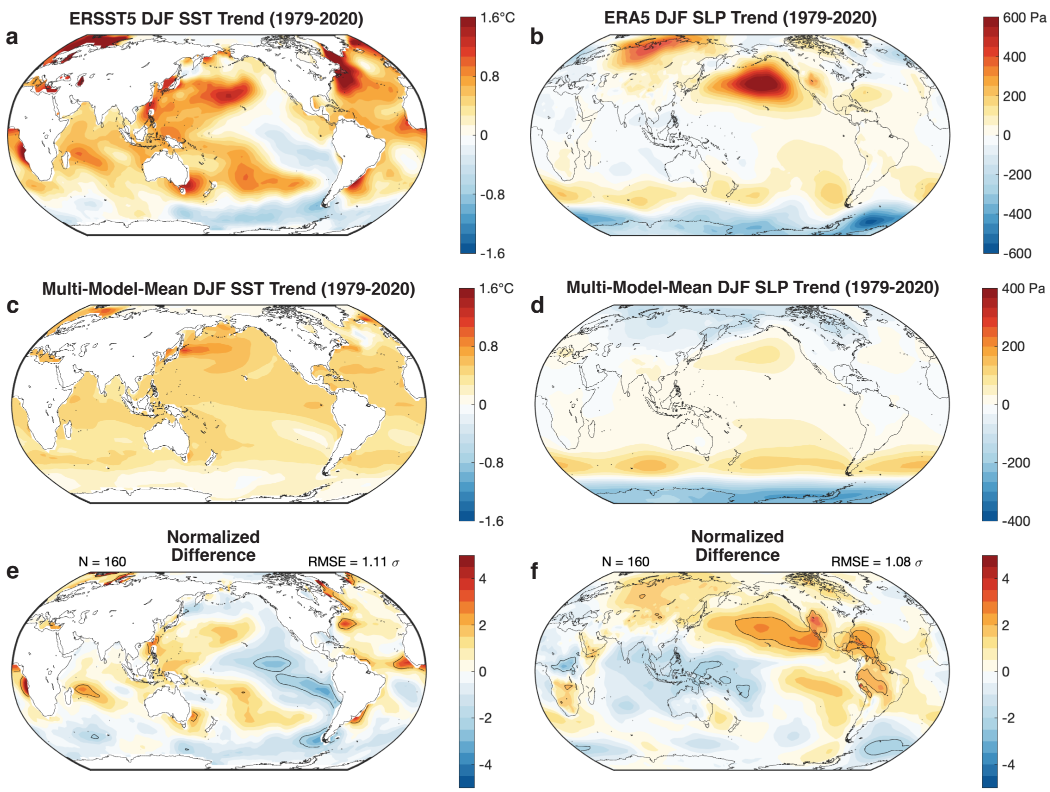

Figure S4. Same as Fig. 1, but for December-January-February (DJF) instead of annual mean.

