Persistent Discrepancies between Observed and Modeled Trends in the Tropical Pacific Ocean

RICHARD SEAGER,[a] NAOMI HENDERSON,[a] AND MARK CANE[a]

[a] *Lamont Doherty Earth Observatory, Columbia University, Palisades, New York*

(Manuscript received 18 August 2021, in final form 2 March 2022)

ABSTRACT: The trends over recent decades in tropical Pacific sea surface and upper ocean temperature are examined in observations-based products, an ocean reanalysis and the latest models from the Coupled Model Intercomparison Project phase six and the Multimodel Large Ensembles Archive. Comparison is made using three metrics of sea surface temperature (SST) trend—the east–west and north–south SST gradients and a pattern correlation for the equatorial region—as well as change in thermocline depth. It is shown that the latest generation of models persist in not reproducing the observations-based SST trends as a response to radiative forcing and that the latter are at the far edge or beyond the range of modeled internal variability. The observed combination of thermocline shoaling and lack of warming in the equatorial cold tongue upwelling region is similarly at the extreme limit of modeled behavior. The persistence over the last century and a half of the observed trend toward an enhanced east–west SST gradient and, in four of five observed gridded datasets, to an enhanced equatorial north–south SST gradient, is also at the limit of model behavior. It is concluded that it is extremely unlikely that the observed trends are consistent with modeled internal variability. Instead, the results support the argument that the observed trends are a response to radiative forcing in which an enhanced east–west SST gradient and thermocline shoaling are key and that the latest generation of climate models continue to be unable to simulate this aspect of climate change.

KEYWORDS: Pacific Ocean; Tropics; Atmosphere-ocean interaction; Sea surface temperature; Thermocline; Climate change; Trends

1. Introduction

The tropical Pacific Ocean plays an outsize role in the global climate system. There is a strong east–west gradient in equatorial sea surface temperature (SST) between the warm pool in the west and the oceanic upwelling-driven cold tongue in the east. Deep convection is focused over the warm pool, the intertropical convergence zone (ITCZ) that extends from the warm pool to the Americas just north of the equator and the South Pacific convergence zone that extends southeastwards from the warm pool. The spatial distribution of convective heating and SSTs powerfully drives the meridional Hadley and zonal Walker circulations in the tropics (Bjerknes 1966; Horel 1982; Brown et al. 2021). Warm pool convection also drives stationary Rossby waves that propagate to the extratropics (Held et al. 2002). As such, the tropical Pacific influences the mean climate worldwide. Further, the eastern tropical Pacific cold tongue has the largest uptake of heat from the atmosphere of any ocean region (Valdivieso et al. 2017) and is the largest source of CO_2 from the ocean to the atmosphere (Takahashi et al. 2002). On interannual to decadal time scales coupled atmosphere–ocean dynamics cause variations in the tropical Pacific Ocean and atmosphere as part of El Niño–Southern Oscillation (ENSO) and the Pacific decadal oscillation (PDO). Zonal variations in the locations of the warm pool and cold tongue and associated atmospheric convection drive atmospheric circulation anomalies that generate climate variability worldwide (Sarachik and Cane 2010). In addition, decadal cold states of the equatorial Pacific, such as that since the 1997/98 El Niño, have been connected to increased ocean heat uptake and reduced global warming (e.g., Meehl et al. 2011; Kosaka and Xie 2013, 2016; Delworth et al. 2015; Deser et al. 2017). Further, during El Niño events, reduced upwelling and a deepened thermocline reduce the partial pressure of CO_2 in the surface ocean and outgassing of CO_2, with the roughly opposite occurring during La Niña events, which combines with the impact of ENSO on terrestrial carbon sinks and sources to form the largest source of interannual variability of atmospheric CO_2 (Feely et al. 1999; McKinley et al. 2004; Zeng et al. 2005; Rödenbeck et al. 2015).

Given this critical role of the equatorial Pacific in the global climate and carbon systems it is important to know how it responds to rising greenhouse gases (GHGs). This has been a matter of uncertainty and debate for decades. Early on, Knutson and Manabe (1995) argued that, because of the nonlinearity of the Clausius–Clapeyron equation, to balance reduced surface longwave cooling with enhanced latent heat loss SSTs should warm in response to rising GHGs more where it is cold than where it is warm. Hence the east–west SST gradient should weaken. In contrast, Clement et al. (1996) and Cane et al. (1997) argued that equatorial upwelling should partially offset warming in the cold tongue and the east–west gradient should strengthen. Following on from Betts (1998), who argued for weakening of the tropical overturning circulation in response

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0648.s1.

Corresponding author: Richard Seager, seager@ldeo.columbia.edu

DOI: 10.1175/JCLI-D-21-0648.1

to rising GHGs, Vecchi and Soden (2007) stated that weakening of the Walker circulation should reduce thermocline tilt, upwelling and the east–west SST gradient. These disagreements sparked a lively debate over the next two decades that has focused on both trends in SST (see Olonscheck et al. 2020, and references therein) and the Walker circulation (see Chung et al. 2019, and references therein) with current opinion saying the observed trends might be within the range of model simulations once internal variability is taken into account.

A different argument is pursued in Seager et al. (2019), who argue that the observed lack of warming, or slight cooling, of the equatorial Pacific cold tongue over the past decades is at or beyond the very cold limit of the ensemble of simulations by multiple models in the Coupled Model Intercomparison Project phase five (CMIP5) and the National Center for Atmospheric Research 40-member Large Ensemble (LENS). In the observations the lack of cold tongue warming is very narrowly confined to the cold tongue and distinct from the more meridionally broad patterns of SST variability familiar from ENSO and the PDO (e.g., Zhang et al. 1997). This equatorial feature is highly suggestive of an association with ocean dynamics and upwelling. The meridional Ekman current is given by $v_{\mathrm{Ek}} = -\tau_x/\rho f H$, where τ_x is zonal wind stress, ρ is seawater density, $f = 2\Omega\cos\phi$ is the Coriolis parameter with Ω the rotation rate of Earth, ϕ latitude, and H the Ekman layer depth. Hence, for a uniform easterly zonal wind stress, there is poleward flow and divergence at the equator but, due to the increase of f and decrease of poleward flow with latitude, there is Ekman convergence immediately north and south, thus restricting upwelling to the immediate equator (Wyrtki 1981). Seager et al. (2019) showed using a simple atmosphere–ocean model they could reproduce the observed trend in response to GHG forcing. Upwelling was key: increased radiative forcing causes more warming in the west Pacific warm pool than in the cold tongue because of the opposing effect of upwelling in the latter. The enhanced east–west SST gradient forces stronger trade winds that cause the thermocline to shoal leading to further suppression of warming in the cold tongue upwelling region.

The model of Seager et al. (2019) imposes aspects of the observed climatology including the SST. The same model could also reproduce the CMIP5 multimodel mean SST trend of enhanced cold tongue warming when the CMIP5 climatology was imposed. In this case the CMIP5 bias toward a cold tongue that has too high relative humidity and too low wind speed makes the SST in the region highly sensitive to forcing leading to strong warming that overwhelms the effect of upwelling. Consequently, Seager et al. (2019) argued that the observed trends of trade wind strengthening, thermocline shoaling and lack of cold tongue warming is a response to rising GHGs while CMIP5 models produce an opposite trend due to their common biases in simulation of the mean tropical Pacific climate—an overdeveloped cold tongue and a tendency to double ITCZs straddling the equator (Li and Xie 2014). Two recent papers, Watanabe et al. (2020) and Olonscheck et al. (2020), which use the CMIP5 and CMIP6 and multimodel LENS (Deser et al. 2020) simulations, have been interpreted to be contrary to the conclusion of Seager et al. (2019) by showing that the observed trends in east–west SST gradient are within the range of climate model simulations once internal variability is more fully sampled by the models. It has also been suggested on the basis of model experiments that the initial response of the system might be to strengthen the east–west gradient but, once the thermocline waters have warmed, the gradient will weaken in the future (Heede et al. 2021; Heede and Federov 2021). Even if so, this does not help explain the model–observations discrepancy in the historical record.

Here we revisit this critical issue focusing on trends to date using updated observations-based products and CMIP6 models and multimodel LENS by (i) examining multiple metrics of equatorial Pacific SST trends, (ii) trends over multiple time periods, and also (iii) related observed and modeled trends in the upper tropical Pacific Ocean temperature. Although a handful of runs out of a total 511 runs do come close, we conclude that the observed trends in SST and thermocline depth cannot easily be reconciled with internal variability of the tropical Pacific atmosphere–ocean system and, instead, are likely a forced response to rising GHGs that state-of-the-art climate models misrepresent.

2. Data, models, and methods

For observed SST we use four SST analysis products: the 1° × 1° Hadley Centre data HadISST (Rayner et al. 2003), the 1° × 1° Japanese Meteorological Agency COBE data (Ishii et al. 2005) and 1° × 1° COBE2 data (Hirahara et al. 2014), and the 2° × 2° National Oceanic and Atmospheric Administration ERSSTv5 data (Huang et al. 2015). COBE only uses in situ data and COBE2 and ERSSTv5 use satellite data to inform on the patterns of SST variations that are used to reconstruct SSTs in data-sparse regions, while HadISST analyses both in situ and satellite data. HadISST, ERSSTv5, and COBE2 are used from 1870 to 2018 and COBE from 1890 to 2018. We also use for SST the European Centre for Medium-Range Weather Forecasts (ECMWF) Ocean Reanalysis 5 (ORAs5) covering the period from 1958 to 2018 at 0.25° × 0.25° resolution (Zuo et al. 2019). ORAs5 assimilates subsurface ocean data and does a correction on SST using HadISST data. The apparent absence of warming in the eastern equatorial Pacific is tightly confined to the equatorial upwelling region (Seager et al. 2019) and for analysis requires observations-based datasets with at least 2° latitude resolution and coverage in the upwelling region. Equatorial upwelling is an essential feature of the coupled ocean–atmosphere dynamics of the Pacific region, a critical element of the Bjerknes feedback, and we are interested in whether the models match observations in the upwelling region. Whatever data products are used, the underlying data are sparse in time and space in the presatellite era, and all data products, including raw data products such as ICOADS (Freeman et al. 2017), must make some assumptions to fill gaps in order to calculate trends. The products we use are in the lineage of Gandin's seminal work on optimal interpolation (OI), and in theory should produce the best representation of trends over data-sparse areas, but

TABLE 1. LENS models used in this study, modeling center, model version, model resolution, full length of simulation, number of ensemble members, and reference.

