How Well Do We Know ENSO's Climate Impacts over North America, and How Do We Evaluate Models Accordingly? 🖉

CLARA DESER, ISLA R. SIMPSON, ADAM. S. PHILLIPS, AND KAREN A. MCKINNON

National Center for Atmospheric Research, Boulder, Colorado

ABSTRACT

The role of sampling variability in ENSO composites of winter surface air temperature and precipitation over North America during the period 1920–2013 is assessed for observations and ensembles of coupled model simulations in which sea surface temperature anomalies in the tropical eastern Pacific are nudged to those of the real world. The individual members of each model ensemble show a surprising amount of diversity in their ENSO composites, despite being constructed from the same observed set of 18 El Niño and 14 La Niña events. For a given model, this ensemble spread can only be due to sampling variability, that is, aliasing of internal variability that is unrelated to ENSO, which in turn is shown to arise from internal atmospheric dynamics rather than coupled ocean–atmosphere processes. Analogous ensemble spread is evident in 2000 synthetic ENSO composites based on observations using random sampling techniques. These synthetic composites provide information on the range of spatial patterns and amplitudes associated with imperfect estimation of the forced ENSO signal in the observational record. In some locations, the amplitude of the estimated ENSO signal can vary by more than a factor of two. This observational uncertainty necessitates an approach to model assessment that considers not only the model's forced response to ENSO, given by its ensemble-mean ENSO composite, but also its representation of internal variability unrelated to ENSO. Such an approach is used to reveal fidelities and shortcomings in the Community Earth System Model, version 1.

1. Introduction

The largest known source of seasonal climate forecast skill over North America is El Niño – Southern Oscillation (ENSO), the leading mode of interannual variability of the tropical ocean–atmosphere system (Shukla et al. 2000; Tippett et al. 2012; L'Heureux et al. 2015). ENSO affects North American climate through changes in the large-scale atmospheric circulation driven by anomalous deep convection and associated latent heat release in the tropical Indo-Pacific (e.g., Bjerknes 1969; Horel and Wallace 1981; Held et al. 1989). These influences generally maximize in boreal winter and early spring when atmospheric conditions are favorable for Rossby wave propagation from the tropics to the Northern Hemisphere and when the SST anomalies in the tropical Pacific are the largest (Ropelewski and

Halpert 1986; Larkin and Harrison 2005; Chiodi and Harrison 2013; L'Heureux et al. 2015; and many others). During ENSO's positive phase (El Niño), anomalous southerly winds advect warmer air over Alaska and Canada while anomalous northerlies bring cooler air to the southeastern United States, and a strengthened and southward-shifted storm track brings above-normal precipitation to the southern tier of the United States and drier conditions to the Ohio and upper Mississippi River valleys. Opposite conditions tend to prevail during ENSO's negative phase (La Niña). While the impacts described above are typical of ENSO events, they do not necessarily occur during every episode. A recent case in point is the El Niño of 2015/16, which failed to bring anticipated and much-needed rains to Southern California and the U.S. Southwest desert, despite the fact that it was as strong as the El Niño events of 1982/83 and 1997/98, which did bring copious amounts of precipitation to this region. Similarly, the strong El Niño events of 1991/92 and 1987/88 also lacked many of the expected impacts over North America, as did the strong La Niña events of 1973/74 and 2007/08 (www.climate.gov). These counterexamples are consistent with the

Corresponding author: Dr. Clara Deser, cdeser@ucar.edu

fact that ENSO generally accounts for <25% of the variance of winter and spring climate anomalies over North America, despite it being a dominant source of predictability (not shown, but see www.climate.gov).

The canonical impacts of ENSO may be obscured during any given El Niño or La Niña event by competing influences from other sources of natural climate variability. In addition, differences in the character of each event (e.g., Rasmusson and Carpenter 1982; Deser and Wallace 1987; Capotondi et al. 2015; Takahashi and Martínez 2018) may affect their atmospheric teleconnections and associated climate impacts over North America (e.g., Garfinkel et al. 2013; Johnson and Kosaka 2016). These issues motivate the question: how well do we know ENSO's canonical influence on North American climate? Empirical studies typically attempt to isolate ENSO-forced signals by compositing over a large number of events or by applying regression analysis to a long period of record. With a sufficiently long dataset (i.e., a large enough sample of ENSO events), the noise due to variability that exists in the absence of ENSO will be minimized, revealing the true forced response. To what extent is the observational record adequate to identify the forced response to ENSO without significant aliasing of unrelated variability?

While empirical studies based on composite or regression analysis almost always include an assessment of statistical significance on the estimated ENSO signals, this information does not necessarily convey the magnitude of the uncertainty at each location, nor does it convey the spatial pattern of the uncertainty. For example, if the uncertainty arises from large-scale atmospheric variability, then the "noise" imparted to observationally derived ENSO signals will also be characterized by large-scale spatial patterns. In view of these issues, Deser et al. (2017, hereafter D17) proposed an approach that integrates information on both pattern and amplitude uncertainty that accompanies any empirical estimate of ENSO response based upon limited data. In addition, they showed the utility of this integrated perspective when evaluating the realism of ENSO signals in climate models, which have the luxury of much larger samples sizes. The focus of D17 was on the NH atmospheric circulation response to ENSO in boreal winter. Here we extend this approach to investigate the surface air temperature (SAT) and precipitation (PR) responses to ENSO over North America. To be consistent with D17, we use the same period of record (1920–2013) and random sampling technique (with replacement) to construct synthetic ENSO composites, each of which could have plausibly happened had a different temporal sequence of natural

variability, unrelated to ENSO, occurred. These synthetic ENSO composites provide important context for, and uncertainty bounds on, the one composite that actually occurred. Issues related to ENSO diversity and nonlinearity within these synthetic composites are also addressed.

We also evaluate ENSO composites of SAT and PR over North America from a multimodel ensemble of coupled simulations (the same used in D17) whose tropical eastern Pacific sea surface temperature anomalies (SSTA) are nudged to observations during 1920–2013. Each model provides an ensemble of simulations starting from slightly different initial conditions. For each, we construct an ENSO composite using the same set of events as in our observational analysis. The resulting range of composites across the individual members of a given model ensemble provides a direct assessment of the uncertainties associated with any single composite sample (i.e., the model's ensemble spread is the counterpart of the spread across the observationally based synthetic composites, for which there is only one actual composite sample). That is, each composite within a given model ensemble represents an estimate of the model's true forced response to ENSO combined with a different sampling of its (unrelated) internal variability (throughout this paper, we shall use the terms "internal" and "natural" interchangeably). Using the uncertainties derived for the observed ENSO composite, we discriminate between true biases in the models' forced response to ENSO and apparent biases that arise from limited sampling of non-ENSO-related natural variability. In this way, we also evaluate whether the spread across the ensemble members of a given model is realistic. We focus on December–February (DJF), the season when ENSO generally has its largest impact over much of North America. We also briefly show results for February–April (FMA).

The rest of this study is organized as follows. The observational datasets, model simulations, and methodology are described in section 2. Results are presented in section 3, beginning with observed and simulated ENSO composites of temperature and precipitation (section 3a), evaluation of the forced ENSO response and internal variability in Community Earth System Model, version 1 (CESM1; section 3b), the range of observationally based synthetic ENSO composites (section 3c), comparison of ENSO composites across models (section 3d), the contributions of ENSO nonlinearity (section 3e) and ENSO diversity (section 3f) to the synthetic observational composites, and ending with observed ENSO composites for late winter (section 3g). Discussion and summary are provided in section 4.

2. Data and methods

a. Observational data

We use SAT from Berkeley Earth Surface Temperature (BEST; Rohde et al. 2013) and PR from Global Precipitation Climatology Centre (GPCC), version 7 (Schneider et al. 2014), both on a 1° latitude × 1° longitude grid. In addition, we make use of sea level pressure (SLP) and PR from the Twentieth Century Reanalysis (20CR), version 2c (Compo et al. 2011), on a 2° latitude × 2° longitude grid, PR from ERA-20C (Poli et al. 2016) on a 1.5° latitude × 1.5° longitude grid, and PR from the Global Precipitation Climatology Project (GPCP), version 2.3 (Adler et al. 2003), on a 2.5° latitude × 2.5° longitude grid. Our analysis is based on the period 1920–2013, except for GPCP, which is based on the years 1979–2013.

b. Model simulations

We use the same model simulations as D17, which are briefly summarized here; additional information is provided in D17. First, we make use of a coordinated set of tropical Pacific pacemaker experiments (referred to as PACE) performed with three state-of-the-art coupled climate models: CESM1; Climate Model, version 2.1 (CM2.1); and MIROC5. These experiments follow the protocol of Kosaka and Xie (2013), in which monthly SSTAs in the eastern tropical Pacific (10°S–10°N, 160°–90°W, with a linearly tapering buffer zone of 10° in latitude and 20° in longitude) are nudged with a 2-day damping time scale to those from the NOAA Extended Reconstruction Sea Surface Temperature, version 3b (ERSSTv3b), dataset; note that the SST mean state in each model is maintained. An ensemble of experiments was conducted with each model (10 for CESM1 and CM2.1, and 5 for MIROC5), produced by randomly perturbing the initial atmospheric temperatures in each member by a small (order 10^{-14} K) amount. These coupled model simulations arguably provide the most realistic setting for evaluating the models' ENSO composites, since only SSTAs in the eastern tropical Pacific are nudged to observations, leaving the rest of the global climate system free to respond in an appropriately coupled manner (e.g., with two-way ocean–atmosphere interaction; Alexander et al. 2002). Our analysis is based on the 1920–2013 period common to each model.

We also make use of a companion 10-member ensemble conducted with the atmosphere–land configuration of CESM1 in which the observed SST evolution is specified throughout the tropics (within 28° latitude and a linearly tapering buffer zone to 35° latitude) and the observed SST climatological seasonal cycle is prescribed elsewhere [the so-called Tropical Ocean and

Global Atmosphere (TOGA) protocol]. While these simulations have the advantage of a realistic tropic-wide distribution of SST anomalies during ENSO (and a realistic SST mean state), they are more idealized than the PACE experiments in the sense that they lack two-way ocean–atmosphere coupling and extratropical SST variability. All simulations (PACE and TOGA) include the historical (and RCP8.5 after 2005) radiative forcing protocols of CMIP5 (Taylor et al. 2012).

Last, we make use of a 2600-yr preindustrial control simulation (1850 radiative conditions) of the atmosphere–land configuration of CESM1 with a prescribed repeating seasonal cycle of SSTs and sea ice conditions taken from the long-term climatology of a companion 2200-yr preindustrial control run of the fully coupled CESM1 (see Kay et al. 2015). This lengthy "atmosphere only" control simulation provides robust statistics on the simulated level of atmospheric circulation variability that exists in the absence of ENSO and other SSTA forcing.

c. Methods

We use the same methodology as D17, of which a summary is given below. We compute monthly anomalies by subtracting the long-term monthly means based on the period 1920–2013 from the corresponding month of each year. We then form DJF and FMA averages from the monthly anomalies and linearly detrend each seasonal time series to reduce potential effects from secular climate change. Following D17, we evaluate statistical significance using a two-sided Student's t test at the 10% significance level as well as a random sampling approach discussed below. We identify 18 El Niño (EN) and 14 La Niña (LN) events during 1920–2013 according to the criterion that the observed detrended DJF Niño-3.4 (5°S–5°N, 120°–170°W) SST index exceeds 1 standard deviation (σ) or falls below -1 σ, respectively [using November–January (NDJ) in place of DJF does not change the event selection]. We form ENSO composites by subtracting the average of the 14 LN events from the average of the 18 EN events. Unless noted otherwise, all results are based on DJF.

To evaluate the influence of sampling variability, we form 2000 synthetic ENSO composites for observations and for each model simulation by randomly sampling with replacement from among the 18 EN events and the 14 LN events, always retaining 18 samples for the former and 14 samples for the latter (these samples will necessarily omit some events and repeat others). As shown in D17, the majority of these "bootstrapped" ENSO composites consist of 11–12 unique EN events and nine unique LN events and a maximum repetition of three events of either sign. We also form 2000 synthetic ENSO composites for each model by drawing from all ensemble
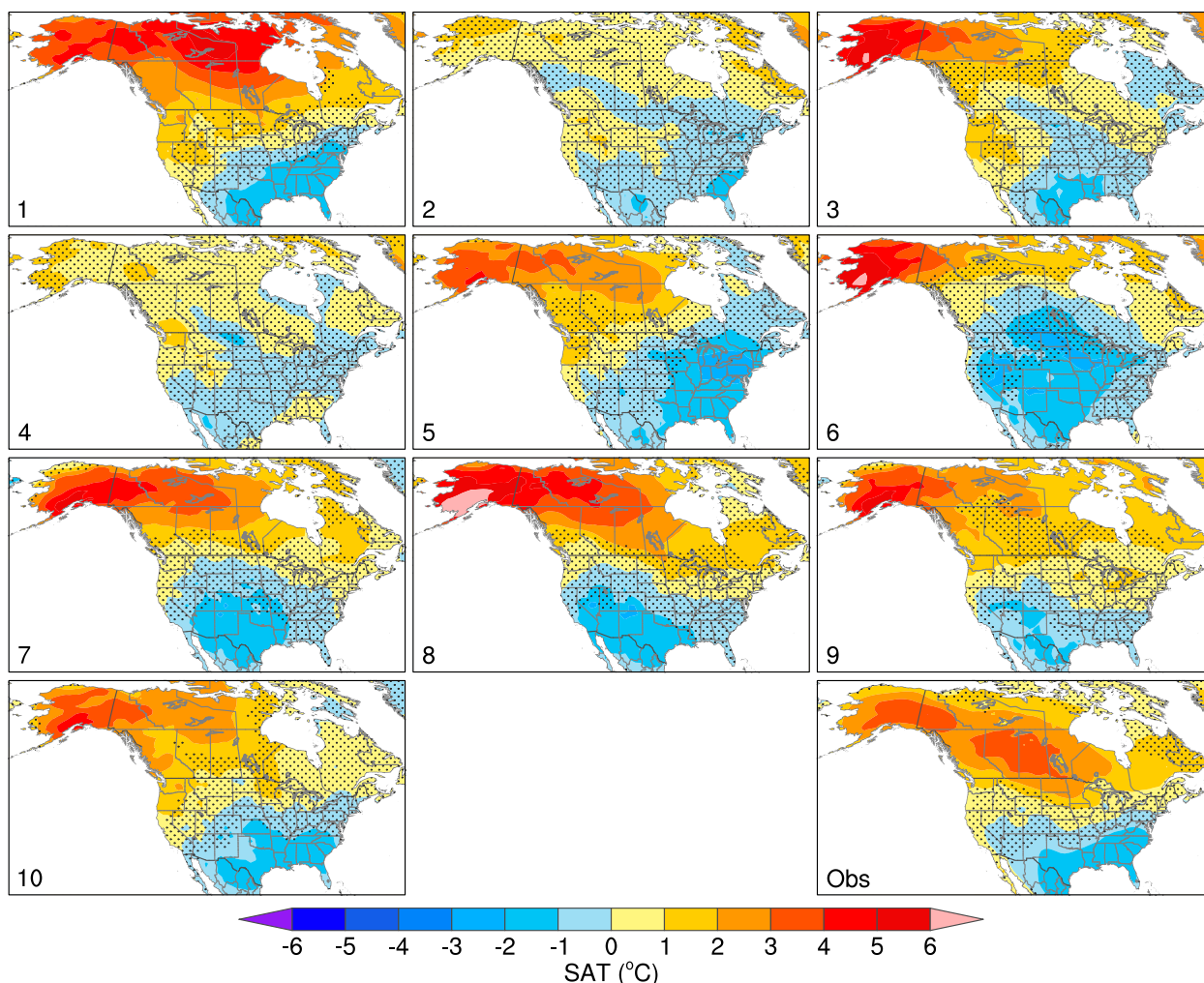
CESM1 Pacemaker ENSO Composites

FIG. 1. ENSO composites of DJF SAT (°C) from each of the 10 CESM1 PACE simulations (labeled 1 through 10) and from (lower right) observations. Each composite is based on the same set of 18 El Niño events minus 14 La Niña events during 1920–2013. Values not significant at the 10% confidence level based on a two-sided *t* test are stippled.

members simultaneously (resulting in a larger number of unique events and lower repetition rate).