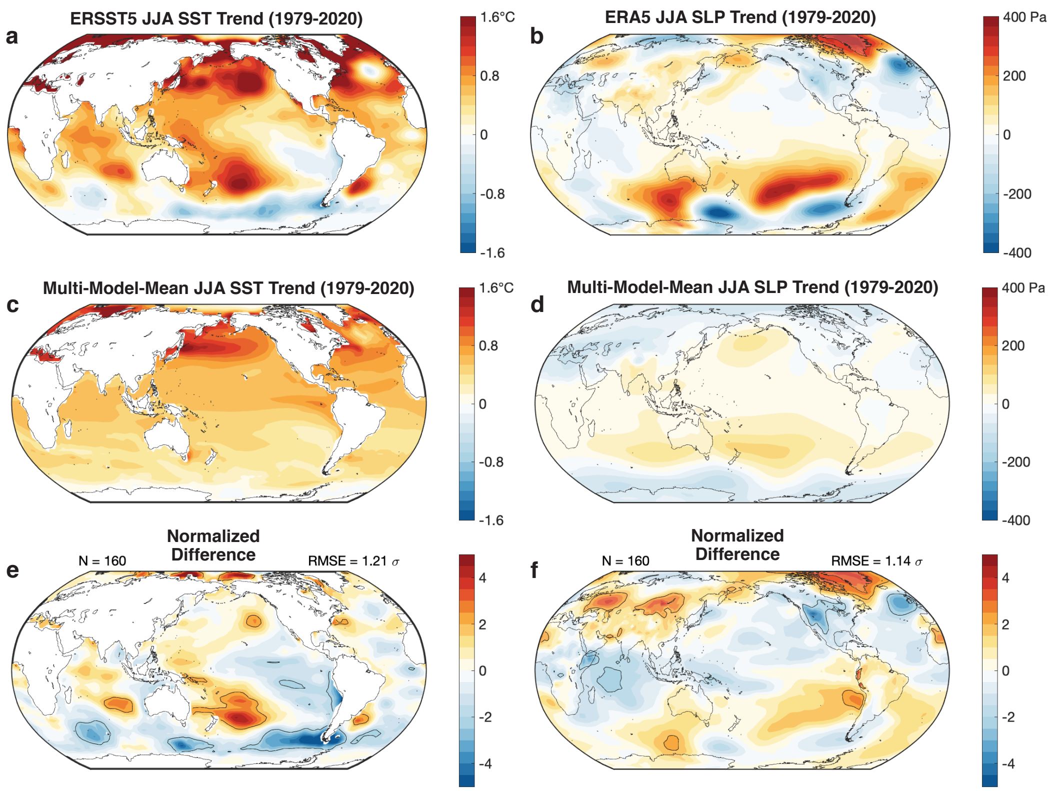

Figure S5. Same as Fig. 1, but for June-July-August (JJA) instead of annual mean.

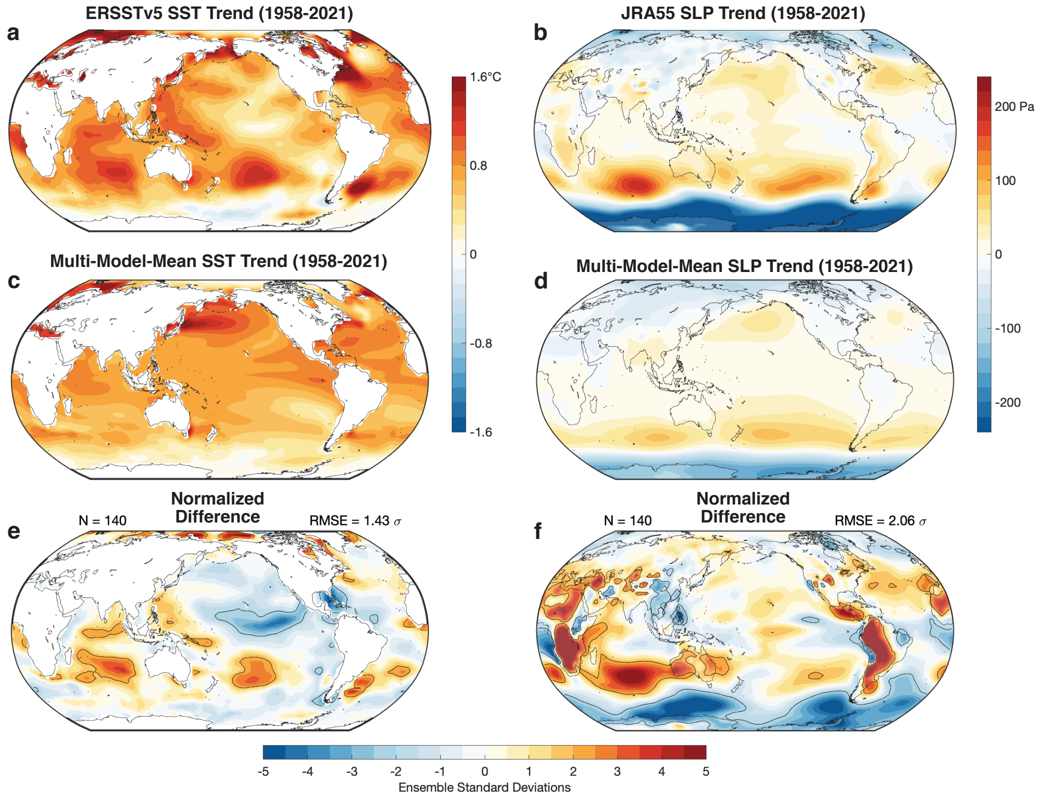

Figure S6. Same as Fig. 1, but for longer-term trends (1958-2021) and using JRA-55 for SLP instead of ERA5.

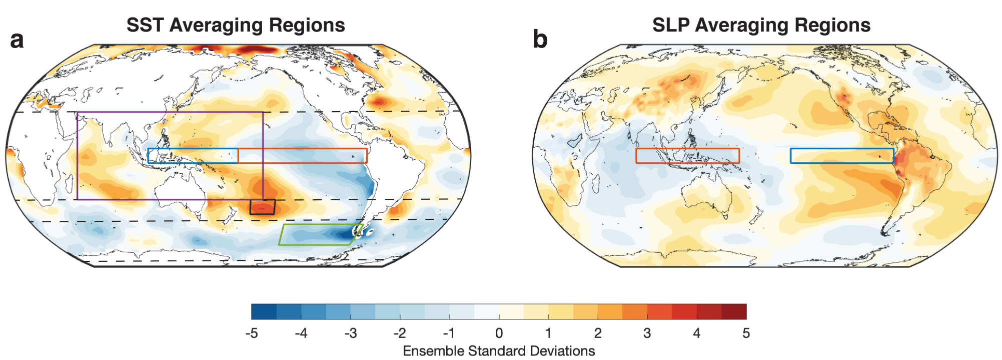

Figure S7. Same as Fig. 1e and f, but with the addition of the averaging regions used in Figs. 2 and S8. Dashed lines show the latitudes of 30°N, 30°S, 45°S, and 75°S.