Modeling center	Model version	Resolution atmosphere/ocean	Years	Members	Reference
CCCma	CanESM2	2.8° × 2.8°/1.4° × 0.9°	1950–2100	50	Kirchmeier-Young et al. (2017)
CSIRO	Mk3.6	1.9° × 1.9°/1.9° × 1.0°	1850–2100	30	Jeffrey et al. (2013)
GFDL	ESM2M	2.0° × 2.5°/1.0° × 0.9°	1950–2100	30	Rodgers et al. (2015)
GFDL	CM3	2.0° × 2.5°/1.0° × 0.9°	1920–2100	20	Sun et al. (2018)
MPI	MPI-ESM-LR	1.9° × 1.9°/nominal 1.5°	1850–2100	100	Maher et al. (2019)
NCAR	CESM1-CAM5	1.3° × 0.9°/nominal 1.0°	1920–2100	40	Kay et al. (2015)

uncertainties surely remain. The collection of observations-based gridded SST datasets we do use represents a variety of different sources of data and methodologies to develop the product and, hence, a range of spatial patterns and magnitudes of equatorial Pacific SST trends to compare models against. ORAs5 is also used for upper ocean temperatures.

For models we use the latest simulations from CMIP6 up to 2018 using the historical simulations extended with the Shared Socioeconomic Pathway 8.5 (SSP-585) (Eyring et al. 2016). We analyze all runs from all models with the necessary SST and upper ocean data (Table 1 in the online supplementary material provides information and citations for the CMIP6 models). In addition, we use simulations from six LENS (Table 1). Collectively this amounts to 511 simulations from 45 CMIP6 plus 6 LENS models. For CMIP6, most models have too few ensemble members for the ensemble mean to effectively isolate the forced response common to all. Instead we average across ensemble members for each model and then average the ensemble means of all models to isolate a CMIP6 multimodel mean forced response. For the six LENS we average across ensemble members to isolate a forced response for each model.

The prime period covered for analysis of trends is 1958–2018 with the start date dictated by the beginning of the ORAs5 ocean reanalysis. However, for SST alone, we look at all trends ending in 2018 and starting at one year increments from 1870 (HadISST, COBE2, ERSSTv5) and 1890 (COBE) up to 1958. Trends are computed by linear least squares regression using annual mean data. Statistical significance is assessed using a two-sided t test.

Prior work has focused on trends in the east–west equatorial Pacific SST gradients (e.g., Coats and Karnauskas 2017) or the SST in the Niño-3.4 region (e.g., Seager et al. 2019). Both metrics are influenced by ENSO and PDO. However, evaluated over periods longer than that of decadal variability, the region of lack of warming is far more confined to the equatorial upwelling region than is typical of ENSO or PDO. To reflect this difference here we use three metrics to measure and compare the observed and modeled SST trends: the east–west gradient [the trend in the SST in a western Pacific box (140°–170°E, 3°S–3°N) minus an east Pacific box (170°–90°W, 3°S–3°N)], a north–south gradient (the trend in the difference in SST between the average of off-equatorial boxes for latitudes 9°–3°S and 3°–9°N minus an equatorial box for latitude 3°S–3°N, all for longitudes 170°–120°W), and finally, the pattern correlation for the entire equatorial Pacific between 10°S and 10°N. The first metric measures the differential warming between the warm pool region of the west where the thermocline is deep and there is only weak or no upwelling and the cold tongue region where the thermocline is shallow and there is strong equatorial upwelling. It is positive if the SST gradient strengthens. The second metric measures the differential warming in the east between the equatorial upwelling region and the waters immediately to the north and south. It is positive for relative cooling at the equator and negative if there is enhanced equatorial warming (as in an El Niño–like response). The third metric measures the similarity between observed and modeled trends across the entire equatorial Pacific.

We also examine trends in observed and modeled thermocline depth h_{tc}, defined as the depth at which the vertical temperature gradient is a maximum in the upper 500 m of the water column. Thermocline depth is computed for ORAs5 and each individual CMIP6 and LENS run for which data were available. Using annual mean data, the trends in depth are then computed. This definition of thermocline depth is preferred over the depth of a representative isotherm because of the direct tie to ocean thermal structure and its physical relevance to SST variations and its lack of sensitivity to the overall ocean warming (see Yang and Wang 2009).

To assess the causes of thermocline depth trends we use the linear shallow water ocean model with two vertical modes as in Seager et al. (2019), described in detail by Israeli et al. (2000) and referred to as the Tropical Climate Ocean Model (TCOM). The model is forced by ECMWF ERA5 wind stresses for 1958–2018 (Hersbach et al. 2020). The model also calculates SST anomalies with the climatological SST and parameterization of the temperature of upwelling water based on ORAs5 as in Seager et al. (2019).

3. Results

a. *Trends in observed and ensemble mean SSTs*

Figure 1 shows maps of trends in tropical Pacific SSTs for 1958–2018 for the five observational gridded datasets and their mean. In the HadISST, COBE, and ORAs5 gridded datasets the trends show widespread warming but a narrow equatorially confined region of cooling in the central to eastern tropical Pacific. In ERSSTv5 and COBE2 the equatorial lack of warming is more confined to the central Pacific and is broader in latitude [see also Santoso et al. (2007) and Coats and Karnauskas (2017) for comparison of spatial patterns of multidecadal tropical Pacific SST across different SST

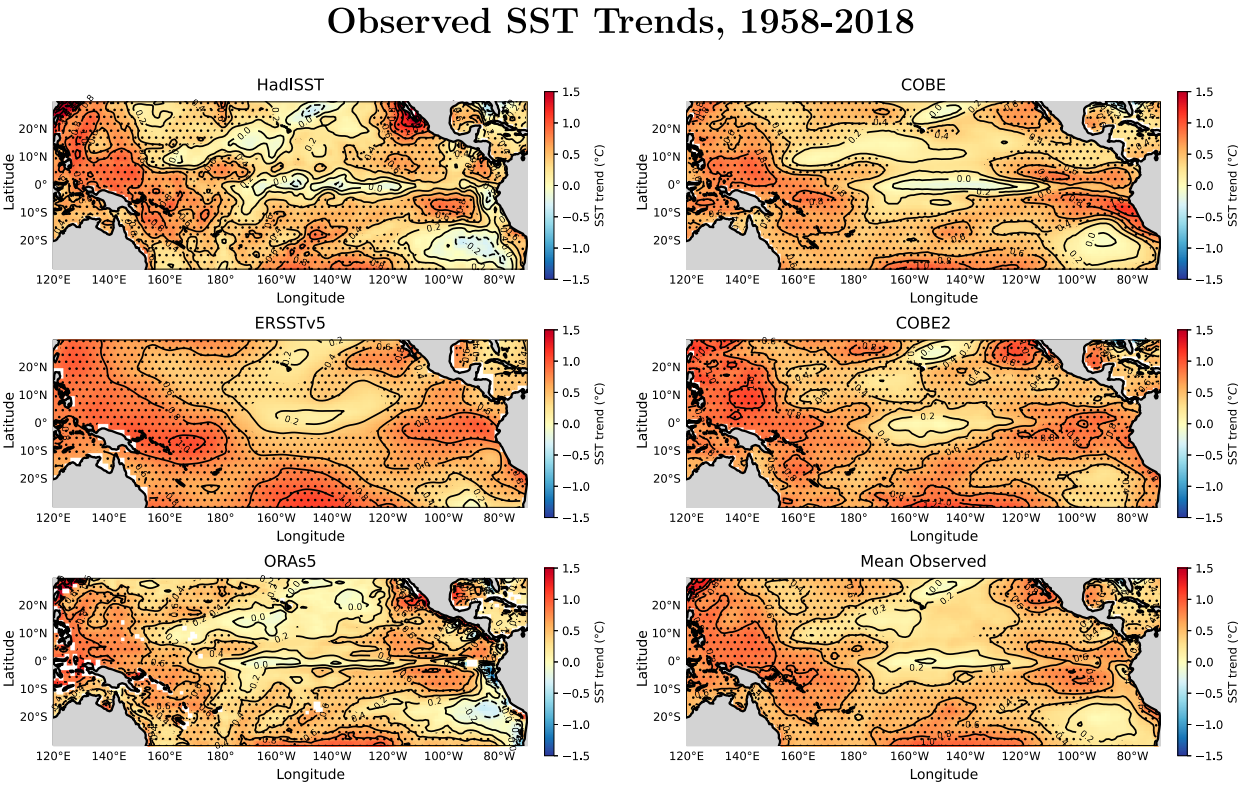

FIG. 1. The SST trend (K per 61 years) from five observations-based products over 1958–2018. Shown are those from (top left) HadISST, (top right) COBE, (middle left) ERSSTv5, (middle right) COBE2, (bottom left) ORAs5, and (bottom right) the mean of these five. Stippling indicates significance at the 95% confidence level according to a two-sided *t* test.

products]. As seen in the observational mean, despite these differences, there is substantial agreement across gridded datasets that the central equatorial Pacific warmed less than the west Pacific warm pool and also less than waters immediately to its north and south. There is no statistically significant trend in the equatorial region of reduced or no warming while elsewhere in the tropical Pacific there is widespread significant warming. It is the very lack of a significant trend that makes the equatorial region of relative cooling important from a climate dynamics and impacts perspective.