3. Results

a. Observed and simulated ENSO composites

The observed ENSO SAT composite shows a statistically significant dipole pattern of positive anomalies over much of Canada and Alaska (maximum values ~4°C) and negative anomalies over the southeastern United States (maximum amplitudes ~2°C; Fig. 1, lower right panel). This pattern is evident to some degree in all of the CESM1 PACE simulations, although the magnitudes and exact locations of significant SAT

anomalies vary considerably (Fig. 1). For example, the warming over western Canada and Alaska is nearly twice as large in simulation 8 compared to simulation 5, and the cooling over the United States is located in the Southeast (as observed) in simulation 1 and over the Southwest in simulation 8. Notably, simulations 2 and 4 show weak (amplitudes <1.5°C) and generally insignificant SAT anomalies throughout North America. A similar level of diversity in ENSO SAT composites is apparent in the GFDL and MIROC PACE (Figs. S1 and S2 in the online supplemental material) and CESM1 TOGA (Fig. S3) ensembles, although the latter shows generally larger amplitudes compared to its PACE counterpart. Recall that each simulated composite is based on the same set of ENSO events as the

CESM1 Pacemaker ENSO Composites
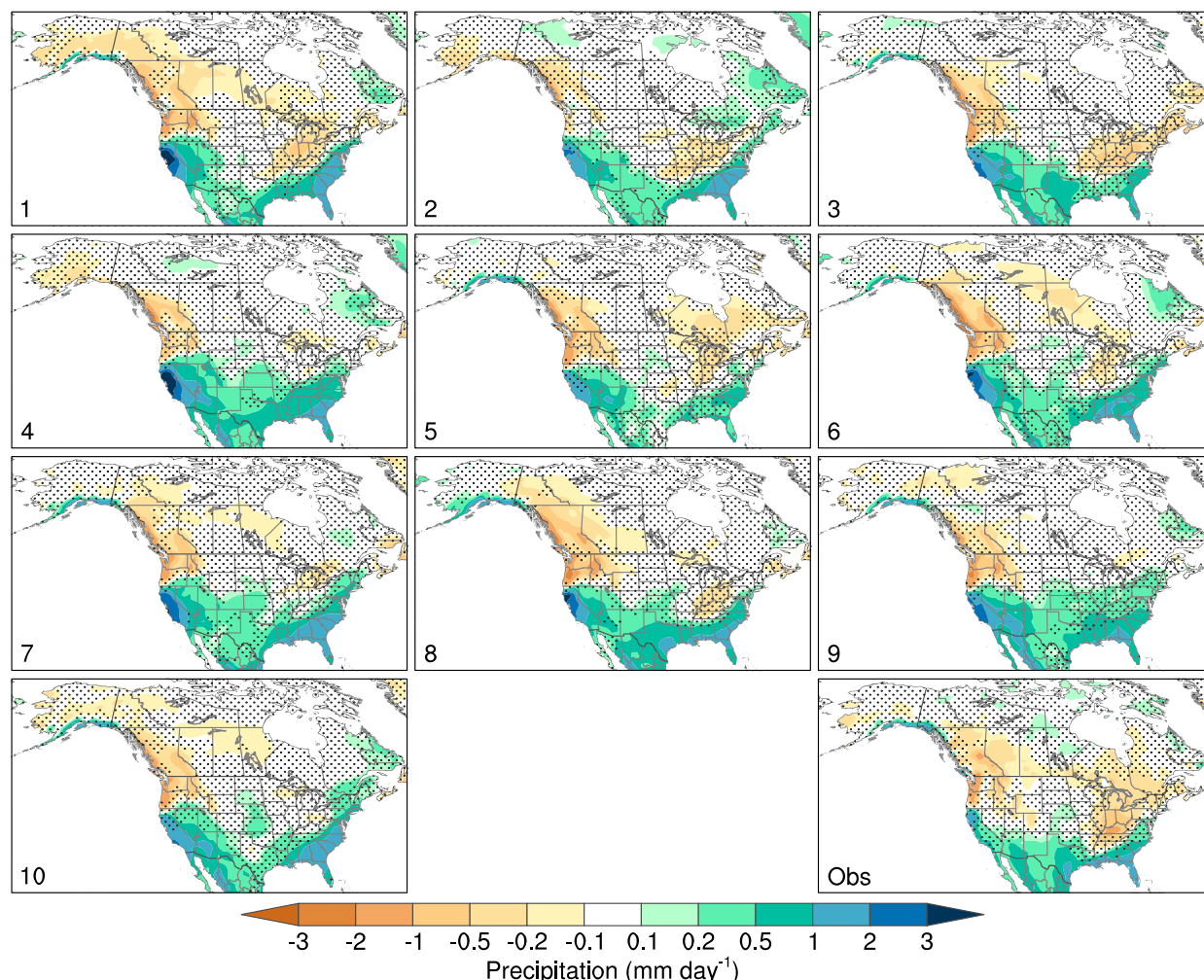
FIG. 2. As in Fig. 1, but for PR (mm day^{-1}).

observed composite, highlighting the role of sampling uncertainty.

The observed ENSO PR composite shows significant positive anomalies over the southernmost U.S. states and significant negative anomalies over southern Ontario, Quebec, and the Ohio Valley–Upper South (the region encompassing Ohio, Michigan, Indiana, Kentucky, Tennessee, and West Virginia), as well as interior British Columbia, Alberta, western Manitoba, and parts of the U.S. Northwest (Fig. 2k). The CESM1 PACE ensemble generally reproduces this observed pattern, but the amplitude and statistical significance of the regional PR anomalies vary across the members. For example, California rainfall anomalies are considerably weaker and more in line with observations in members 2, 5, and 10 compared to members 1, 4, 6, and 8. Similarly, the spatial extent, amplitude, and significance of

drying over western Canada range considerably across the simulations (note the contrast between members 2 and 8), as well as over the Ohio Valley–Upper South (note the absence of drying in members 4, 9, and 10). A similar level of diversity in ENSO PR composites is evident in the other models and the CESM1 TOGA ensemble (Figs. S4–S6).

These results highlight that a sample size of 18 EN and 14 LN events may be insufficient to accurately determine the ENSO-forced SAT and PR responses in models because of the presence of unrelated internal variability that may obscure the ENSO signal. They also raise the related issue of how well the observed response to ENSO is known, even with 94 years of data (1920–2013). Finally, the results underscore the challenge of evaluating the ENSO response in models given sampling uncertainty in both the observational target
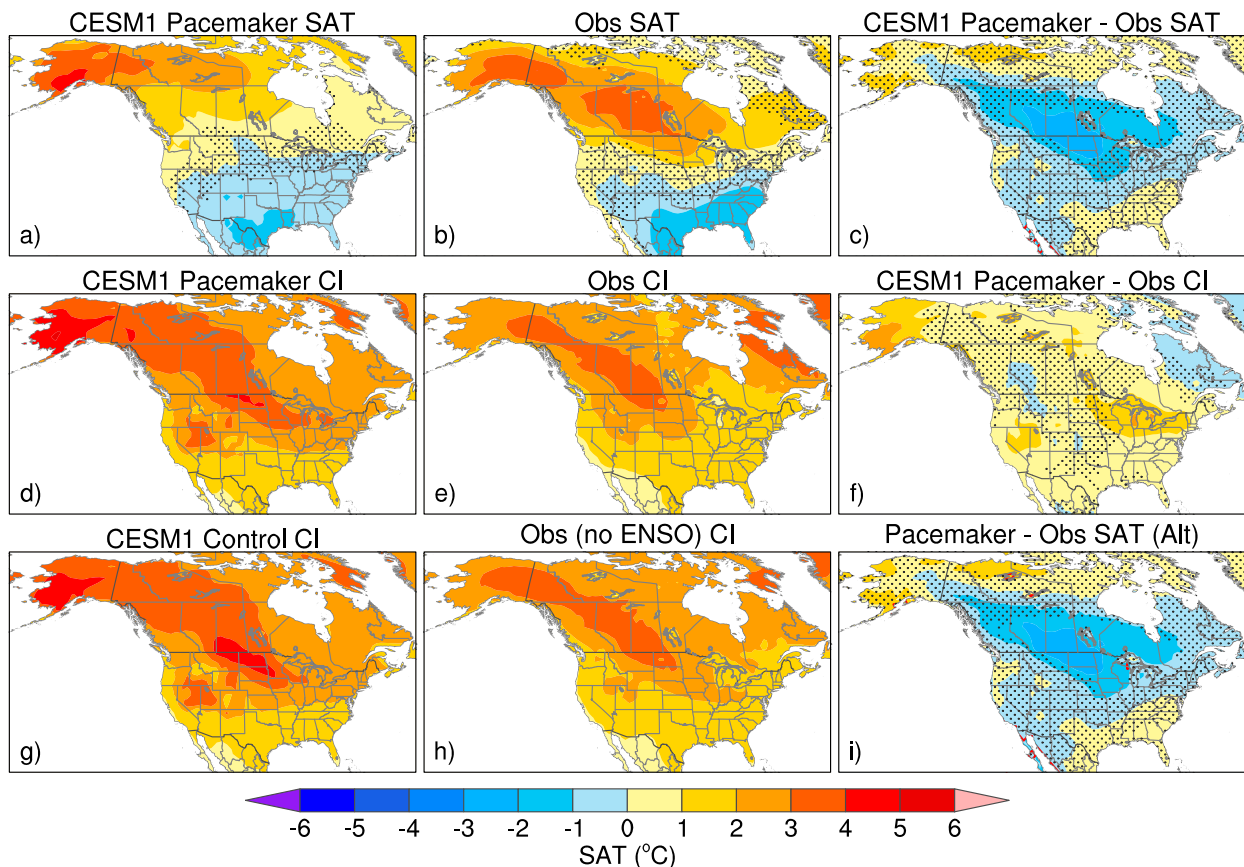
ENSO Composites

FIG. 3. (top row) ENSO composites of DJF SAT (°C) for (a) ensemble mean of the 10 CESM1 PACE simulations, (b) observations, and (c) their difference. Stippling in (a) and (b) indicates values not significant at the 10% confidence level based on a two-sided *t* test. Stippling in (c) indicates that the observed value lies outside of the 5%–95% values from the model's 2000 bootstrapped composites, and red contours (which occur only over Baja California) indicate the observed value lies outside any of the model's bootstrapped composites. (middle row) 5%–95% CIs (°C) on the SAT ENSO composites based on 2000 bootstrapped samples for (d) CESM1 PACE simulations, (e) observations, and (f) their difference. Stippling in (f) indicates regions where the observed value falls within the spread of values from the individual simulations. (bottom row) As in middle row, but for (g) the CAM5 atmospheric control simulation and (h) observations after removing ENSO. (i) As in (c), but applying the observed CIs to the model simulations. See text for details.

and in each model simulation. We address these issues next.

b. Evaluating CESM1's response to ENSO

1) TEMPERATURE

The ensemble mean of the 10 CESM1 PACE ENSO composites provides a robust estimate of the model's true response to ENSO, as it is based on a total of 180 EN events and 140 LN events. Accordingly, the ensemble-mean SAT response is significant over most of the continent, with warming in Canada and Alaska (maximum values >3°C in the Yukon and Alaska) and weaker-amplitude cooling over most of the contiguous United States (Fig. 3a). The pattern and amplitude of

the ensemble-mean response is generally similar to the observed composite except that the simulated warming does not penetrate as far southeastward into the central Canadian provinces and the north-central United States (cf. Figs. 3a and 3b). (Note however that individual realizations of the model, most notably simulations 1 and 8, do show a more southerly extension than the ensemble mean; recall Fig. 1). Differencing the observed composite from the ensemble-mean composite reveals a significant cold bias in the model's ENSO response in the central Canadian provinces and U.S. Upper Midwest; all other regions show insignificant differences (Fig. 3c). Here, differences are deemed statistically significant if the value of the observed composite is lower than the 5th percentile or greater than the 95th percentile of the

2000 model bootstrapped ENSO composite values obtained by randomly sampling from among all 10 PACE ensemble members.

To properly evaluate whether the model has a realistic forced response to ENSO, one must also assess its internal variability. This is because a bias in the model's forced response might not be detected if the simulated internal variability is overestimated (i.e., the spread among ENSO composites is too large), or it might be falsely detected if the simulated internal variability is underestimated (i.e., the spread among ENSO composites is too small). Here we assess the model's internal variability by comparing the spread between the 5th and 95th percentiles of the 2000 bootstrapped ENSO composites for observations and CESM1 PACE (the latter computed using all ensemble members).