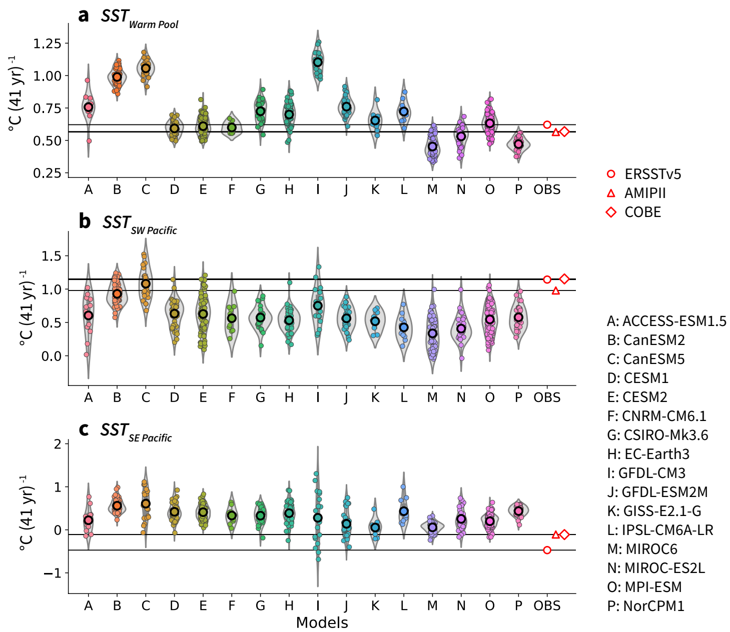

Figure S8. Comparison of observed trends (1979–2020) in key SST indices with those in all ensemble members from 16 LEs: SST in the (a) Warm Pool (30°S-30°N, 50°E-160°W); (b) southwest Pacific (30°S-40°S, 170°W-150°W); and (c) southeast Pacific (47°S-62°S, 140°W-70°W). The southwest and southeast Pacific 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.

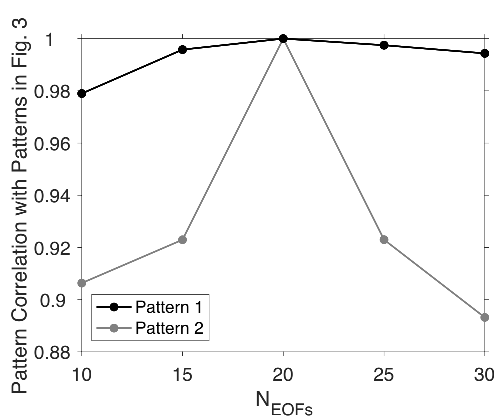

Figure S9. Robustness of the analysis shown in Fig. 3 to the number of EOFs included. For each choice of the number of EOFs (N_{EOFs}), the pattern correlation with the patterns in Fig. 3 are shown for the pattern with the maximum pattern correlation. Only the absolute value of the pattern correlation is considered.

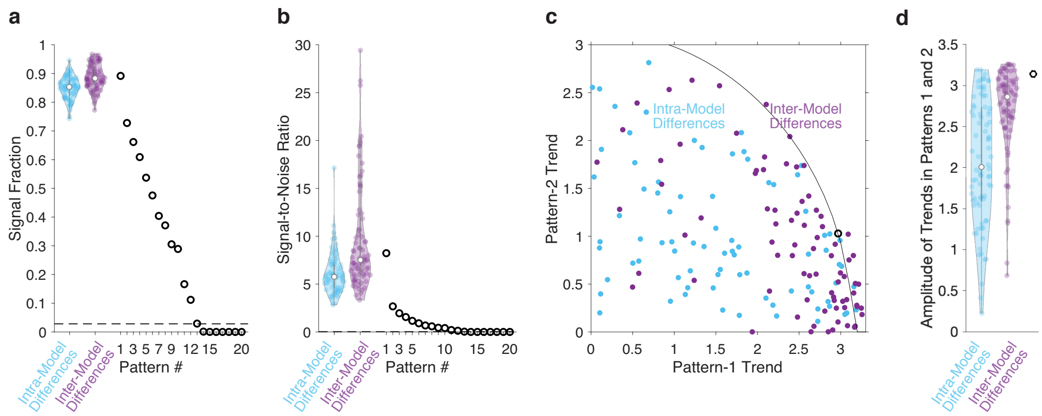

Figure S10. Eigenvalue spectrum of the signal-to-noise maximizing pattern analysis shown in Fig. 3, shown both in terms of (a) signal fraction s and (b) signal-to-noise ratio SNR $= s/(1-s)$ (black circles). Values that have a 5% chance of occurring due to random sampling of differences within the multi-model 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.