Figure 2 shows the modeled forced responses as derived from the CMIP6 multimodel mean, the ensemble mean for each LENS and the mean across the LENS ensemble means. The observed pattern of relative equatorial cooling is not captured as the response to changes in forcing in either the CMIP6 multimodel ensemble mean nor any of the six LENS. The multimodel CMIP6 and LENS ensemble means are rather spatially uniform in the equatorial Pacific. Among the ensemble means of the LENS models, CanESM2 and GFDL-CM3 have El Niño–like responses with enhanced warming in the eastern equatorial Pacific. In contrast, CSIRO-Mk3.0-6-0 and MPI-ESM tend to have a stronger zonal SST gradient. Despite these differences, none of the CMIP6 or LENS ensemble means have an SST trend pattern that is visually similar to the observed ones with either narrow equatorial cooling from the central to eastern equatorial Pacific (HadISST, COBE, ORAs5) or a broad region of relative cooling in the central equatorial Pacific (COBE2, ERSSTv5). Indeed, the LENS models with a stronger gradient connect lesser warming in the eastern equatorial Pacific to a wider region of reduced warming in the southeast subtropical Pacific (He et al. 2017). It is striking that the forced responses of all the models have weak warming in the southeast Pacific Ocean. The observations-based products also show an area of reduced warming in the southeast Pacific but separate it from the equatorial cooling by a zonal band of warming west of Peru. Because these ensemble and multimodel mean trends are evaluated on averages across many simulations in which the internal variability has been greatly reduced all of the trends, even the smallest ones, are significant at the 95% confidence level.

b. *Trends in SST in individual models runs*

These results make clear that the models do not have a response to radiative forcing that resembles the observed trend. Typically, recognizing this, the discrepancy between the observed and modeled trends is explained in terms of the observed trend being strongly influenced by internal variability. To examine this possibility we next consider individually all the trends in all the runs of all the models. Figure 3 shows all three metrics of SST trend, with the east–west gradient

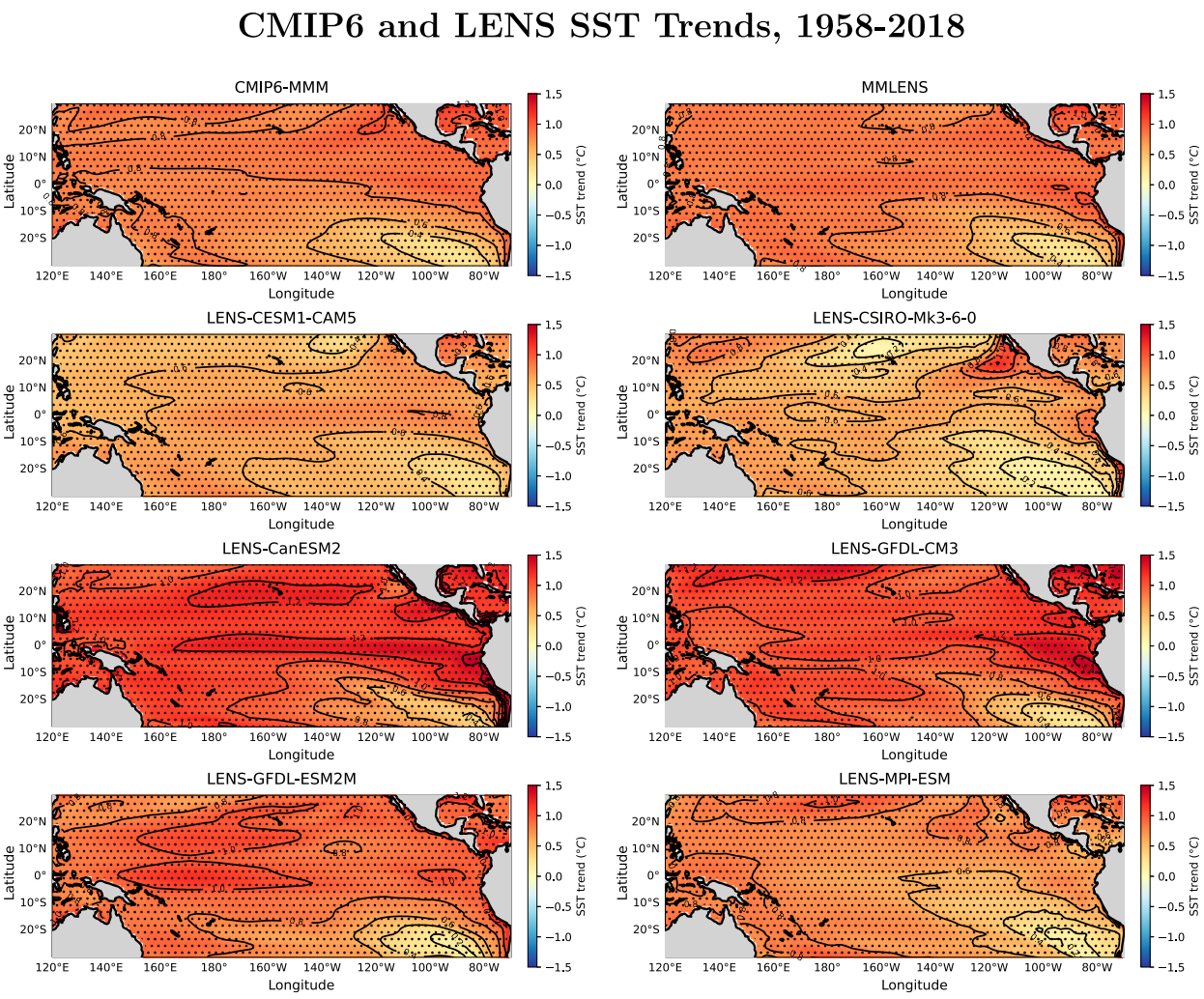

FIG. 2. The SST trend (K per 61 years) from climate models over 1958–2018. Shown are the multimodel mean trends of (top left) CMIP6 and (top right) LENS and, below, the ensemble mean trends of six individual LENS models.

trend on the horizontal axis, the north–south gradient trend on the vertical axis (both in °C), and the pattern correlation with HadISST indicated by the color of the dot. The five observations-based results are also indicated on each panel (the pattern correlation for HadISST is 1 by construction). Results are shown for each LENS model ensemble, the CMIP6 models with the multimodel ensemble mean marked, and the ensemble means of the LENS models.

From the point of view of all three metrics, the observed trends tend to stand apart from the model ensembles. This is definitely the case for all three metrics for the forced responses as measured by the CMIP6 multimodel mean, the individual LENS ensemble means, and the LENS multimodel mean. Of the LENS, the MPI-ESM is the only model that has a forced response with both a positive east–west and north–south temperature gradient as in all the observations-based products. Of the observations-based products ERSSTv5 is the one that models find it easiest to agree with but, even so, no forced response matches even its muted SST gradient changes.

For the individual model runs which contain full modeled internal variability it is also the case that none can match the amplitude of the gradient changes seen in the HadISST and COBE gridded datasets. Further, only a handful of 511 CMIP6 and LENS individual runs can match the trends seen in the other three observations-based gridded datasets. Once more, the MPI-ESM stands out as the model whose individual runs can best match the observed gradient changes. In general, while some models can reproduce the trend in the east–west gradient, they are less likely to reproduce the trend in the north–south gradient (i.e., the equatorially confined cooling) and even MPI-ESM does not match this as seen in the HadISST and COBE gridded dataset. Notably, no individual runs have a pattern correlation with the HadISST trend above 0.63, while the other observational gridded datasets have pattern correlations with HadISST ranging from 0.65 (ERSSTv5) to 0.87 (COBE).

We rank, relative to the HadISST data, the other four observations-based gridded datasets and the individual