The spatial distribution of the model's confidence intervals (CIs) (Fig. 3d) is similar to observations (Fig. 3e), with the largest values extending southeastward from Alaska toward the U.S. Great Lakes. However, their amplitudes exceed the observed values in western Alaska, the far western United States, and the eastern third of the continent by up to 1°–2°C (nonstippled regions in Fig. 3f). In these regions, the observed CI is less than the minimum CI of any of the 10 PACE runs, indicating that the model has enhanced variability compared to the real world (the CIs for each PACE simulation are shown in Fig. S7). The larger CIs in the model might obscure a true bias in the model's forced response. To address this possibility, we apply the observed CIs to the model's ensemble-mean ENSO SAT composite and reevaluate the significance on the difference between the SAT values from the ensemble-mean composite and the observed composite (Fig. 3i). The area with statistically significant differences expands slightly to encompass the southern Great Lakes region compared to using the model's CIs (cf. Figs. 3c and 3i). This additional region is thus an area where there is a true bias in the model's forced ENSO response, which had been obscured by the model's overestimated CI. The model's underestimated CIs over far northeastern Canada do not affect assessment of its forced ENSO response (cf. Figs. 3c and 3i). The observed CIs will also be subject to uncertainty, in analogy with the range of CIs across the individual members of CESM1 PACE, but this has not been investigated here.

Next, we make use of the 2600-yr atmospheric control simulation of CESM1 to assess the contribution of internal atmospheric variability to the CIs in CESM1 PACE. To do so, we randomly select two groups of years (with replacement) from the control run, one consisting of 18 winters and the other of 14 winters. We then average the SAT anomaly fields within each group and take their difference, in analogy with how we formed the ENSO composites. We repeat this procedure 2000 times, and use these 2000 random samples to compute the CIs. The CIs from the atmosphere-only control run (Fig. 3g) are very similar to those from the coupled model's PACE simulations (Fig. 3d), indicating that internal atmospheric variability accounts for most of the uncertainty obtained from the model's bootstrapped ENSO composites. This means that the spread in ENSO SAT composites across the members of the CESM1 PACE ensemble is primarily due to the superposition of random (i.e., inherently unpredictable on interannual time scales) internal atmospheric circulation anomalies on the forced ENSO response.

While we cannot isolate the contribution of internal atmospheric dynamics to the observed CIs, we can evaluate the contribution of non-ENSO-related SAT variability. To do this, we compute CIs by randomly sampling from all 93 winters during 1920–2013 after linearly regressing out the Niño-3.4 SST index (Fig. 3h) or by computing CIs from the 61 ENSO-neutral years (not shown). The CIs based on the ENSO residual sample (Fig. 3h) are very similar to those on the ENSO sample (Fig. 3e), indicating that internal atmospheric variability likely underlies the uncertainty in the observed ENSO composite in analogy with the model-based results. We note that the differences between the CIs from the atmospheric control simulation (Fig. 3g) and the ENSO-residual observations (Fig. 3h) are even smaller than those in Fig. 3f (not shown).

2) PRECIPITATION

The ensemble-mean PR composite in CESM1 PACE shows statistically significant drying in British Columbia, the U.S. Pacific Northwest, and the Ohio Valley, and statistically significant wetting along the southern coast of Alaska and throughout the southern United States, with maximum amplitudes in California and the Southeast (Fig. 4a). This pattern resembles the observed composite (Fig. 4b), except for the lack of pronounced drying in the Ohio Valley–Upper South; however, as noted earlier, this region is subject to large member-to-member variation (recall Fig. 2). The larger area of statistically significant PR anomalies in the model composite compared to observations is likely a result of averaging over 10 times as many ENSO events. Differencing the observed composite from the ensemble-mean composite reveals that the model significantly overestimates the amplitude of the PR response to ENSO in Southern California, Nevada, western Utah, and coastal British Columbia, and significantly underestimates it in the Ohio Valley–Upper South and Florida (Fig. 3c).

ENSO Composites
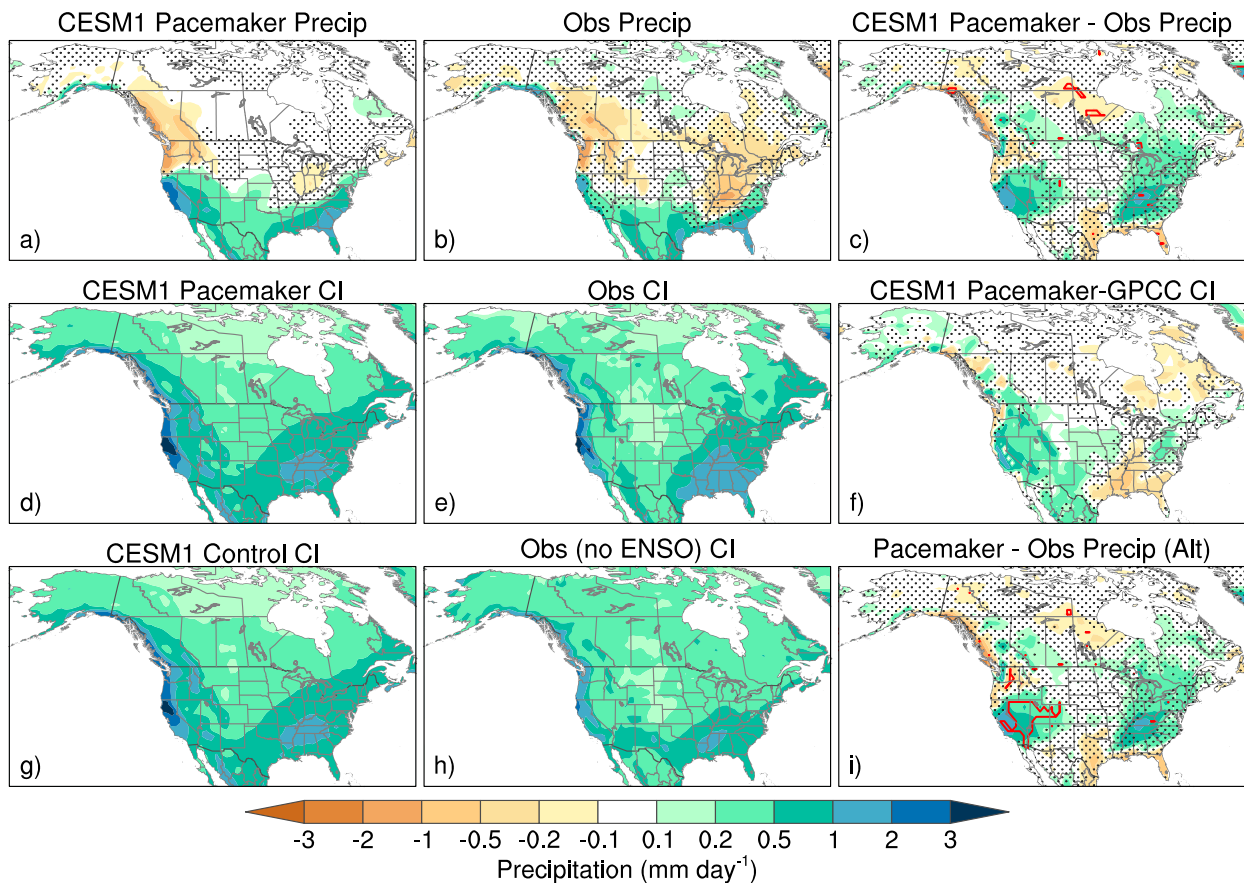
FIG. 4. As in Fig. 3, but for PR (mm day^{-1}).

The modeled (Fig. 4d) and observed (Fig. 4e) CIs on the ENSO PR composites show similar patterns, with the largest amplitudes along the Pacific coast (maximum values ~2–3 mm day^{-1}) and the southeast United States (maximum values ~1–2 mm day^{-1}). However, the model's CIs are larger over most of the western United States and the Ohio Valley, and smaller over portions of the southeast United States (Fig. 4f) compared to observations. In these regions of overestimation (underestimation), the minimum (maximum) CI of any of the 10 CESM1 PACE simulations exceeds (is less than) the observed CI, indicating that the model's variability is likely different from that of the real world in these locations (see Fig. S8 for the CIs of each individual simulation).

Applying the observed CIs to the model's ensemble-mean PR composite accentuates the statistical significance of the model's overestimated PR response to ENSO in the western United States (California, Nevada, Utah, western Colorado, and northwestern Arizona), as the ensemble-mean composite PR values are found to

be larger than those in any of the 2000 observed bootstrapped composite samples (area outlined in red in Fig. 4i). Similarly, the model's underestimated PR response in the Ohio Valley and southern portions of Ontario and Quebec becomes statistically significant because of the smaller observed CIs. This additional region is thus an area where there is a true bias in the model's forced ENSO response, which had been obscured by the model's overestimated CI.

Finally, internal atmospheric variability accounts for almost all of the uncertainty on the model's ENSO PR response (cf. Figs. 4g and 4d). Likewise, non-ENSO-related variability (which we posit stems mostly from internal atmospheric variability) accounts for most of the uncertainty on the observed ENSO PR composite (cf. Figs. 4h and 4e).

c. Range of ENSO composites using bootstrapped observations

How well do we know the spatial pattern and amplitude of SAT and PR responses to ENSO in the real

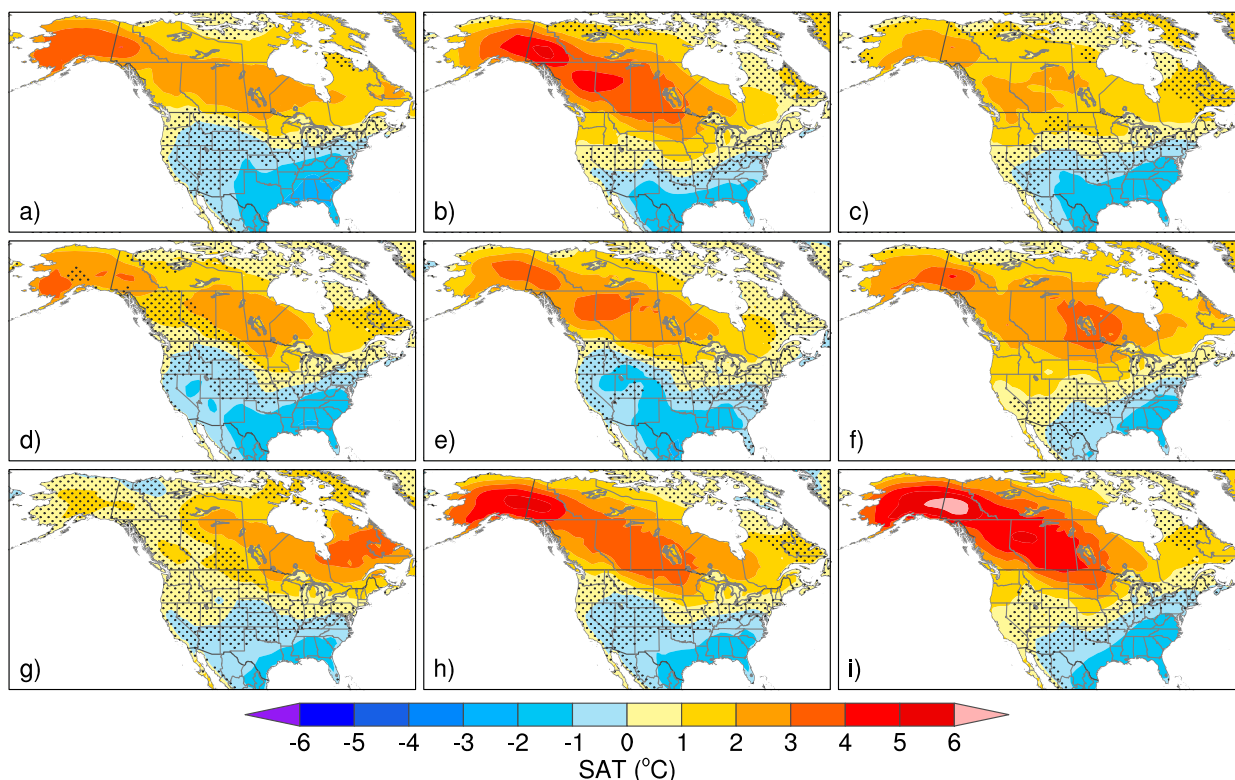## Observational Bootstrapped ENSO Composites

FIG. 5. (a)–(i) Ten randomly selected bootstrapped ENSO composites of DJF SAT (°C) based on observations. Values not significant at the 10% confidence level based on a two-sided *t* test are stippled.

world? As discussed above, the number of ENSO events available for compositing during 1920–2013 may be insufficient to accurately separate the ENSO-forced response from unrelated climate variability. We make use of the 2000 bootstrapped ENSO composites based on observations to address this question. To begin, we show nine of these composites selected at random for SAT (Fig. 5) and PR (Fig. 6). In analogy with the 10 PACE simulations shown in Figs. 1 and 2, these synthetic observational composites display a range of amplitudes and patterns. This range arises from a combination of the different sets of ENSO events in each observational composite (recall the bootstrapping methodology outlined in section 2) and the different sample of climate anomalies unrelated to ENSO in each composite (recall that only the latter contributes to the spread within a given model's PACE simulations). Although the general patterns are similar across the nine synthetic composites, there is considerable variation in amplitude and level of statistical significance. For example, statistically significant warming (4–6°C) over western Canada and Alaska is found in one randomly selected composite (Fig. 5i) but not in another (<2°C; Fig. 5g). Similarly,

significant wetting occurs over Northern California (>3 mm day^{-1}) in one composite (Fig. 6g) but not in another (<0.5 mm day^{-1}; Fig. 6d), while pronounced drying is widespread over the Ohio Valley–Upper South and southern Quebec in one composite (Fig. 6f) but not in another (Fig. 6g).