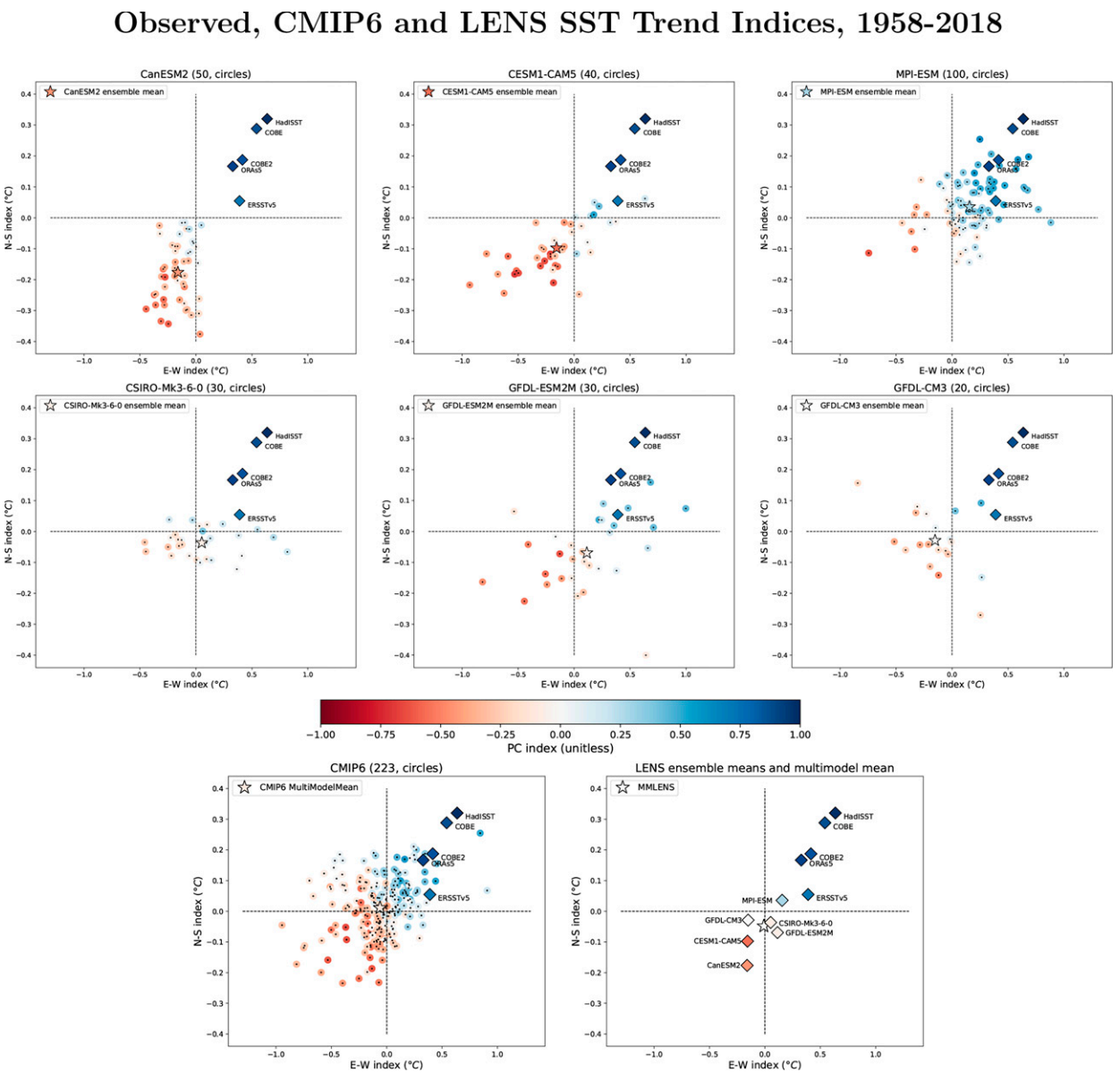

FIG. 3. Trends (K per 61 years) in the east–west SST gradient (horizontal axis), north–south near-equatorial SST gradient (vertical axis), and the pattern correlation between the HadISST observations-based and modeled SST gradient in the Pacific Ocean 10°S–10°N domain (color coding of dots), for 1958–2018. Results are shown for (top),(middle) the six LENS ensembles and (bottom left) for individual CMIP6 runs and the multimodel ensemble mean and (bottom right) for the six LENS ensemble means plus the multimodel LENS ensemble mean. Observations-based estimates for HadISST, COBE, COBE2, ERSSTv5, and ORAs5 are also shown in each panel.

models runs by each metric: the amplitudes of the pattern correlation, the east–west and north–south gradient trends and then assign a multimetric rank by equally weighting the three ranks. Results are shown in Table 2 for the top and bottom eight model runs. We also show the change in the tropical Pacific mean SST and the area-weighted root-mean-square difference between the SST trends and the HadISST trend in the 10°S–10°N, 140°E–90°W domain. COBE ranks higher than any model run, 3 model runs rank higher than COBE2, 3 more runs higher than ORAs5, and 21 more higher than ERSSTv5. Interestingly, the GFDL-ESM2M model contributes ensemble members to both the best matching and worst matching groups, indicating large amplitude variability. Next we examine in closer detail the six model runs that best match the observations-based products.

Figure 4 shows the 1958–2018 SST trends for these six runs (which come from three models). All clearly have an enhanced east–west SST gradient with muted warming in the cold tongue region. However, unlike the observations-based gridded datasets, the runs from the MPI-ESM and MIROC-ES2L models connect this muted warming in the cold tongue to the muted warming in the southeast subtropical Pacific

TABLE 2. Observations-based, CMIP6, and LENS individual run SST trends for 1958–2018 compared to the HadISST observations-based trend by three metrics showing the eight best matching (top of table) and eight worst matching (bottom of table) runs. Listed are name of model and run identifier, whether from CMIP6 or LENS, the trend in the tropical Pacific mean SST and root-mean-square error between the trend in the run and the HadISST one in the 10°S–10°N, 140°E–90°W domain (both in °C), the three metrics: pattern correlation and east–west and north–south gradient change (°C), and, finally, overall rank. Trends significant at the 95% (90%) confidence level are marked in bold (italics).

			Statistics		Ranked indices			
Obs or model	Ensemble member	Type	Mean (°C)	RMSE (°C)	PC	NS (°C)	EW (°C)	Combined rank
HadISST		OBS	0.44	0.00	1.00	**0.32**	**0.64**	1
COBE		OBS	0.45	0.13	0.87	**0.29**	*0.54*	2
MPI-ESM	r008i2005p3	CMIP5-LENS	0.59	0.28	0.62	0.20	**0.68**	2
MIROC-ES2L	r6i1p1f2	CMIP6	0.67	0.39	0.50	0.25	*0.84*	4
MPI-ESM	r073i2005p3	CMIP5-LENS	0.52	0.25	0.62	0.19	*0.59*	5
COBE2		OBS	0.58	0.22	0.77	0.19	0.42	6
MPI-ESM	r090i2005p3	CMIP5-LENS	0.54	0.28	0.51	0.16	0.57	7
GFDL-ESM2M	r19i1p1	CMIP5-LENS	1.02	0.64	0.46	0.16	0.68	8
MPI-ESM	r033i2005p3	CMIP5-LENS	0.35	0.27	0.50	0.17	0.42	9
ORAs5		OBS	0.42	0.15	0.82	0.17	0.33	10
MPI-ESM	r059i2005p3	CMIP5-LENS	0.55	0.24	0.63	0.25	0.25	11
MPI-ESM	r075i2005p3	CMIP5-LENS	0.33	0.31	0.45	0.21	0.35	12
ERSSTv5		OBS	0.61	0.28	0.65	0.05	0.39	34
CESM1-CAM5	r19i1p1	CMIP5-LENS	0.70	0.49	−0.56	−0.17	*−0.52*	509
GFDL-ESM2M	r17i1p1	CMIP5-LENS	1.15	0.82	−0.54	−0.23	−0.44	509
CESM1-CAM5	r33i1p1	CMIP5-LENS	0.71	0.51	−0.43	−0.24	**−0.63**	511
CanESM2	r25i1p1	CMIP5-LENS	1.35	1.00	−0.55	−0.25	*−0.37*	512
CESM1-CAM5	r12i1p1	CMIP5-LENS	0.70	0.57	−0.47	−0.22	**−0.93**	513
CESM1-CAM5	r27i1p1	CMIP5-LENS	0.92	0.67	−0.63	−0.18	**−0.50**	514
CanESM2	r21i1p1	CMIP5-LENS	1.28	0.93	−0.55	−0.30	*−0.45*	515
CESM1-CAM5	r35i1p1	CMIP5-LENS	0.74	0.52	−0.69	−0.18	**−0.53**	516

Ocean. The run from the GFDL-ESM2M model, in contrast, appears quite realistic in the sense that it has an equatorially confined region of muted warming in the central to eastern Pacific. This model, however, provides runs that were also among the lowest ranked in terms of matching the observed trends, so the good match clearly arises from model internal variability as opposed to a forced response.

How unusual are the observed trends when viewed from the perspective of the CMIP6 and LENS? Figure 5 shows histograms for all 511 model runs of the east–west and north–south gradient changes, and pattern correlation with HadISST, for 1958–2018 as well as the five observational values. For the east–west gradient change 15 (3%) (relative to HadISST) and 19 (4%) (relative to COBE) runs equal or exceed the observed values. For the north–south gradient change no runs equal or exceed the observed values in these two observations-based gridded datasets. In the terminology of the Intergovernmental Panel on Climate Change uncertainty language (Mastrandrea et al. 2011), this means that the east–west gradient change as observed by HadISST is *extremely likely*, and the north–south change to be *virtually certain*, to be outside the range of what CMIP6 and LENS models can produce for this period as a combination of internal variability plus forced change. For the pattern correlation coefficient with HadISST no LENS or CMIP6 models runs exceed 0.66. Of the observations-based gridded datasets, ERSSTv5 is the one that model runs come closest to being in agreement with. The change in north–south SST gradient in ERSSTv5 is well within the model ensemble but this dataset clearly stands out as an outlier in this metric within the observations-based gridded datasets. However, even for ERSSTv5 only a handful more model runs (23 or <5%) exceed its east–west SST gradient change. The distributions of the multimodel ensemble for the east–west and north–south gradients and the pattern correlation are all skewed negative, with a mean close to zero or negative, quite distinct from the observed positive values.

c. *Trends in equatorial Pacific thermocline depth and upper ocean temperature*

Seager et al. (2019) argued that the observed lack of cold tongue warming was due to thermocline shallowing in response to stronger trade winds that themselves were driven by the enhanced east–west SST gradient. Examining the observed and modeled thermocline depth trends provides another check on the realism of the modeled tropical Pacific climate trends. Figure 6 shows the trends in thermocline depth and SST for the Niño-3.4 region for the CMIP6 and LENS ensembles. There is a general tendency in the models toward SST warming and thermocline shoaling. This is consistent with GHG-driven warming being surface trapped and not penetrating to depth (Fig. 7). As such the depth of the

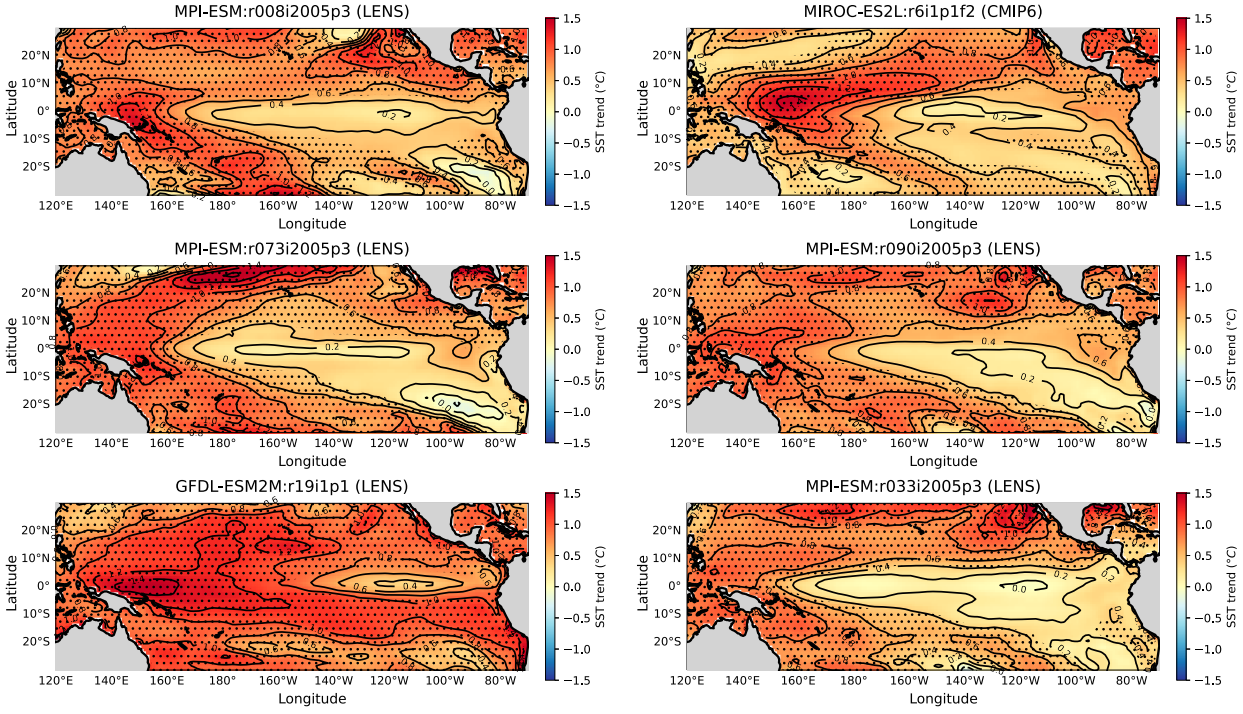

FIG. 4. The SST trends (K per 61 years) in the six runs from the CMIP6 and LENS ensembles that best match the observations-based HadISST trend over 1958–2018, according to an equal weighting of the east–west gradient, north–south gradient, and pattern correlation metrics.

maximum vertical temperature gradient—the thermocline—tends to move upward. This is a thermodynamic (diabatic) change to the thermocline depth. Once more, the observations-based results from ORAs5 are distinct from the models with strong thermocline shoaling going along with only weak warming of the SSTs, suggesting a dynamical (adiabatic) shoaling. The plots also show the thermocline and SST changes computed with the linear shallow water ocean model of Seager et al. (2019) forced by observed winds. In this model the thermocline depth is only influenced by winds since the thermal structure is held fixed. The model thermocline depth trend reasonably well matches the observed one, confirming the dynamical origins of the observed trend. The SST in the ocean model (which, while depending in part on upwelling and thermocline depth, does not influence the specified vertical thermal structure) also matches the observed SST trend.

The vertical temperature profile trends, averaged over 2°S–2°N, are shown in Fig. 7 and emphasize just how different the observed trends are from the models. ORAs5 has strong (and statistically significant in a broad region) cooling in the upper ocean driven by thermocline shoaling. While the models do have some dynamical lifting of the thermocline as indicated by the subsurface cooling, which is widely statistically significant for the ensemble and multimodel means and has been explained in terms of changes to wind stress curl (DiNezio et al. 2009), the greater dominance of dynamical driving in the observations-based ocean reanalysis is clear. The difference between ORAs5 and models is most stark for the CMIP6 and LENS multimodel means. We also show in Fig. 7 the trend for five of the six best-matching runs (GFDL does not make ocean data available so only the MPI-ESM and MIROC results are shown). They have upper ocean temperature trends more akin to ORAs5 but the dynamically induced thermocline shoaling and subsurface cooling is weak compared to the observations-based ocean reanalysis.

d. Trends in the east–west and north–south SST gradients over the last century and a half

In Fig. 8 we show how the east–west SST gradient trend has evolved over time in observations-based products, the top six model runs and the complete CMIP6 plus LENS multimodel ensemble. Trends over all periods ending in 2018 are shown beginning in 1870 (HadISST, COBE2, ERSSTv5) and 1890 (COBE) with the most recent one being 1958–2018. Model trends are shown for 1870 on except for GFDL-ESM2M which begins in 1950. The trend toward an increasing east–west gradient in HadISST has been positive and between 0.3° and 0.6°C over the entire observational record. At the other end of the range of observations-based products, the trend in ERSSTv5 has also been consistently positive but varying between 0.0°C (for 1870–2018) and 0.5°C. The variations over time in the observations-based products are consistent with amplification or partial cancellation of a steady trend by internal PDO variability (Zhang et al. 1997). When

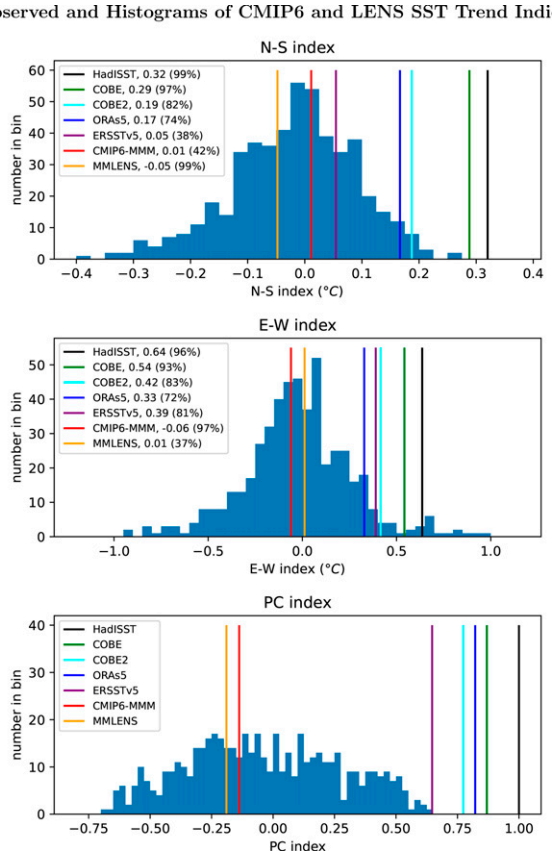

FIG. 5. Histograms of CMIP6 plus LENS SST trend (1958–2018) indices with observations-based values and CMIP6 and LENS multimodel ensemble means marked. (top) The north–south SST gradient index; (middle) the east–west SST gradient index; (bottom) the pattern correlation with HadISST, all for trends (K per 61 years) over 1958–2018. The CMIP6 and LENS multimodel means are marked as red and orange vertical bars and the HadISST, COBE, COBE2, ORAs5, and ERSSTv5 observations-based estimates as black, green, light blue, dark blue, and purple vertical bars. In the legend the numbers and percents are the trend values and significance levels against the null hypothesis that the trend is zero.

the trend begins during a cold tropical Pacific state of the PDO the trend from then to 2018 is least positive (e.g., the 1940s and 1950s) and when the trend begins in a warm PDO state the trend from then to 2018 is more positive (e.g., the 1920s and 1930s). This steadiness with small variations is not at all the case for the six model runs. Although they match the observations-based products trends for the most recent 60 year period (which is dictated by the selection procedure), their trends have either been generally weaker, or even toward a reduced east–west gradient in the MIROC and GFDL models, when evaluated from earlier periods. Consequently even these best matching model runs do not well match the observed record of the east–west SST gradient when viewed over the whole period of observations.

The discrepancy between models and three of the four observations-based products with long records is more pronounced for the north–south gradient changes shown in Fig. 9. These observed changes are beyond the 95th percentile of the CMIP6 plus LENS ensemble spread for all trend periods and, very often, outside the entire ensemble spread. While not as strong as the observations-based products, of the best-matching six runs from all models, one run of the MPI-ESM model is typically outside the 95th percentile of the model ensemble spread of the north–south gradient. The other MPI-ESM runs also have persistent positive trends in the north–south gradient though much weaker than in the observations-based products. The ERSSTv5 dataset is an outlier among the observations-based gridded products for the north–south gradient change and the models are consistent with it.