Next, we perform a more systematic investigation of the 2000 observational bootstrapped ENSO composites, sorting them according to their area-weighted amplitudes in selected regions of interest. For SAT, these regions are the Northwest (NW: 54°–70°N, 175°–98°W) and Southeast (SE: 25°–37°N, 102°–85°W) portions of North America where the actual observed composite shows significant warming and cooling, respectively. For PR, we select the Pacific Northwest (PNW: 42°–58°N, 125°–112°W), Gulf states (GULF: 25°–34°N, 100°–77°W), and California (CA: 32°–42°N, 125°–119°W), regions within which the majority of grid boxes show significant rainfall signals in the actual observed composite (drying for PNW and wetting for GULF and CA). These regions are depicted in Fig. 7 (SAT) and Fig. 9 (PR). For illustration purposes, we display the 10th- and 90th-percentile composite samples based on each

Observational Bootstrapped ENSO Composites
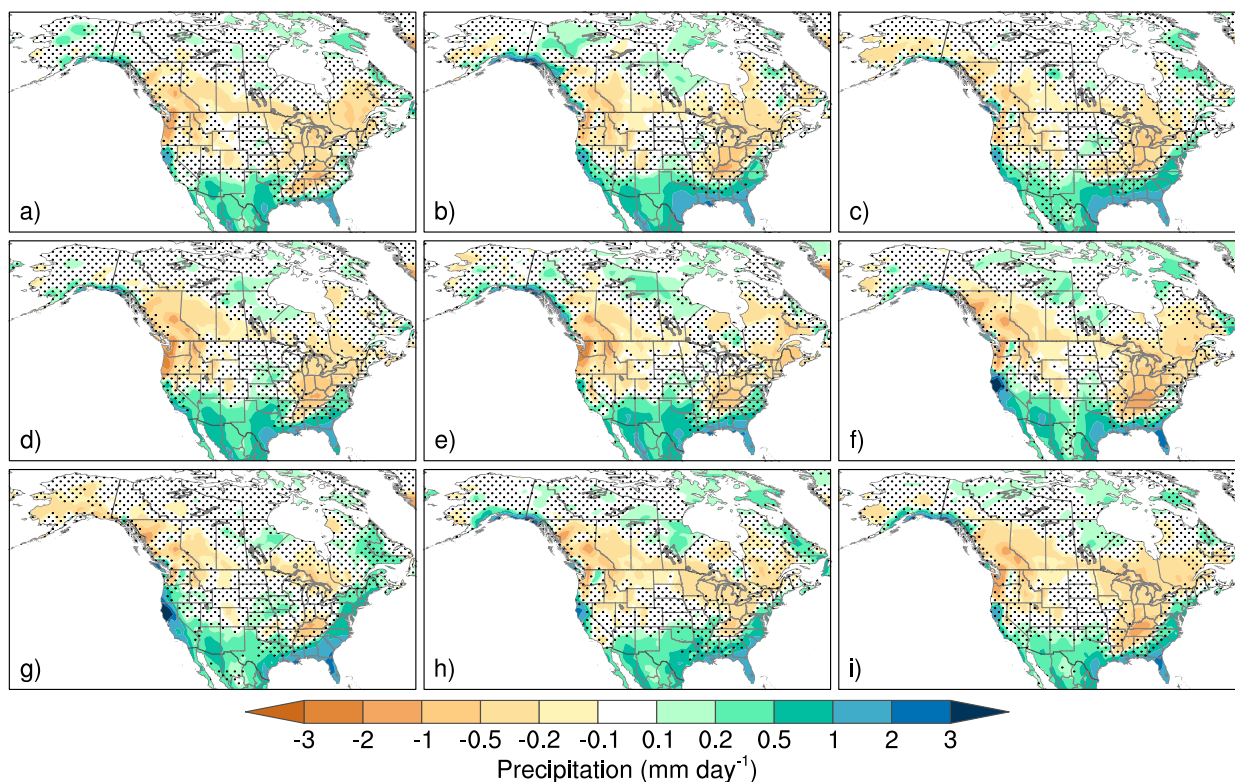
FIG. 6. As in Fig. 5, but for PR (mm day^{-1}).

regional index. Table S1 lists the particular EN and LN events and the number of times they are sampled for each of the 10th- and 90th-percentile composite samples for each regional index. A minimum of 11 distinct EN events and 7 distinct LN events comprise each composite sample, and no single event is sampled more than 4 times (Table S1).

Figures 7a and 7b (7c and 7d) show the observed SAT bootstrapped ENSO composites that lie at the 10th and 90th percentiles, respectively, based on the NW (SE) SAT index. As expected, the index regions show clear differences in SAT anomaly amplitude between the lower- and upper-percentile composites on which they are based. For example, the warming across western Canada and Alaska ranges from ~1°–3°C in the 10th-percentile NW composite (Fig. 7a) compared to ~3°–6°C in the 90th-percentile NW composite (Fig. 7b). Similarly, the cooling over the southeastern United States reaches −3°C in the 10th-percentile SE composite (Fig. 7c) compared to −2°C in the 90th-percentile composite (Fig. 7d). In addition to these local differences in SAT amplitude, there are differences in magnitude, pattern, and statistical significance over the rest of continent. For example, the 90th-percentile SE

composite, which has low-amplitude cooling over the SE United States, shows high-amplitude warming over western Canada and Alaska. This is in contrast to the 90th-percentile NW composite, which shows similar-magnitude warming over western Canada and Alaska but stronger and more widespread areas of significant cooling over the southern United States compared to the 90th-percentile SE composite. Similarly, the amplitude and spatial extent of the cooling over the southern United States is comparable between the 90th-percentile NW and 10th-percentile SE composites, yet the warming over Canada and Alaska is much larger and more widespread in the former compared to the latter. These results illustrate that composite SAT amplitudes in one region may be decoupled from those in another.

To the extent that our random sampling methodology does not introduce additional diversity due to differences among ENSO events (addressed below), the range of ENSO SAT composites in Fig. 7 illustrates what nature might have produced given a different sequence of internal variability independent of ENSO. That is, even with a sample of 18 EN and 14 LN events, the amplitude and to a lesser extent the pattern of the North American
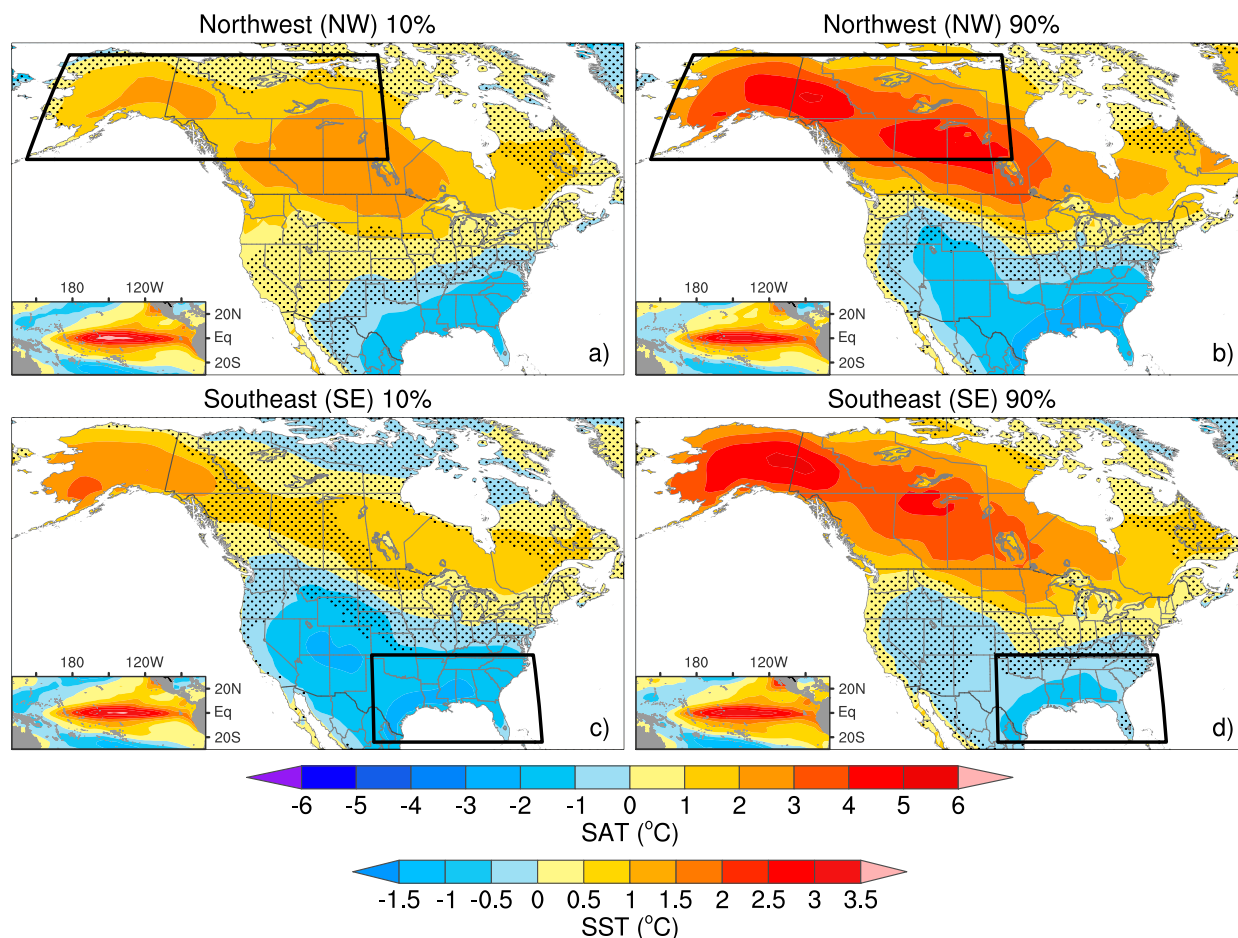
FIG. 7. Tenth- and 90th-percentile bootstrapped ENSO composites of observed DJF SAT (°C) based on the (a),(b) NW and (c),(d) SE SAT indices (index regions outlined in black). Values not significant at the 10% confidence level based on a two-sided t test are stippled. Panel insets show the accompanying SST (°C) composites in the tropical Pacific.

SAT response to ENSO are subject to considerable uncertainty.

To what extent does the different composition of ENSO events in each of the observed bootstrapped composites shown in Fig. 7 contribute to their different SAT anomalies? As a first step in addressing this question, we show maps of the composite SST anomalies in the tropical Pacific that accompany each SAT composite (insets in Fig. 7). All four SST composites show similar patterns and amplitudes, with positive anomalies along the equator (maximum values ~3°–4°C in the central basin 165°–105°W) flanked by weaker-amplitude (<1 C) negative values to the Northwest and Southwest. If anything, the 90th-percentile composite based on the NW SAT index, which features higher-amplitude warming across Canada and Alaska, shows weaker SST anomalies in the central equatorial Pacific compared to its

10th-percentile counterpart (cf. Figs. 7a and 7b). However, this difference is likely a result of random chance, since there is no systematic relationship between the Niño-3.4 SST and NW SAT indices across the 2000 observed composite bootstrapped samples as shown by the scatterplot in Fig. 8a. The large scatter indicates that the precise pair of values of any particular composite, and by extension the pair of spatial patterns shown in Figs. 7a and 7b, is likely due to chance.

For example, for a given value of the NW SAT index such as 3.1°C, which is close to the value of the 90th-percentile sample (3.2°C), there is a wide range of possible composite Niño-3.4 SSTA values (2.2° to 3.0°C) across the 2000 bootstrapped ENSO composites (Fig. 8a). Thus, the small difference in Niño-3.4 SST values (0.2°C) between the 10th- and 90th-percentile NW SAT index composites is unlikely to be the cause of
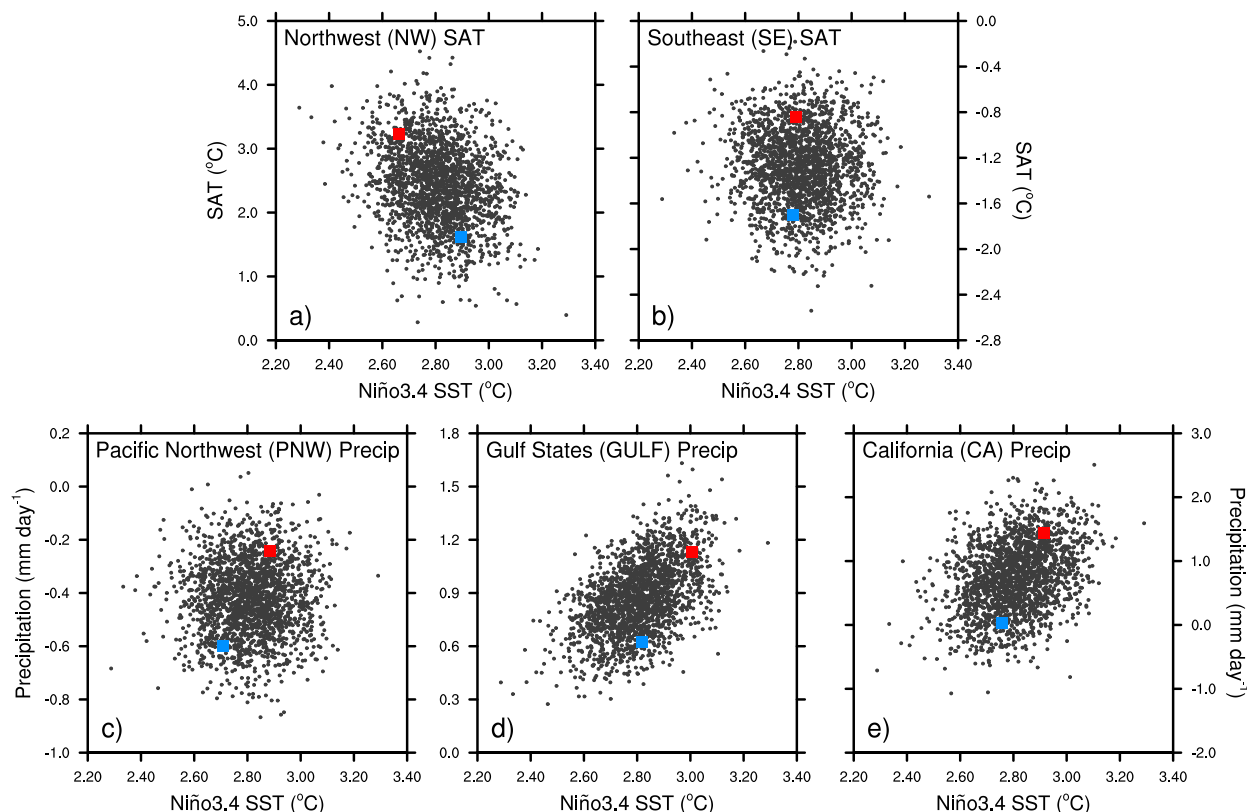
FIG. 8. Scatterplots of ENSO composite values of the Niño-3.4 SST index against the (a) NW SAT index, (b) SE SAT index, (c) PNW PR index, (d) GULF PR index, and (e) CA PR index based on 2000 observed bootstrapped samples. Blue and red symbols show the 10th- and 90th-percentile values, respectively, of each index. SAT and SST values are in °C and PR values are in mm day^{-1}.

the approximately twofold difference in their NW SAT values (1.6°° vs 3.2°C). Conversely, for a given value of the Niño-3.4 SST index, say 2.6°C, the NW SAT index can range from 1° to 4°C. Similar remarks apply to the 10th- and 90th-percentile composites based on the SE SAT index (Fig. 8b). Finally, although there is a weak linear dependence of the NW SAT index on the Niño-3.4 SST index across the 2000 bootstrapped ENSO composites (correlation coefficient = −0.20), and to a lesser extent of the SE SAT index on Niño-3.4 (correlation coefficient = 0.01), removing this dependency via linear regression analysis has virtually no effect on the results (not shown), underscoring that differences among the 2000 individual bootstrapped composites are unlikely to be the result of sampling slightly different sets of ENSO events.

To extend this analysis to all of North America, we calculate the contribution to the observed CI that arises from the linear dependence of the SAT composite values at each grid box upon the Niño-3.4 composite values across the 2000 bootstrapped samples. To obtain this "ENSO contribution," we first compute the CIs using the 2000 SAT values of the bootstrapped composites

from which the composite Niño-3.4 SST index has been linearly removed via regression analysis, and then subtract it from the original CIs. This ENSO contribution is <0.5°C at all grid boxes, corresponding to <5% of the total CI, except near the Great Lakes and a few isolated locations where it can reach up to 5%–15% (Fig. S9).

Repeating these analyses for PR, Fig. 9 shows the observed bootstrapped ENSO composites that lie at the 10th and 90th percentiles based on the PNW, GULF, and CA PR indices defined above. All six composites show similar spatial patterns, consisting of PR increases over the Gulf states and along the Pacific coast of Canada and Alaska, and PR decreases over the Pacific Northwest and Ohio Valley–Upper South, similar to the ensemble-mean composite (recall Fig. 4a). However, their magnitudes and areas of statistical significance vary considerably. For example, drying over the U.S. Northwest and Upper South is greater in amplitude and area of significance in the 10th-percentile PNW composite compared to its 90th-percentile counterpart (Figs. 9a and 9b, respectively). Similarly, PR increases over the southern United States are larger and extend farther north in the
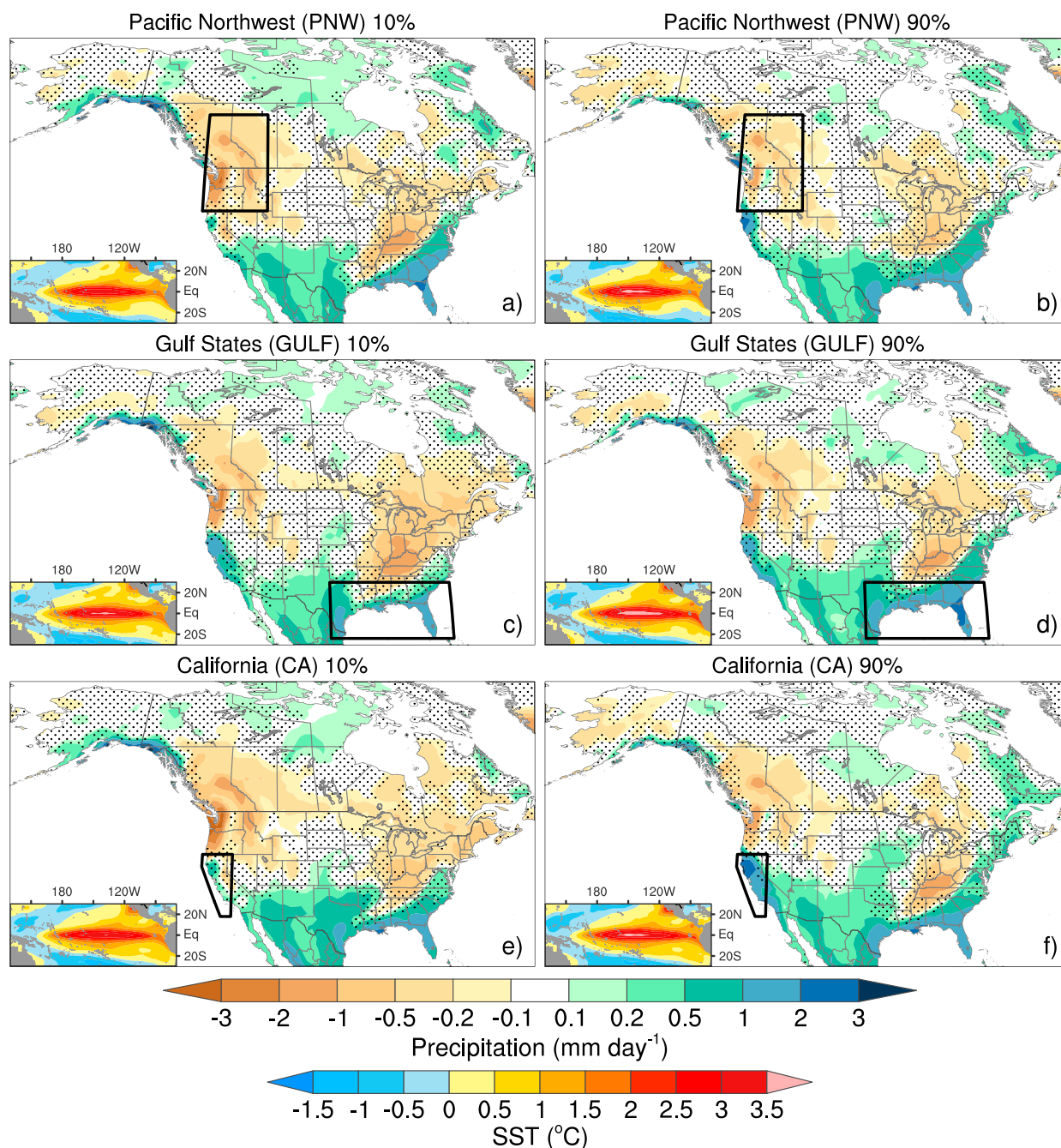
Observational Bootstrapped ENSO Composites

FIG. 9. As in Fig. 7, but for observed PR (mm day^{-1}). Tenth- and 90th-percentile composites based on the (a),(b) PNW, (c),(d) GULF, and (e),(f) CA PR indices (index regions outlined in black).

90th-percentile GULF composite compared to the 10th-percentile composite (Figs. 9d and 9c, respectively). Finally, the only case with significant wetting over all of California is the one based on the 90th percentile of the CA PR index (Fig. 9f). The 10th–90th percentile PR range within each index region is −0.60 to −0.24 mm day^{-1}

for PNW, 0.62 to 1.13 mm day^{-1} for GULF, and 0.03 to 1.43 mm day^{-1} for CA.

Similar tropical SSTA patterns accompany each PR composite, with small variations in amplitude (panel insets in Fig. 9). The 10th-percentile PNW composite (which has larger drying over the Pacific Northwest)
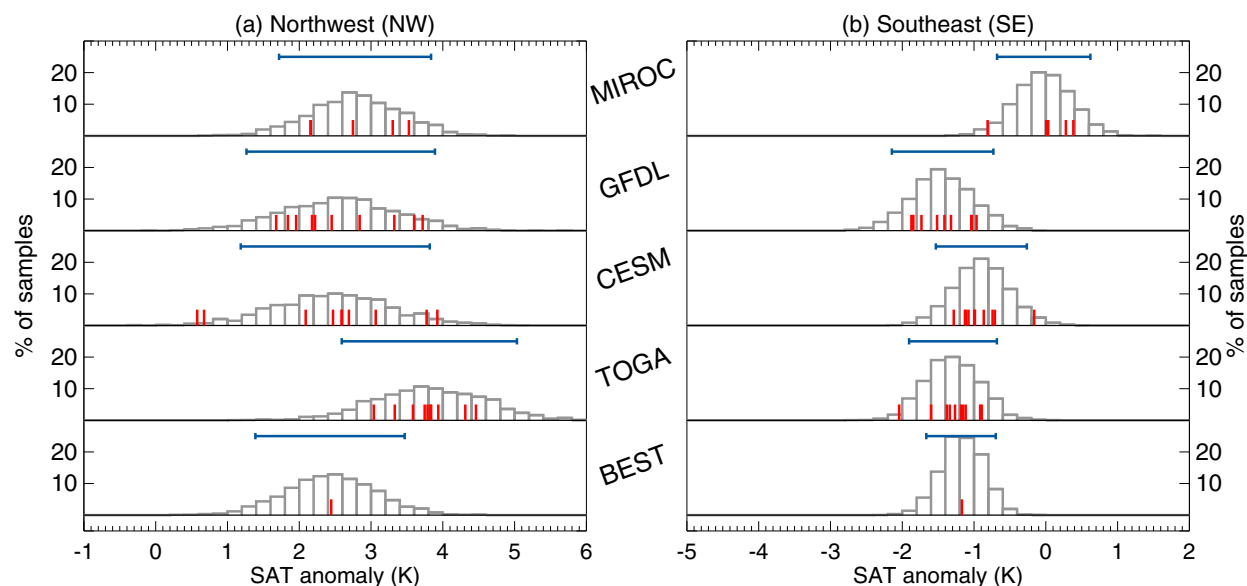
FIG. 10. Histograms of ENSO composite values of the (a) NW and (b) SE SAT indices (°C). Gray bars denote values from the 2000 bootstrapped samples and red bars indicate actual values. From top to bottom: PACE simulations from MIROC (5 members), GFDL CM2.1 (10 members) and CESM1 (10 members), and CAM5 TOGA simulations (10 members). Horizontal blue bar above each histogram indicates the 5%–95% CI range based on 2000 bootstrapped samples.

shows slightly weaker SSTA in the central equatorial Pacific (160°–130°W), compared to its 90th-percentile counterpart (Figs. 9a and 9b, respectively). However, this is likely a result of random chance since the PNW index shows no systematic dependence on the Niño-3.4 SST index across the 2000 bootstrapped composite samples (correlation coefficient = 0.01; Fig. 8c). The GULF and CA PR indices show slightly stronger dependencies on Niño-3.4 SST (correlation coefficients of 0.51 and 0.30, respectively), confirming the visual impression from the scatterplots (Figs. 8d,e) and consistent with the slightly larger SSTA in the central equatorial Pacific in the 90th-percentile composite sample (Figs. 9d and 9f, respectively) compared to the 10th-percentile sample (Figs. 9c and 9e, respectively). However, the magnitude of the ENSO contribution to the CIs on the observed PR composite is <10% over the Gulf states and California, and <5% everywhere else except southern Florida, where it reaches 20%–25% (Fig. S9).

Collectively, these results demonstrate that while there is some effect associated with sampling different sets of EN and LN events in the observed bootstrapped composites, it does not make a large contribution to the uncertainty in the SAT and PR ENSO composites, with some regional exceptions as noted above. Thus, the diversity of amplitudes, patterns, and degree of statistical significance among the ENSO composites shown in Figs. 5–9 is primarily due to internal variability rather than slightly different samples of ENSO events. In this

context, it is worth recalling the diversity in SAT and PR composites across the individual members of the PACE ensembles for which the set of ENSO events is identical.

d. Comparison across models

We summarize the amplitudes of the regional SAT and PR indices across all 2000 bootstrapped ENSO composites from observations and models in the histograms shown in Figs. 10 and 11. The actual composite values are shown as red bars: one for observations; 10 for each member of the CESM1 PACE, CESM1 TOGA, and CM2.1 PACE ensembles; and five for each member of the MIROC5 PACE ensemble. While the observed value must lie in the middle of its bootstrapped samples by construction, this need not be the case for the models since their bootstrapped samples were constructed by drawing from among all ensemble members (although the average across all members will lie at the peak of the distribution of the bootstrapped samples for a given model). The horizontal blue bar above each dataset indicates the 5%–95% CI range based on the bootstrapped samples.

As expected based on the results already presented, the CESM1 PACE ensemble shows a realistic mean value of the NW SAT index but slightly overestimates the width of its distribution (Fig. 10a). Also consistent with Fig. 1, two of the CESM1 PACE ensemble members are obvious outliers, falling in the lowest 1% of the distribution (by chance). The width and mean value of the NW SAT histograms based on the CM2.1 and MIROC5 PACE
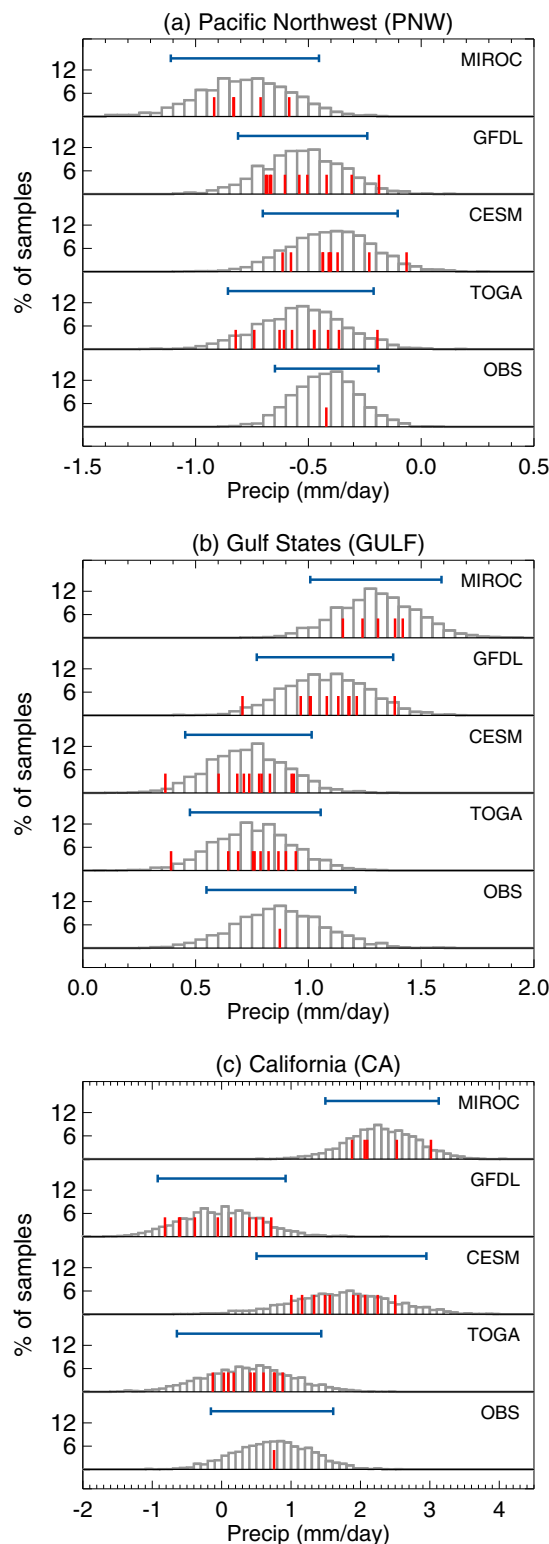
FIG. 11. As in Fig. 10, but for the (a) PNW, (b) GULF, and (c) CA PR indices (mm day^{-1}). Note the wider x-axis range in (c) compared to (a) and (b).

ensembles are similar to those from the CESM1 PACE ensemble, but the individual members are more evenly distributed across the range of bootstrapped samples than those in CESM1 (Fig. 10a). The NW SAT histogram based on the CESM1 TOGA ensemble is shifted to the right of those based on observations and PACE simulations, overestimating the observed value (2.4°C) by more than 50% in the ensemble mean (3.7°C). For the SE SAT index, the models generally show realistic distributions and CIs, except for MIROC, which simulates an ensemble-mean value close to zero that is significantly different from the observed value of −1.2°C (Fig. 10b). In all cases, the widths of the distributions are considerably smaller for the SE SAT index compared to the NW SAT index.