4. Conclusions

We have examined tropical Pacific climate change in the latest generations of CMIP climate models and the multimodel large ensembles. As for prior generations of models, these models still fail to reproduce the observed history of SST in the equatorial Pacific Ocean. This is a matter of concern since SST changes in this region exert a strong influence on climate worldwide, global ocean heat uptake and air–sea exchange of CO_2. The forced response of the CMIP6 models, as represented by the multimodel mean, and the forced response of the individual model large ensembles, as represented by the ensemble mean of each, are emphatically different from the trend in four observations-based gridded datasets and still different from the one dataset closest to the models (ERSSTv5). The difference between trends on observations-based products and models cannot easily be explained away by appealing to the observations-based trend being strongly influenced by internal climate variability. By using three metrics of the pattern of SST trend, only a handful of the 511 individual model runs come close to matching the observations-based SST trend pattern and only a few of these have the observed pattern of meridionally narrow equatorial cooling surrounded by warming. Instead, some of the best matching model runs erroneously connect the lack of cold tongue warming with a wider area of muted warming in the southeast Pacific. Furthermore, even those individual runs that best match the observations-based trend for the 1958–2018 period do not simulate how the trend has evolved over the period of instrumental records. In the observations-based products, trends to 2018 beginning in the nineteenth century and continuing up to the most recent (1958–2018), consistently show a strengthening east–west SST gradient. While the best matching models agree with the observations-based products by this single metric for the most recent period they do not have a gradient that strengthens as consistently and strongly as observed from the nineteenth century until now. The north–south gradient also is consistently strengthening in all trends to 2018 starting from 1870 to 1958 in four gridded datasets. Only one model run (with MPI-ESM) does this, though more weakly than observed. However, the ERSSTv5 dataset is a clear outlier for this metric and models easily agree with it.

The cooling or lack of warming of the cold tongue in observations-based gridded products has been linked to

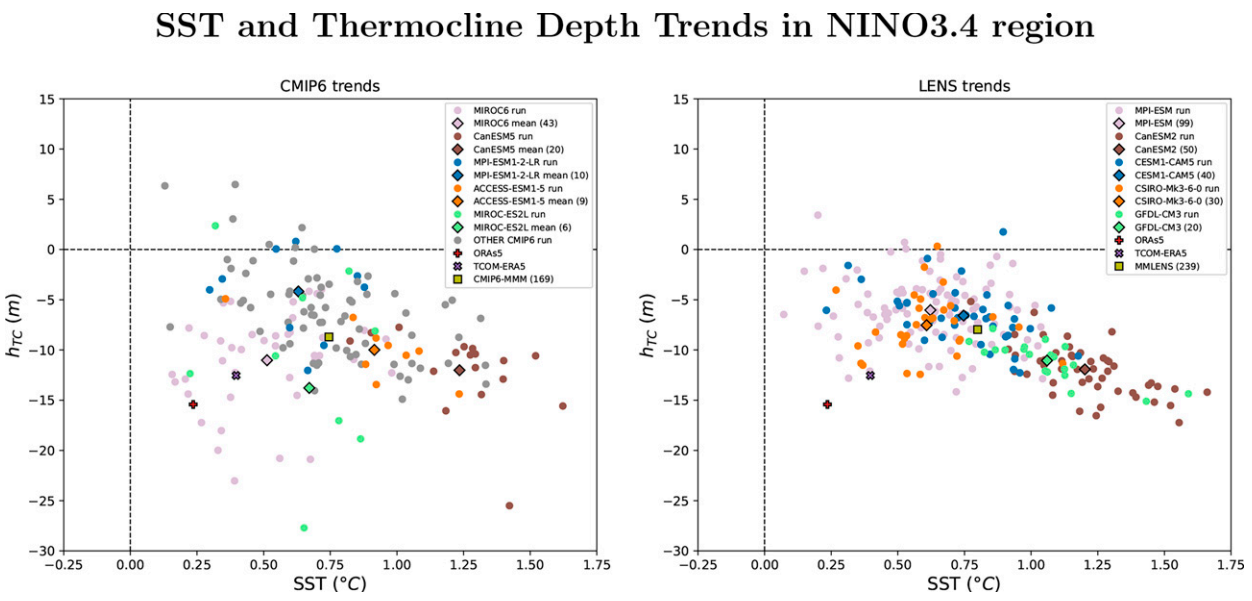

FIG. 6. Trends (K per 61 years) over 1958–2018 in SST (horizontal axis) and thermocline depth (vertical axis) in the Niño-3.4 region of the central to eastern equatorial Pacific Ocean for (left) CMIP6 and (right) LENS with individual runs shown as dots and ensemble mean as diamonds for LENS and CMIP6 models with more than 6 ensemble members, and the CMIP6 and LENS multimodel means as squares. The ORAs5 trends are shown as a plus sign and the trends simulated by a linear shallow water model (TCOM) forced by ERA5 observations-based wind stresses are shown as a cross. The right panels show trends in the vertical temperature profile at the equator for (top) ORAs5 and the multimodel mean of (middle) CMIP6 and (bottom) LENS.

thermocline shoaling driven by enhanced trade winds. Consistent with the disagreement on SST trends, none of the CMIP6 (with the exception of the very scattered MIROC6 model)[1] or LENS models simulate the observed combination of strong thermocline shoaling and muted cold tongue SST change. Instead the models shoal the thermocline primarily for thermodynamic reasons in response to surface warming. In contrast the observed shoaling can be explained adiabatically in terms of dynamical lifting in response to wind stress trends.

These results are not inconsistent with those of Olonscheck et al. (2020) and Watanabe et al. (2020). The latter worked with a 1951–2010 period, comparable to that here in its ability to average over the dominant time scales of internal variations. However, to estimate the trends in zonal SST gradient, ΔSST_{eq}, they use a latitude range of 5°S–5°N whereas we use the narrower 3°S–3°N to emphasize the equatorial upwelling region that is so clear in the observed trend pattern and of significant climatic and biogeochemical importance. Their results for CMIP5 models and six observational products are summarized in their Fig. 1a. They state, "While CMIP5 models show a large inter-model spread, the average of the observed ΔSST_{eq} trends lies outside the 5%–95% range of the CMIP5-based trends." We note that their average of six observed trends includes two raw data products; all of the four analysis products have stronger trends than any of the 27 CMIP5 models in their sample. The trends in the two raw data products also show a lesser increase in the zonal gradient, while most CMIP5 models have it decreasing. Though we prefer the analyses, which use covariance information and so in principle ought to be superior, even with all six observational products included it remains highly unlikely that the CMIP5 models simulate the observed trend.

Watanabe et al. (2020) then examine four models with large ensembles, concluding that two (IPSL-CM6A-LR and MIROC6) give only marginal agreement, in one (CESM1) 10% of the simulations are close to observations, and one (MPI-ESM1.1) gives agreement 40% of the time. A rigorous statistician might object that, in the absence of an a priori reason to single out the models giving good agreement, the only fair thing is to consider the entire aggregate of simulations, in which case the small number of total successes is not impressive. In this study we consider the latest generation CMIP6 models and also the available LENS and also find that only the MPI-ESM model simulations are close enough to the observations-based products to be deemed potentially realistic. Olonscheck et al. (2020) work with CMIP5, CMIP6, and LENS models and do use the narrower latitude range but focus on much shorter time periods for trend evaluation than we use and which are more influenced by the natural variability of the PDO.

An inescapable limitation of our investigation is the uncertainty in the observations. Leaving aside errors in each datum, data coverage in the tropical Pacific is too sparse to allow full

[1] The MIROC6 model, while simulating in the Niño-3.4 region a combination of thermocline shoaling and SST change similar to ORAs5, also simulates a region of relative SST cooling in the western equatorial Pacific while the thermocline shoaling is also shifted west of that in ORAs5. Hence this model ranks low in the metrics of SST trend patterns (not shown).

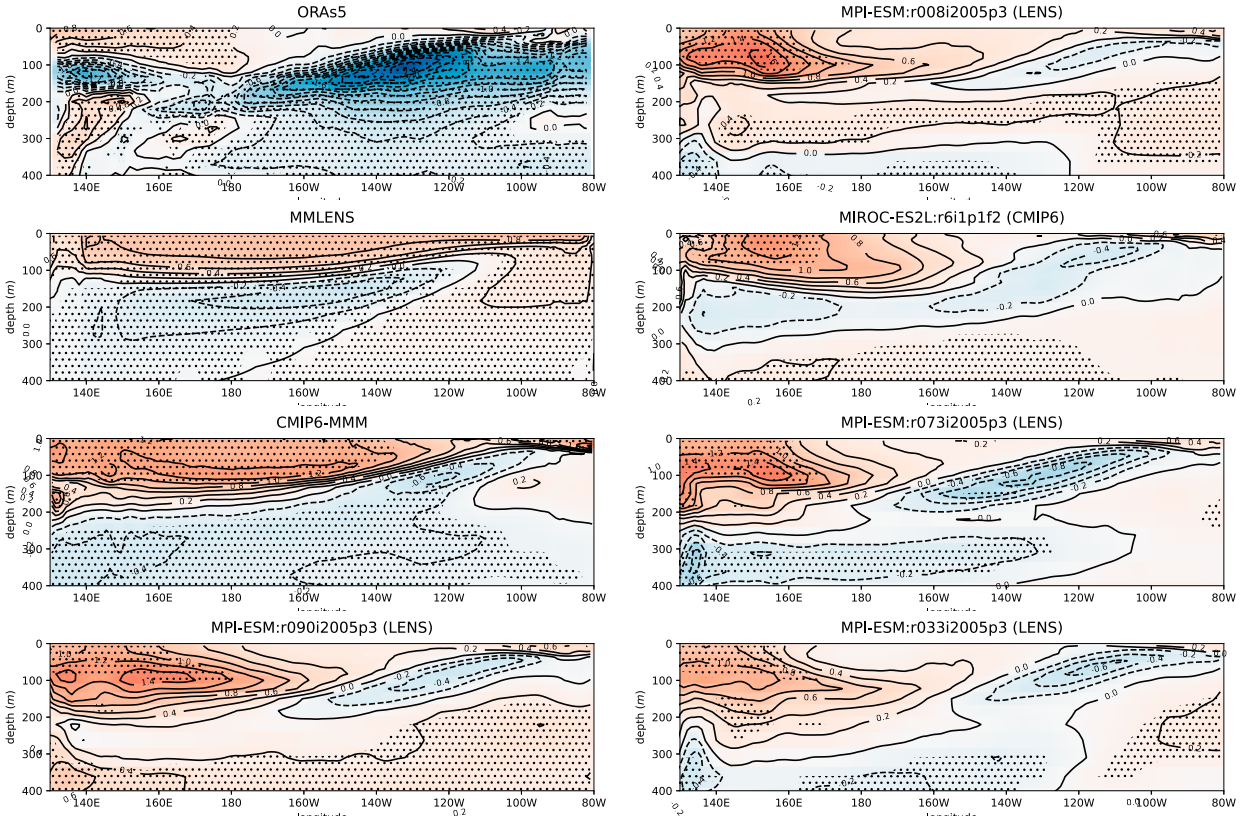

FIG. 7. Trends (K per 61 years) in the equatorial upper ocean temperature for ORAs5, the LENS multimodel mean, and CMIP6 multimodel mean (first three panels in the left column) and four of the six best matching individual runs from the CMIP6 plus LENS ensemble (remaining panels). Trends significant at the 95% level are stippled.