All model ensembles show realistic distributions for PR in the PNW region, although MIROC5 is on the drier side (Fig. 11a). There is more variation across models for PR in the GULF and CA regions, with all members of MIROC5 substantially overestimating the observed wetting, although there is overlap between the simulated and observed bootstrapped CIs (Figs. 11b,c). The most realistic distributions for GULF are those from CESM1 PACE and TOGA (Fig. 11b), while those for CA are CESM1 TOGA and CM2.1 (Fig. 11c). Like CESM1 PACE and TOGA, CM2.1 shows one member that falls at the very dry end of the distribution for GULF (Fig. 11b). The models generally simulate realistic CIs for all three PR indices, with the possible exception of CESM1 PACE for CA, which is considerably larger than observed, although there is some member-to-member variation (Fig. S8).

These portrayals of the bootstrapped ENSO composites for selected regional climate indices highlight the need for large model ensembles, since a single simulation can alter the mean value, and to a lesser extent the width, of the distribution just by chance, confounding model evaluation and model intercomparison.

e. El Niño versus La Niña composites

Up to now, we have focused on the linear component of ENSO. Here we examine whether there are any appreciable nonlinearities aside from polarity by comparing observed composites of the 18 EN events and the 14 LN events separately. For ease of comparison with EN, we show the LN composite with inverted sign (denoted −LN). While the SAT composites show regional differences in amplitude associated with a southward displacement of the continental-scale dipole pattern in EN (Fig. 12a) compared to −LN (Fig. 12b), these differences are not significant except near the Great Lakes (Fig. 12c). Similarly, the PR composites show local differences in magnitude associated with southward-shifted dipoles over the eastern United
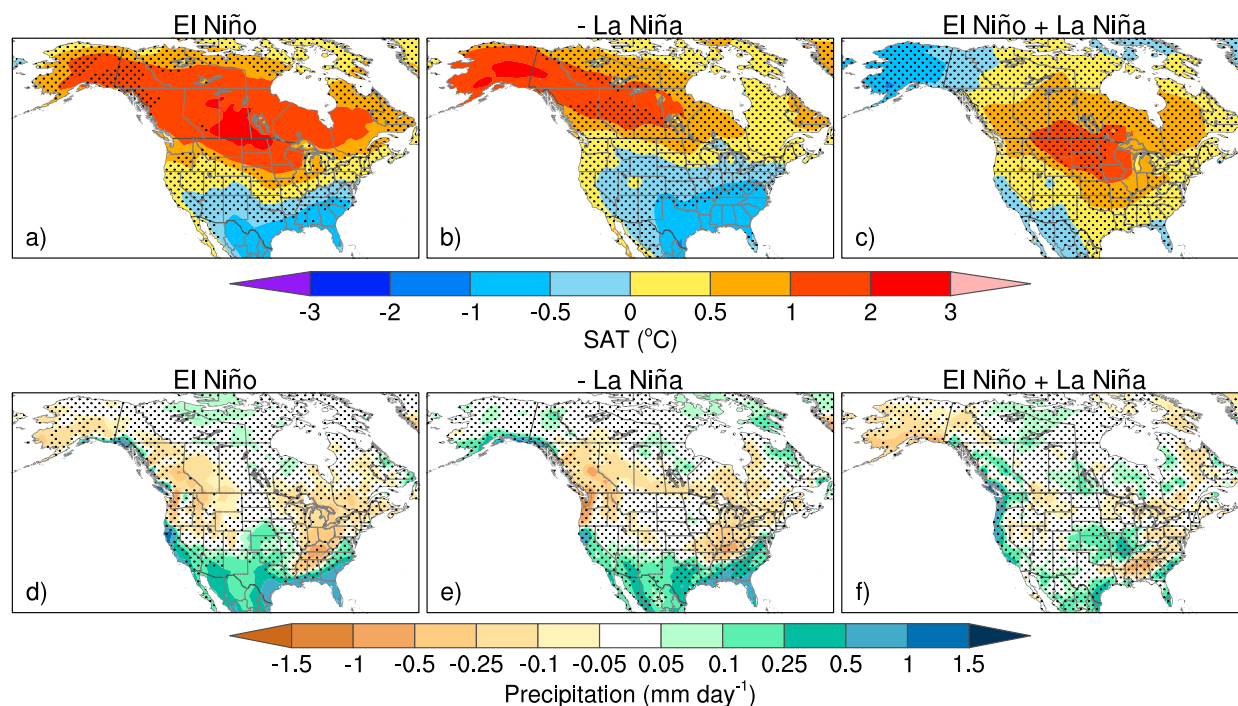
Observed ENSO Composites

FIG. 12. Observed DJF SAT (°C) and PR (mm day^{-1}) composites for (a),(d) 18 El Niño events, (b),(e) 14 La Niña events (with reversed sign), and (c),(f) their difference (El Niño + La Niña). Values not significant at the 10% confidence level based on a two-sided t test are stippled.

States and along the Gulf of Alaska in EN (Fig. 12d) compared to −LN (Fig. 12e), but none are significant except those along the southern coast of mainland Alaska (Fig. 12f). Our EN and LN composites are consistent with the analogous one-sided regression maps in Hoerling et al. (2001); however, that study did not assess differences between their one-sided regressions.

f. Flavors of El Niño

To what extent might different "flavors" of El Niño affect the range of SAT and PR anomalies across the 2000 observed bootstrapped composites? In particular, if we sample only east Pacific (EP) or central Pacific (CP) El Niño events [defined according to the consensus method of Yu et al. (2012); see also Graf and Zanchettin (2012) and Yu et al. (2015)] in our ENSO composites, do we obtain significantly different anomalies and CIs? To address this issue, we construct two additional 2000-member sets of bootstrapped composites, which differ from the original set by restricting the random sampling of all 18 El Niño events to those that fall in the EP category (7) and to those that fall in the CP category (11); nothing is changed for the sampling of La Niña events. Note that we maintain a total of 18 El Niño (and 14 La Niña) events in these new CP and EP sets of

bootstrapped composites for consistency with the original "all El Niño" bootstrapped composites.

The CI maps based on the 2000 CP and 2000 EP bootstrapped composites are very similar in pattern and amplitude for both SAT (Figs. 13a,b) and PR (Figs. 13c,d). The slightly larger SAT CI values over Canada in CP compared to EP are well within the range of what could be expected by chance, at least according to the individual members of CESM1 PACE, whose CIs are based on the same set of ENSO events (Fig. S7). Similar results are found by sampling only east Pacific nonconvecting (EPN) or east Pacific convecting (EPC) El Niño events (Johnson and Kosaka 2016) in our 2000 observed bootstrapped ENSO composites (not shown). Taken together, the results shown above reinforce the notion that ENSO diversity, whether in the form of differences in magnitude, nonlinearities between EN and LN, or different "flavors" of El Niño, does not have an appreciable effect on our quantification of uncertainty on the observed ENSO SAT and PR composites.

A separate but related question is whether the actual observed CP and EP composites show significantly different SAT and PR anomalies. Both composites display the familiar SAT dipole pattern across North America, but EP exhibits larger statistically significant warming
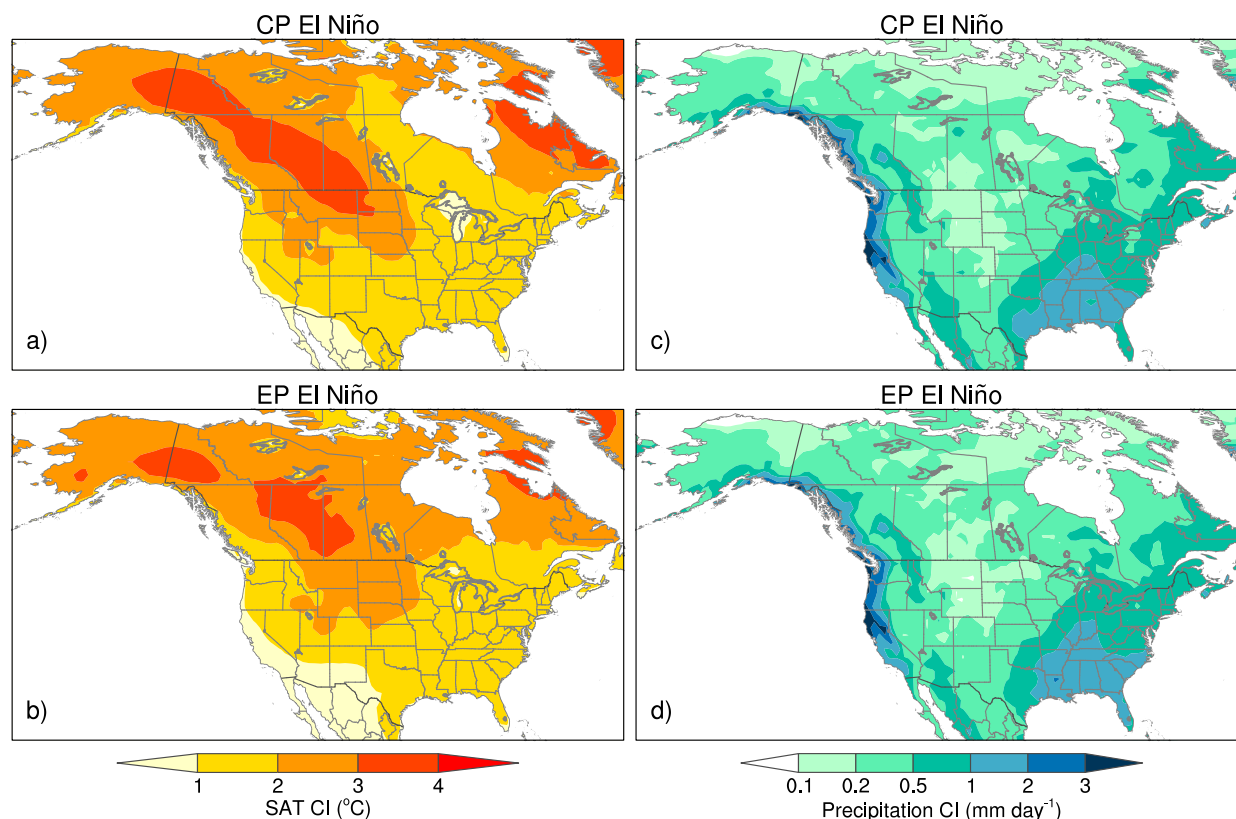
Observed ENSO Composites (CI)

FIG. 13. CIs of 5%–95% on observed ENSO composites of DJF (a),(b) SAT (°C) and (c),(d) PR (mm day^{-1}) using only (a),(c) CP and (b),(d) EP El Niño events during 1920–2013. See text for details.

over Canada and Alaska compared to CP (maximum values 4°–5°C vs 2°–3°C, respectively; Figs. 14a,b). In addition, the region of significant cooling is confined to the southern tier of U.S. states in CP but penetrates into the Mid-Atlantic states in EP. The area extending southeastward from northern Saskatchewan to the U.S. central Atlantic coast shows statistically significant differences between EP and CP, but the rest of Canada does not, despite the nearly twofold difference in composite SAT amplitudes (Fig. 14c). PR shows a very similar pattern between the two sets of composites, with somewhat larger amplitudes for EP compared to CP (Figs. 14e and 14d, respectively), but these differences are not statistically significant except at a few locations (Fig. 14f). SAT and PR differences between EPN and EPC composites are also generally not statistically significant over North America, as shown by Johnson and Kosaka (2016).

g. Late-winter ENSO composites

While the primary focus of this study is on DJF, we briefly report on FMA, as rainfall over Southern California shows a larger ENSO signal in this season (L'Heureux et al.

2015; Jong et al. 2016). Repeating our observational ENSO composites for late winter (shown in Fig. S10), we confirm that in addition to Southern California, positive PR anomalies occur with larger amplitudes (statistically significant values of 0.2–0.5 mm day^{-1}) in FMA compared to DJF over the U.S. Southwest desert and portions of Kansas, Nebraska, and South Dakota (cf. Fig. S10b with Fig. 4b). The increased PR in FMA is accompanied by stronger cooling (statistically significant amplitudes of 1°–2°C) over Arizona, New Mexico, and parts of Colorado and Kansas (Fig. S10a) compared to DJF (Fig. 3b). Elsewhere, ENSO composite values are considerably weaker in late winter than in midwinter, both for PR and SAT. For example, warming over Canada and Alaska is only half as strong in FMA compared to DJF, and drying over the interior Pacific Northwest is weak and insignificant in late winter.