confidence in computed trends. The resolution we require restricts the "observational" datasets we use to a subset of the available reanalysis schemes, which use covariance information to fill data gaps and are not unadulterated observations. In common with most but not all users, we prefer these to more raw data products such as ICOADS because in principle the added information should be helpful. Our experience confirms this; for example, the ENSO forecasts in Chen et al. (2004) are initialized with SST reanalyses and are quite successful in data-poor decades such as the late nineteenth century even when initialized from times with no obvious strong signal. Nonetheless we cannot rule out the possibility that these reanalysis schemes misuse the additional information so that it degrades the raw data, as suggested by Deser et al. (2010) in their study of trends using ICOADS data. We note, however, that not even ICOADS data are made from raw observations but have been processed to try to eliminate biases and changes in observing systems. As monthly gridded data there is also some homogenization in time and space. Moreover, as is true of the reanalysis products, a consequence of incomplete coverage is a relaxation toward climatology, which impacts estimates of trends.

As noted above, due to the limited spatial coverage and resolution of ICOADS data, we cannot make a direct comparison between reanalysis and ICOADS versions of our gradient metrics. The closest fit we have found in the literature is Fig. 1a of Watanabe et al. (2020), which compares six observational products (four reanalyses and two raw data) and 26 CMIP5 model runs. Compared with the present paper, they use latitudinally broader boxes with somewhat different longitudes and their calculated trends are from 1951 to 2010 instead of from 1958 to 2018 here so our period ends with 8 more years marked by a relatively strong east–west gradient. In their figure both raw products (ICOADS and HadSST3; Kennedy et al. 2011a,b) have trends that show a strengthening of the east–west gradient, though neither is as strong as any of the four reanalyses [HadISST, Kaplan (Kaplan et al. 1998), COBE2, ERSSTv6]. While no CMIP5 model runs analyzed have a trend as large as those four, there are four model runs with stronger trends than ICOADS (and six stronger than HadSST3). These findings strongly suggest that those who believe ICOADS or HADSST3 to be more reliable than the reanalyses are less likely to be persuaded by our results. For the reasons stated, we are not, however, of that persuasion.

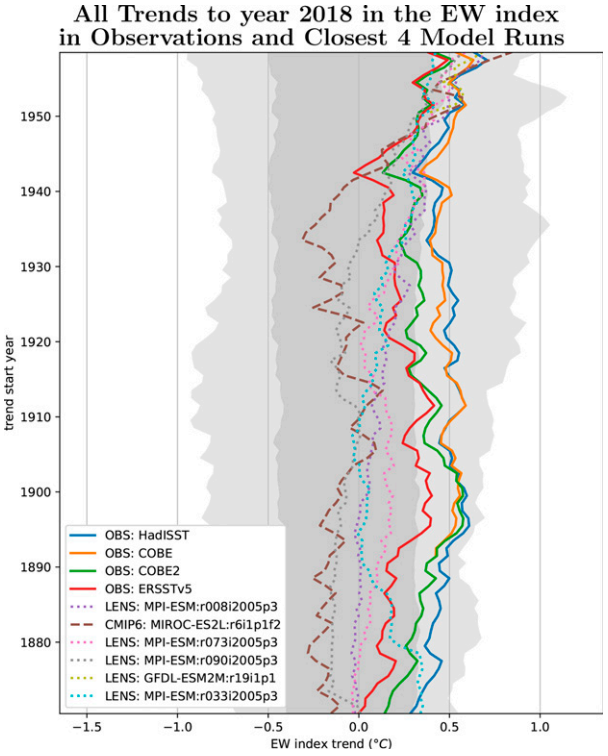

FIG. 8. Trends to 2018 from start year (K per length of trend period) shown on the vertical axis in the east–west SST gradient observations-based products (blue: HadISST; orange: COBE; green: COBE2; red: ERSSTv5) and the six models runs that best match the HadISST observations-based product with the 95th percentile (dark shading) and full range (light shading) of the entire CMIP6 plus LENS ensembles.

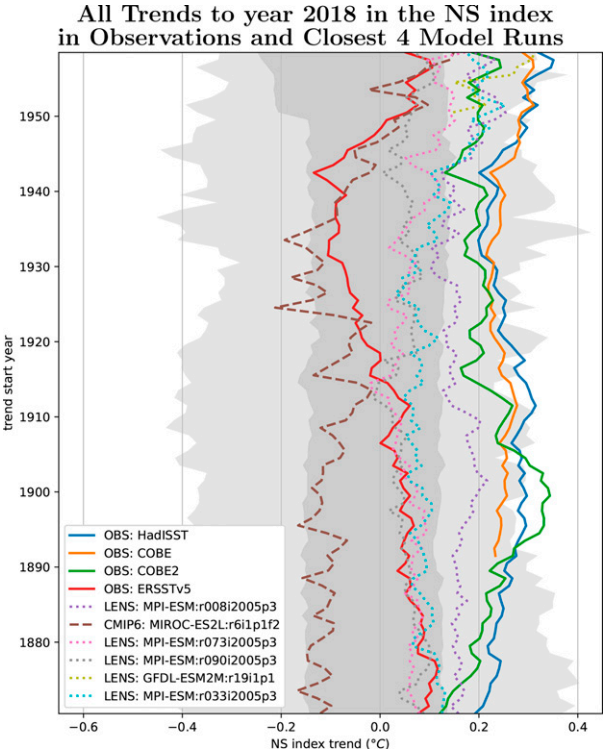

FIG. 9. As in Fig. 8, but for the north–south SST gradient index.

If we do put faith in the SST reanalyses, this collection of evidence from observations-based products and models supports the conclusions that (i) during the time of rising GHGs, the observations-based products show a persistent increase of the east–west SST gradient that has been associated with a wind-driven shoaling of the thermocline that enhances cooling due to upwelling in the cold tongue; (ii) models have forced responses to rising GHGs that are quite different from the observed trend; (iii) taken together, the observed SST and thermocline trends are extremely likely (in IPCC usage) to be outside the range of internal variability in the models. Only one model (MPI-ESM) comes close in some of its runs to the observed trends, both recent ones and those evaluated over longer time periods, though its thermocline shoaling is weaker than observed. It is still possible that the observed tropical Pacific trends are a result of internal variability akin to that in MPI-ESM (and also possibly GFDL-ESM2M, but in the absence of ocean model data it is hard to tell). The character of the observed trends could be artifacts of changes in observation density and observing practice or the analysis methods. Also, perhaps the ERSSTv5, though an outlier among observations-based gridded datasets, is closer to reality than the other four gridded datasets reducing, but not eliminating, the observations–models discrepancies. However, the parsimonious explanation of all the above results is that there is a discrepancy between models and the real world and it arises from models misrepresenting the forced response to rising greenhouse gases, most likely due to model biases in climatological SSTs and winds (e.g., Seager et al. 2019).

Acknowledgments. This work was supported by NSF Awards OCE 1657209, OCE 1740648, and AGS 1934363. CMIP6 data were retrieved from the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/cmip6/) and the LENS data are from the Mutimodel Large Ensembles Archive (https://www.cesm.ucar.edu/projects/community-projects/MMLEA/). We thank two anonymous reviewers and the editor for constructive and insightful reviews.

REFERENCES

Betts, A. K., 1998: Climate-convection feedbacks: Some further issues. *Climatic Change*, **39**, 35–38, https://doi.org/10.1023/A:1005323805826.

Bjerknes, J., 1966: A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, **18**, 820–829, https://doi.org/10.3402/tellusa.v18i4.9712.

Brown, J., and Coauthors, 2021: South Pacific convergence zone dynamics, variability and impacts in a changing climate. *Nat. Rev. Earth Environ.*, **1**, 530–543, https://doi.org/10.1038/s43017-020-0078-2.

Cane, M. A., A. C. Clement, A. Kaplan, Y. Kushnir, D. Pozdnyakov, R. Seager, S. E. Zebiak, and R. Murtugudde, 1997: Twentieth century sea surface temperature trends. *Science*, **275**, 957–960, https://doi.org/10.1126/science.275.5302.957.

Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the last 148 years. *Science*, **428**, 733–736, https://doi.org/10.1038/nature02439.

Chung, E., A. Timmermann, B. Soden, K. Ha, and V. John, 2019: Reconciling opposing Walker circulation trends in observations and model projections. *Nat. Climate Change*, **9**, 405–412, https://doi.org/10.1038/s41558-019-0446-4.