4. Discussion and summary

This study has evaluated the role of sampling variability in ENSO composites of winter SAT and PR over
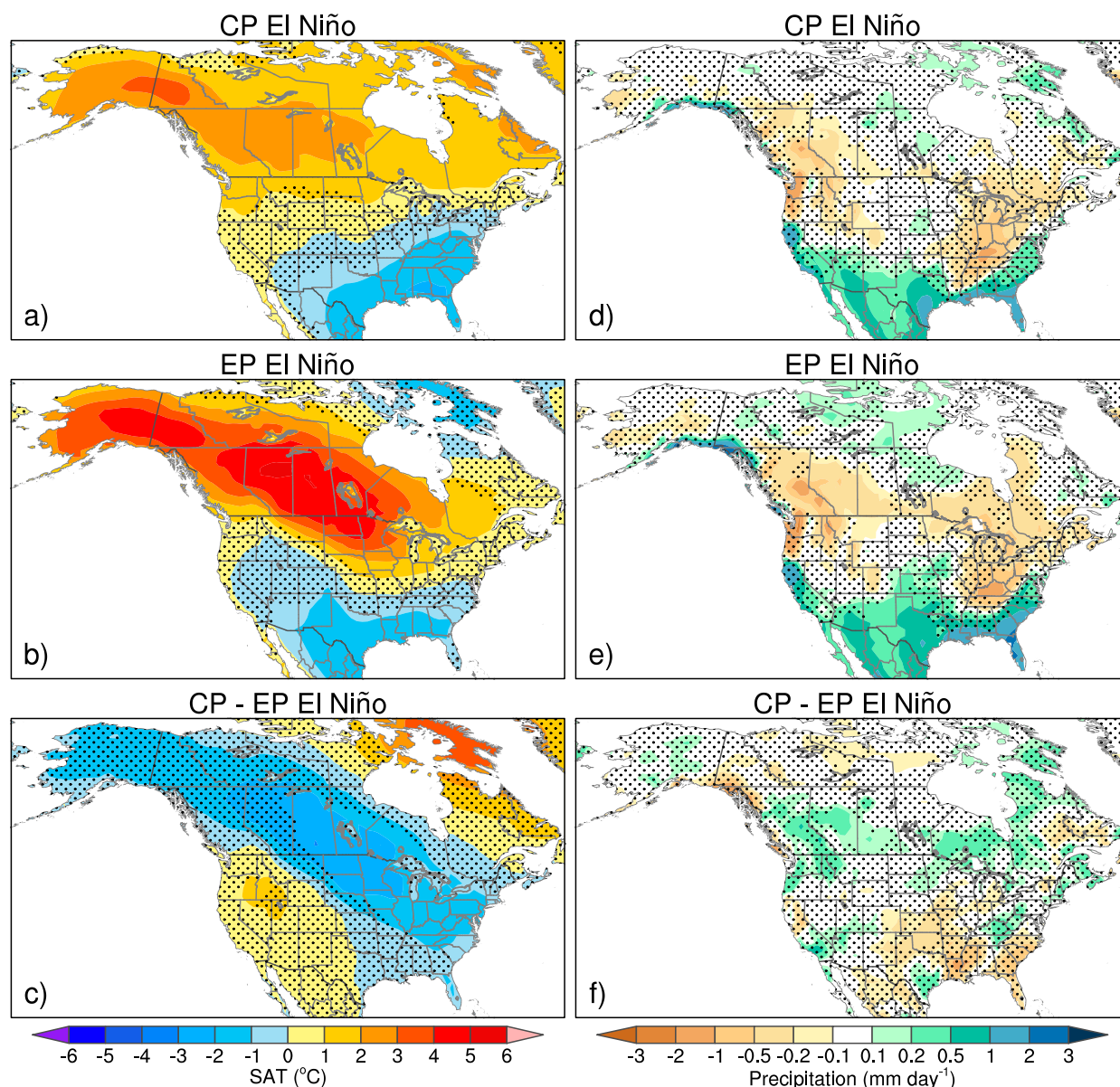
Observed ENSO Composites

FIG. 14. Observed ENSO composites of DJF SAT (°C) and PR (mm day^{-1}) using only (a),(d) CP and (b),(d) EP El Niño events during 1920–2013, and (c),(f) their difference (CP–EP). Values not significant at the 10% confidence level based on a two-sided t test are stippled.

North America during the period 1920–2013 in observations and ensembles of "Tropical Pacific Pacemaker" coupled model simulations with CESM1, CM2.1, and MIROC5. The individual members of each model ensemble show a surprising amount of diversity in their ENSO composites, despite the fact that they are constructed from the same observed set of 18 EN and 14 LN events. For a given model, this ensemble spread can only be due to sampling variability, that is, aliasing of internal

variability that is unrelated to ENSO. In the case of CESM1, for which a lengthy atmosphere-only control simulation is available, we showed that this sampling variability arises from internal atmospheric dynamics rather than coupled ocean–atmosphere processes. Similar ENSO composite spread is evident in an uncoupled (atmosphere-only) model ensemble with observed time-varying tropical SSTs prescribed at the lower boundary (the CESM1 TOGA ensemble).

Are the observed ENSO composites subject to a similar level of uncertainty as those in the Pacemaker ensembles? What might the observed ENSO composite have looked like under a different permutation of natural variability unrelated to ENSO? To address these questions, we constructed 2000 synthetic ENSO composites from the observations using random sampling techniques. These synthetic composites provide information on the range of spatial patterns and amplitudes associated with imperfect estimation of the forced ENSO signal. The observed SAT composite shows a statistically significant dipole pattern of positive anomalies over western Canada and Alaska and negative anomalies over the southeastern United States. But although all 2000 synthetic ENSO composites show positive SAT values in the NW and negative SAT values in the SE, their amplitudes vary by approximately a factor of 2.5 between the 5th- and 95th-percentile composite samples (1.3° to 3.4°C for the NW and −0.7° to −1.7°C for the SE). The observed PR composite shows significant wetting over the Gulf states and parts of California, Arizona, and New Mexico, and significant drying over the Ohio Valley–Upper South and parts of the interior Pacific Northwest. However, the 5%–95% uncertainty range on the magnitudes of these regional composite PR anomalies is substantial: $0.55 \, \text{mm day}^{-1}$ to $0.89 \, \text{mm day}^{-1}$ in the GULF; $-0.15 \, \text{mm day}^{-1}$ to $-0.69 \, \text{mm day}^{-1}$ in the PNW (and also the Ohio Valley–Upper South); and $-0.07 \, \text{mm day}^{-1}$ to $+1.18 \, \text{mm day}^{-1}$ in CA. While previous studies highlight that the strong EN events of 1957/58, 1982/83, and 1997/98 each brought copious amounts of rainfall to CA (Siler et al. 2017; Lee et al. 2018), our results are not unduly influenced by the number of times these events are sampled in our synthetic composites. In particular, these three events account for <16% of the ENSO events sampled in 89% of the synthetic ENSO composites, consistent with the results shown in Figs. 8, 9, and S9, and make up 9% and 12.5% of the events sampled in the 10th- and 90th-percentile PR composites based on CA PR, respectively (Figs. 9e,f).

Although the synthetic ENSO composites based on observations are necessarily constructed from different combinations of EN and LN events, differences in magnitude of the composite Niño-3.4 SST index make only a minor (<5%) contribution to their spread over most of North America, with slightly higher values (up to 10%) for SAT near the Great Lakes and PR over California and portions of the SE United States, and up to 25% for PR over central Florida. Removing this dependence on the composite Niño-3.4 values results in a slight (i.e., on the order of a few grid boxes) expansion of the regions covered by robust ENSO signals, but does not quantitatively affect the results (Fig. S11). Other forms of ENSO

diversity, such as nonlinearities between EN and LN or different "flavors" of El Niño (EP vs CP), also do not appreciably affect our quantification of uncertainty on the observed ENSO SAT and PR composites.

Our results have implications for ENSO reconstructions based on paleoclimate proxy records of SAT and PR over North America. In particular, such ENSO reconstructions will also be subject to uncertainties associated with sampling variability, even if the proxies are perfect indicators of winter climate anomalies. Judicious choices of proxy record locations based on the uncertainties provided here may help to narrow this range; another ameliorating factor may be if the proxy records integrate climate signals over a broader seasonal window, as this may help to reduce aliasing from natural variability unrelated to ENSO (although it may also weaken the ENSO signal).

Our results have broad implications for how to evaluate the realism of ENSO signals in models. In particular, uncertainty in the pattern and amplitude of the observed ENSO composite necessitates an approach to model assessment that considers not only the model's forced response to ENSO, but also its representation of internal variability unrelated to ENSO. In the Pacemaker ensembles, we can determine the forced response by averaging ENSO composites across all members of a given model. Using the 2000 synthetic ENSO composites constructed for each model simulation and for observations, we can discriminate between true model biases in the forced ENSO response and apparent model biases that arise from limited sampling of internal variability unrelated to ENSO.

Applying this approach to the CESM1 Pacemaker ensemble, we find that the model significantly overestimates internal variability (and hence ENSO composite spread) of SAT over Alaska and parts of the eastern and southwestern United States, and also significantly overestimates (underestimates) internal variability of PR over the western (southeastern) United States. Taking these differences in internal variability into account, we are able to reveal true biases in the model's forced ENSO response, including a significant underestimation of warming over the central Canadian provinces and U.S. Upper Midwest, a significant underestimation of wetting (drying) over Florida (Ohio Valley–Upper South), and a significant overestimation of wetting (drying) over California and Nevada (coastal British Columbia). Somewhat different model biases in the forced ENSO response are apparent in the uncoupled CESM1 TOGA ensemble for reasons discussed in the appendix. Observational uncertainty in tropical SSTs used as boundary forcing for the models represents an additional potential source of discrepancy between
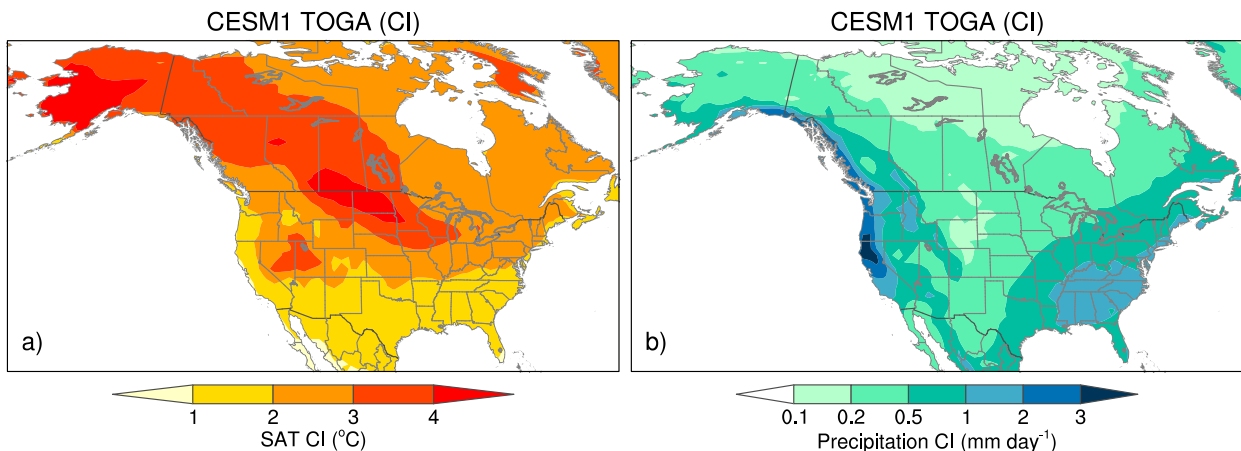
CESM1 TOGA (CI)

CESM1 TOGA (CI)

FIG. A1. CIs of 5%–95% on ENSO composites of DJF (a) SAT (°C) and (b) PR (mm day^{-1}) from the CESM1 TOGA ensemble based on 2000 bootstrapped samples.

the observed and simulated ENSO composites, and merits investigation.

In summary, even with nearly a century of observations, quantification of the canonical influence of ENSO on North American climate is subject to considerable uncertainty due to aliasing of unrelated climate variability. This observational uncertainty must be properly accounted for when evaluating ENSO responses in climate models. In particular, discriminating between true model biases in the forced response to ENSO, and apparent model biases that arise from limited sampling of internal variability unrelated to ENSO, is essential.

ENSO Composites
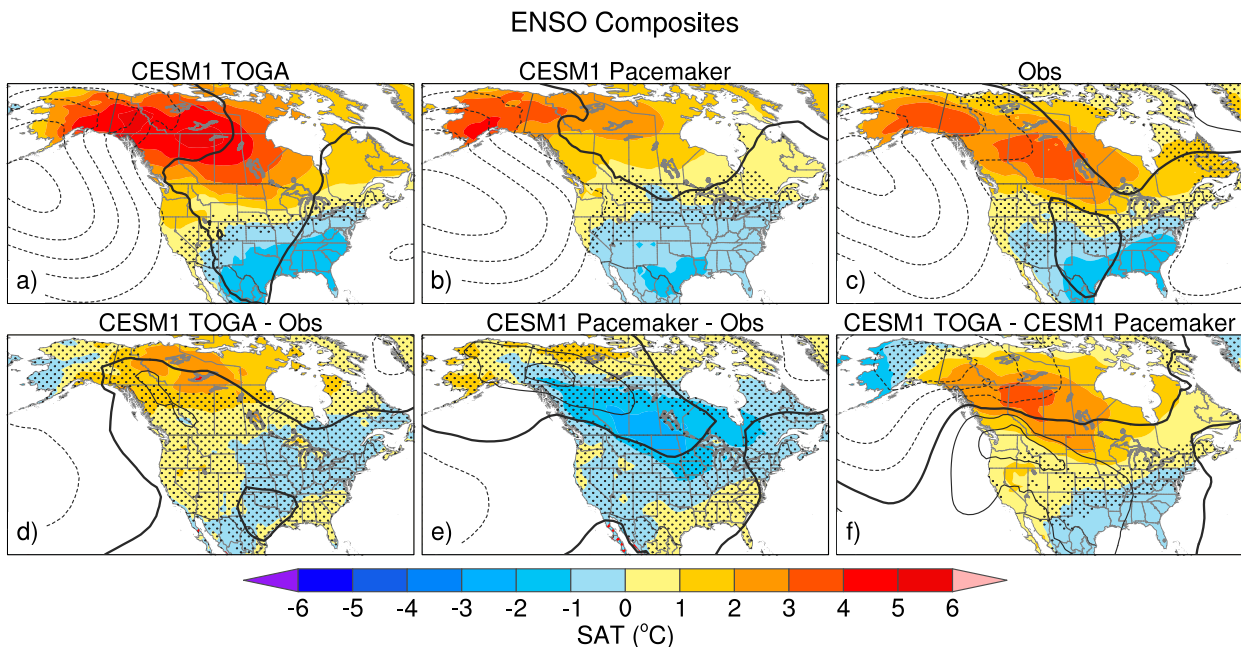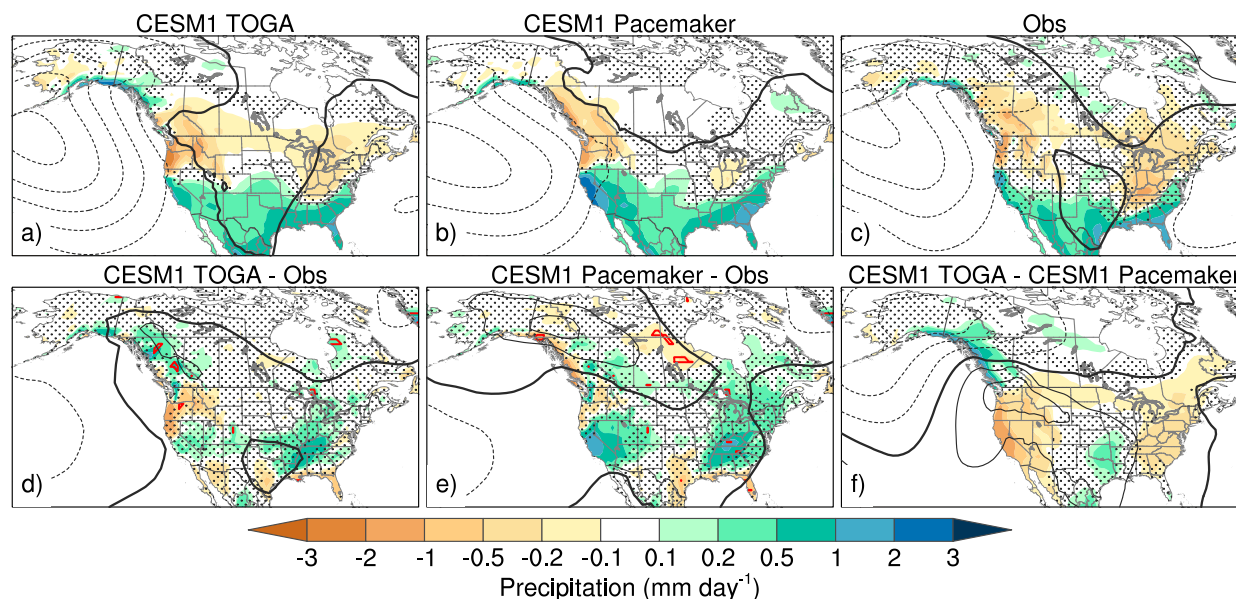

FIG. A2. ENSO composites of DJF SAT (°C) and SLP (hPa; contours) for the (a) CESM1 TOGA ensemble mean, (b) CESM1 PACE ensemble mean, (c) observations, and differences between (d) TOGA ensemble mean and observations, (e) PACE ensemble mean and observations, and (f) TOGA and PACE ensemble means. Contour interval is 2 hPa, with negative values dashed and the zero contour thickened. Stippling in (a),(b),(c),(f) indicates SAT values not significant at the 10% confidence level based on a two-sided *t* test. Stippling in (d),(e) indicates that the observed value lies outside of the 5%–95% values from the model's 2000 bootstrapped composites, and red contours indicate the observed value lies outside any of the model's bootstrapped composites.