Clement, A. C., R. Seager, M. A. Cane, and S. E. Zebiak, 1996: An ocean dynamical thermostat. *J. Climate*, **9**, 2190–2196, https://doi.org/10.1175/1520-0442(1996)009<2190:AODT>2.0.CO;2.

Coats, S., and K. Karnauskas, 2017: Are simulated and observed twentieth century tropical Pacific sea surface temperature trends significant relative to internal variability? *Geophys. Res. Lett.*, **44**, 9928–9937, https://doi.org/10.1002/2017GL074622.

Delworth, T., F. Zeng, A. Rosati, G. A. Vecchi, and A. T. Wittenberg, 2015: A link between the hiatus in global warming and North American drought. *J. Climate*, **28**, 3834–3845, https://doi.org/10.1175/JCLI-D-14-00616.1.

Deser, C., A. S. Phillips, and M. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.*, **37**, L10701, https://doi.org/10.1029/2010GL043321.

——, R. Guo, and F. Lehner, 2017: The relative contributions of tropical Pacific sea surface temperatures and atmospheric internal variability to the recent global warming hiatus. *Geophys. Res. Lett.*, **44**, 7945–7954, https://doi.org/10.1002/2017GL074273.

——, and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

DiNezio, P., A. C. Clement, G. A. Vecchi, B. J. Soden, B. P. Kirtman, and S. Lee, 2009: Climate response of the equatorial Pacific to global warming. *J. Climate*, **22**, 4873–4892, https://doi.org/10.1175/2009JCLI2982.1.

Eyring, V., S. Bony, G. Meehl, C. Senior, B. Stevens, R. Stouffer, and K. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Feely, R. A., R. Wanninkhof, T. Takahashi, and P. Tans, 1999: Influence of El Niño on the equatorial Pacific contribution to atmospheric CO_2 accumulation. *Nature*, **398**, 597–601, https://doi.org/10.1038/19273.

Freeman, E., and Coauthors, 2017: ICOADS release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.*, **37**, 2211–2232, https://doi.org/10.1002/joc.4775.

He, C., B. Wu, L. Zou, and T. Zhou, 2017: Responses of the summertime subtropical anticyclones to global warming. *J. Climate*, **30**, 6465–6479, https://doi.org/10.1175/JCLI-D-16-0529.1.

Heede, U. K., and A. Federov, 2021: Eastern equatorial Pacific warming delayed by aerosols and thermostat response to CO_2 increase. *Nat. Climate Change*, **11**, 696–703, https://doi.org/10.1038/s41558-021-01101-x.

——, ——, and N. Burls, 2021: A stronger versus weaker Walker: Understanding model differences in fast and slow tropical Pacific responses to global warming. *Climate Dyn.*, **57**, 2505–2522, https://doi.org/10.1007/s00382-021-05818-5.

Held, I., M. Ting, and H. Wang, 2002: Northern winter stationary waves: Theory and modeling. *J. Climate*, **15**, 2125–2144, https://doi.org/10.1175/1520-0442(2002)015<2125:NWSWTA>2.0.CO;2.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, https://doi.org/10.1175/JCLI-D-12-00837.1.

Horel, J. D., 1982: On the annual cycle of the tropical Pacific atmosphere and ocean. *Mon. Wea. Rev.*, **110**, 1863–1878, https://doi.org/10.1175/1520-0493(1982)110<1863:OTACOT>2.0.CO;2.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, https://doi.org/10.1175/JCLI-D-14-00006.1.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analysis of sea surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. *Int. J. Climatol.*, **25**, 865–879, https://doi.org/10.1002/joc.1169.

Israeli, M., N. Naik, and M. A. Cane, 2000: An unconditionally stable scheme for the shallow water equations. *Mon. Wea. Rev.*, **128**, 810–823, https://doi.org/10.1175/1520-0493(2000)128<0810:AUSSFT>2.0.CO;2.

Jeffrey, S., L. Rotstayn, M. Collier, S. Dravitski, C. Hamalainen, C. Moeseneder, K. Wong, and J. Syktus, 2013: Australia's CMIP5 submission using the CSIRO-Mk3.6 model. *Aust. Meteor. Oceanogr. J.*, **63**, 1–13, https://doi.org/10.22499/2.6301.001.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature: 1856–1991. *J. Geophys. Res.*, **103**, 18 567–18 589, https://doi.org/10.1029/97JC01736.

Kay, J. E., and Coauthors, 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. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Kennedy, L. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations since 1850: 1. Measurement and sampling errors. *J. Geophys. Res.*, **116**, D14103, https://doi.org/10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, https://doi.org/10.1029/2010JD015220.

Kirchmeier-Young, M., F. Zwiers, and N. Gillett, 2017: Attribution of extreme evens in Arctic sea ice extent. *J. Climate*, **30**, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1.

Knutson, T. R., and S. Manabe, 1995: Time-mean response over the tropical Pacific to increased CO_2 in a coupled ocean–atmosphere model. *J. Climate*, **8**, 2181–2199, https://doi.org/10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2.

Kosaka, Y., and S. Xie, 2013: Recent global warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, https://doi.org/10.1038/nature12534.

——, and ——, 2016: The tropical Pacific as a key pacemaker of the variable rates of global warming. *Nat. Geosci.*, **9**, 669–673, https://doi.org/10.1038/ngeo2770.

Li, G., and S. Xie, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765–1780, https://doi.org/10.1175/JCLI-D-13-00337.1.

Maher, N., and Coauthors, 2019: The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst.*, **11**, 2050–2069, https://doi.org/10.1029/2019MS001639.

Mastrandrea, M., K. J. Mach, G. Plattner, O. Edenhofer, T. Stocker, C. Field, K. Ebi, and P. Matschoss, 2011: The IPCC guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change*, **108**, 675, https://doi.org/10.1007/s10584-011-0178-6.

McKinley, G. A., M. Follows, and J. Marshall, 2004: Mechanisms of air-sea CO_2 flux variability in the equatorial Pacific and the North Atlantic. *Global Biogeochem. Cycles*, **18**, GB2011, https://doi.org/10.1029/2003GB002179.

Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. Trenberth, 2011: Model-based evidence of deep-ocean heat uptake during surface temperature hiatus periods. *Nat. Climate Change*, **1**, 360–364, https://doi.org/10.1038/nclimate1229.

Olonscheck, D., M. Rugenstein, and J. Marotzke, 2020: Broad consistency between observed and simulated trends on sea surface temperature patterns. *Geophys. Res. Lett.*, **47**, e2019GL086773, https://doi.org/10.1029/2019GL086773.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Rödenbeck, C., and Coauthors, 2015: Data-based estimates of the ocean carbon sink variability—First results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM) *Biogeosciences*, **12**, 7251–7278, https://doi.org/10.5194/bg-12-7251-2015.

Rodgers, K., J. Lin, and T. Frohlicher, 2015: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, **12**, 3301–3320, https://doi.org/10.5194/bg-12-3301-2015.

Santoso, A., M. McPhaden, and W. Cai, 2007: The defining characteristics of ENSO extremes and the strong 2015/16 El Niño. *Rev. Geophys.*, **55**, 1079–1129, https://doi.org/10.1002/2017RG000560.

Sarachik, E. S., and M. A. Cane, 2010: *The El Niño-Southern Oscillation Phenomenon*. Cambridge University Press, 384 pp.

Seager, R., M. A. Cane, N. Henderson, D. Lee, R. Abernathey, and H. Zhang, 2019: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat. Climate Change*, **9**, 517–522, https://doi.org/10.1038/s41558-019-0505-x.

Sun, L., M. Alexander, and C. Deser, 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**, 7823–7843, https://doi.org/10.1175/JCLI-D-18-0134.1.

Takahashi, T., and Coauthors, 2002: Global sea-air CO_2 flux based on climatological surface ocean pCO_2 and seasonal biological and temperature effects. *Deep-Sea Res.*, **2002**, 1601–1622, https://doi.org/10.1016/S0967-0645(02)00003-6.

Valdivieso, M., and Coauthors, 2017: An assessment of air-sea heat fluxes from ocean and coupled reanalyses. *Climate Dyn.*, **49**, 983–1008, https://doi.org/10.1007/s00382-015-2843-3.

Vecchi, G. A., and B. J. Soden, 2007: Global warming and the weakening of the tropical circulation. *J. Climate*, **20**, 4316–4340, https://doi.org/10.1175/JCLI4258.1.

Watanabe, M., J. Dufresne, Y. Kosaka, T. Mauritsen, and H. Tatebe, 2020: Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nat. Climate Change*, **11**, 33–37, https://doi.org/10.1038/s41558-020-00933-3.

Wyrtki, K., 1981: An estimate of equatorial upwelling in the Pacific. *J. Phys. Oceanogr.*, **11**, 1205–1214, https://doi.org/10.1175/1520-0485(1981)011<1205:AEOEUI>2.0.CO;2.

Yang, H., and F. Wang, 2009: Revisiting the thermocline depth in the equatorial Pacific. *J. Climate*, **22**, 3856–3863, https://doi.org/10.1175/2009JCLI2836.1.

Zeng, N., A. Mariotti, and P. Wetzel, 2005: Terrestrial mechanisms of interannual CO_2 variability. *Global Biogeochem. Cycles*, **19**, GB1016, https://doi.org/10.1029/2004GB002273.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like decade-to-century scale variability: 1900–93. *J. Climate*, **10**, 1004–1020, https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zuo, H., M. Balmaseda, S. Tietsche, K. Mogensen, and M. Mayer, 2019: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Sci.*, **15**, 779–808, https://doi.org/10.5194/os-15-779-2019.