ENSO Composites

FIG. A3. As in Fig. A2, but for PR (mm day^{-1}).

APPENDIX

CESM1 Pacemaker versus TOGA Simulations

Are there any systematic differences in ENSO composites from the coupled (PACE) and uncoupled (TOGA) simulations with CESM1, and if so, why? The spread across the 2000 bootstrapped ENSO composites is very similar between TOGA and PACE, for both SAT (cf. Fig. A1a and Fig. 3d) and PR (cf. Fig. A1b and Fig. 4d), consistent with the dominant role of internal atmospheric variability on ENSO composite uncertainty discussed in section 3b. However, the ensemble-mean ENSO composite (i.e., the forced response to ENSO) differs somewhat between TOGA and PACE. In particular, the ensemble-mean SAT composite shows oppositely signed biases, with TOGA significantly overestimating the

observed warming over northern Canada by up to 3°C, and PACE significantly underestimating it over south-central Canada by about the same amount (Fig. A2). Thus, neither model configuration is clearly superior in terms of SAT amplitude, although the spatial pattern is more realistic in TOGA than PACE [the pattern correlation (r) of the ensemble-mean SAT composite against the observed SAT composite is 0.91 in TOGA compared to 0.75 in PACE, and the lowest r of any of the individual TOGA ensemble members (0.82) exceeds the highest r from any of the individual PACE ensemble members (0.81)]. The TOGA ensemble-mean PR composite shows realistic magnitudes of wetting over Southern California and Nevada and drying over coastal British Columbia, areas where PACE was biased high; however, TOGA overestimates the drying over parts of Oregon, Washington, and Montana and underestimates it in interior British Columbia, areas where PACE was not significantly biased (Fig. A3). While the spatial pattern of the TOGA ensemble-mean PR composite bears a closer resemblance to the observed PR composite than does PACE (r = 0.76 for TOGA and 0.63 for PACE), there is overlap between the lowest pattern correlation in TOGA (0.64) and the highest in PACE (0.68) across the individual ensemble members.

In summary, there are systematic differences in the ENSO-forced SAT and PR responses between the TOGA and PACE configurations of CESM1, with TOGA showing an improved representation of the spatial pattern but not amplitude of SAT, and of the PR
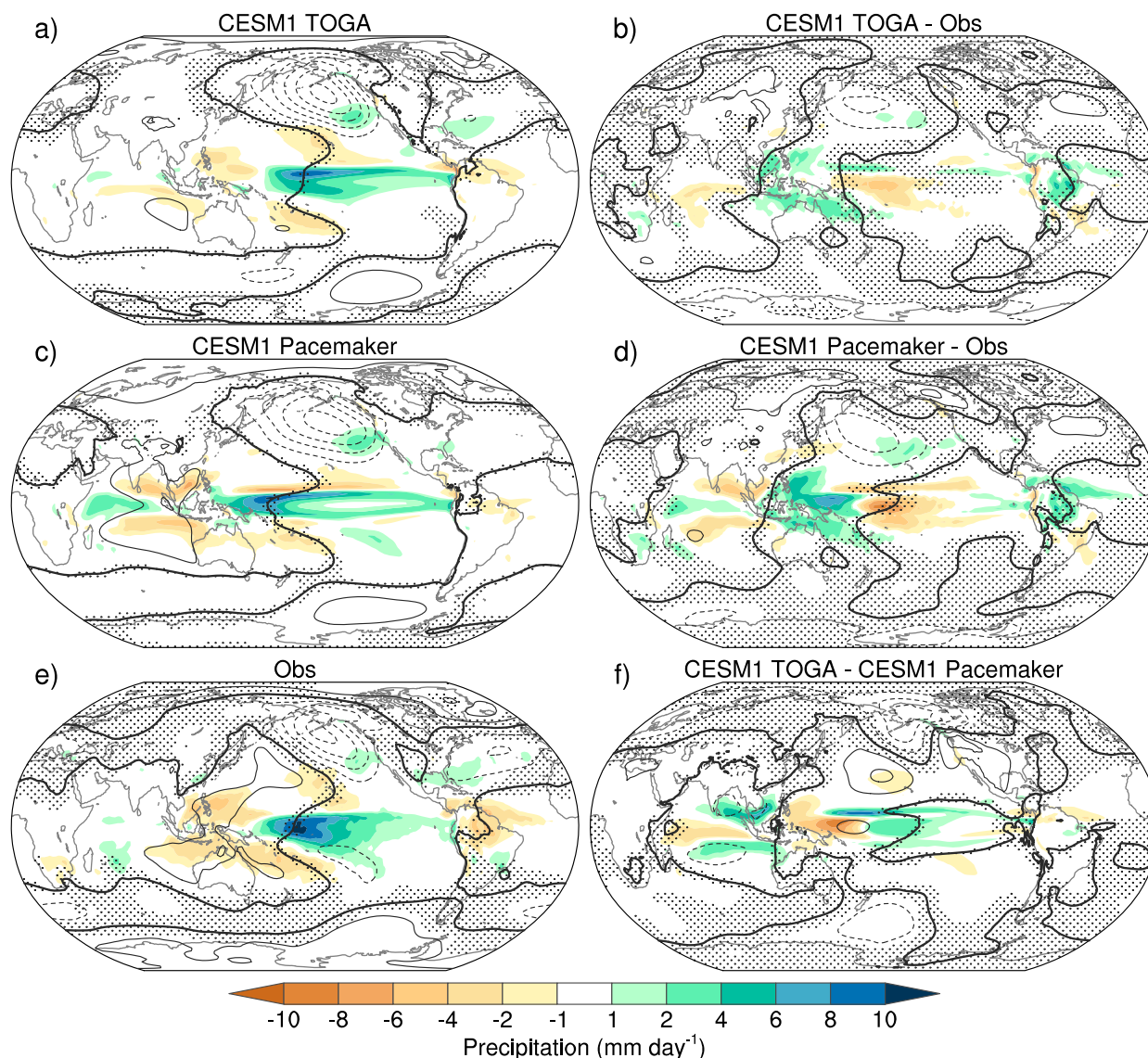
ENSO Composites

FIG. A4. ENSO composites of DJF PR (mm day^{-1}) and SLP (hPa; contours) for the (a) CESM1 TOGA ensemble mean, (c) CESM1 PACE ensemble mean, (e) observations (20CR), and differences between (b) TOGA ensemble mean and observations, (d) PACE ensemble mean and observations, and (f) TOGA and PACE ensemble means. Contour interval is 2 hPa, with negative values dashed and the zero contour thickened. Stippling in (a),(c),(e),(f) indicates SLP values not significant at the 10% confidence level based on a two-sided t test. Stippling in (b),(d) indicates that the observed SLP value lies outside of the 5%–95% values from the model's 2000 bootstrapped composites.

magnitudes over Southern California and Nevada and coastal British Columbia, compared to PACE. A 10-member CESM1 ensemble with specified observed time-evolving SSTs (and sea ice) over the entire globe yields virtually identical results to TOGA (not shown); thus, the differences between PACE and TOGA are unlikely to result from ENSO-related SST anomalies in the extratropics.

What is the origin of the systematic differences in ENSO responses between TOGA and PACE? To address this question, it is helpful to view the surface climate impacts of ENSO within the context of the large-scale atmospheric circulation that drives them. The ensemble-mean ENSO composites of SLP from TOGA and PACE show negative anomalies over the North Pacific, with maximum values of 8–10 hPa near
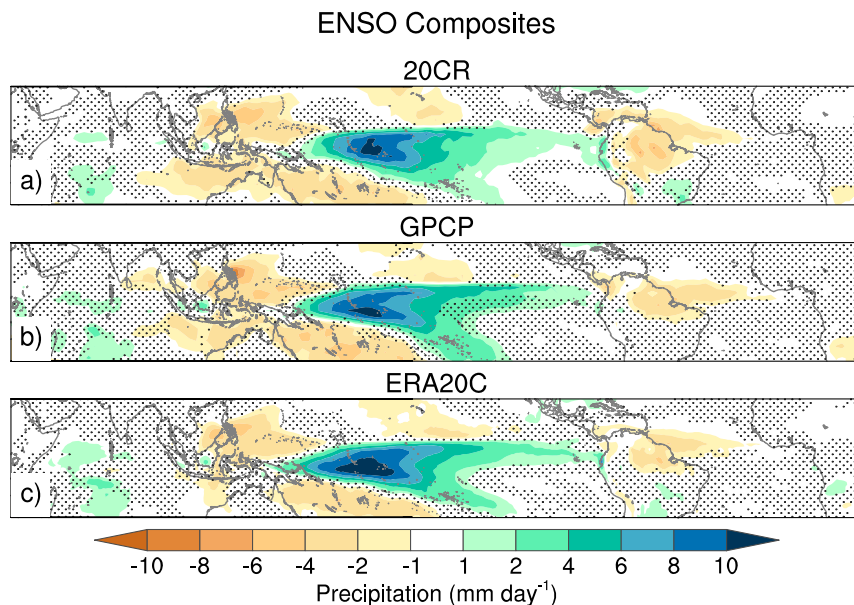
ENSO Composites

FIG. A5. ENSO composites of DJF PR (mm day^{-1}) for (a) 20CR, (b) GPCP, and (c) ERA20C. Stippling indicates values not significant at the 10% confidence level based on a two-sided t test. Composites based on 1920–2013 for (a),(c), and 1979–2013 for (b).

the Aleutian Islands, similar to observations (Figs. A2 and A3; see also D17). However, the orientation of the isobars is more zonal in PACE compared to the NW–SE tilt evident in observations and TOGA. The SLP difference between TOGA and PACE indicates onshore flow of mild maritime air into western Canada, which may account for the greater warming (Fig. A2f) and wetting (Fig. A3f) in this region in TOGA relative to PACE. Farther south, the offshore flow component in TOGA compared to PACE is likely responsible for the reduced wetting over California and neighboring states (Fig. A3f).

The circulation differences in TOGA and PACE, in turn, may be linked to differences in their tropical PR responses via Rossby wave dynamics as shown in Fig. A4. The tropical PR response in TOGA shows wetting over the central Pacific and drying over the far western Pacific, similar to observations but with reduced amplitude (Figs. A4a,e). Here, the observed tropical PR composite is based on 20CR, but similar results are found using GPCP (limited to the satellite period starting in 1979) and ERA20C (Fig. A5). In PACE, this entire pattern is shifted to the west and more equatorially confined, with maximum wetting over the western equatorial Pacific and maximum off-equatorial drying over the eastern Indian Ocean (Fig. A4c). This westward displacement likely reflects the influence of mean-state biases in the fully coupled CESM1, in particular a westward-extended equatorial SST "cold tongue" that anchors a narrow "double ITCZ" on either side of the equator (not shown; similar

mean-state biases are found in CM2.1 and MIROC5). Thus, the PACE protocol is not a panacea because of the influence of mean-state biases on CESM1's response to observed SSTA. The difference in tropical PR responses between TOGA and PACE shows negative values in the western Pacific of up to -10 mm day^{-1} and weaker-amplitude positive values in the central Pacific (Fig. A4f). This is accompanied by an arching SLP wave train over the North Pacific that appears to emanate from the negative precipitation center in the western tropical Pacific, possibly indicative of a "short wavelength" Rossby wave response. This circulation response, in turn, drives the surface climate response differences noted above.

REFERENCES

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231, https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172, https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

Capotondi, A., and Coauthors, 2015: Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, **96**, 921–938, https://doi.org/10.1175/BAMS-D-13-00117.1.

Chiodi, A. M., and D. E. Harrison, 2013: El Niño impacts on seasonal U.S. atmospheric circulation, temperature, and

precipitation anomalies: The OLR-event perspective. *J. Climate*, **26**, 822–837, https://doi.org/10.1175/JCLI-D-12-00097.1.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https://doi.org/10.1002/qj.776.

Deser, C., and J. M. Wallace, 1987: El Niño events and their relationship to the Southern Oscillation: 1925–1986. *J. Geophys. Res.*, **92**, 14 189–14 196, https://doi.org/10.1029/JC092iC13p14189.

——, I. R. Simpson, K. A. McKinnon, and A. S. Phillips, 2017: The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *J. Climate*, **30**, 5059–5082, https://doi.org/10.1175/JCLI-D-16-0844.1.

Garfinkel, C. I., M. M. Hurwitz, D. W. Waugh, and A. H. Butler, 2013: Are the teleconnections of central Pacific and eastern Pacific El Niño distinct in boreal wintertime? *Climate Dyn.*, **41**, 1835–1852, https://doi.org/10.1007/s00382-012-1570-2.

Graf, H.-F., and D. Zanchettin, 2012: Central Pacific El Niño, the "subtropical bridge," and Eurasian climate. *J. Geophys. Res.*, **117**, D01102, https://doi.org/10.1029/2011JD016493.

Held, I. M., S. W. Lyons, and S. Nigam, 1989: Transients and the extratropical response to El Niño. *J. Atmos. Sci.*, **46**, 163–174, https://doi.org/10.1175/1520-0469(1989)046<0163:TATERT>2.0.CO;2.

Hoerling, M. P., A. Kumar, and T. Xu, 2001: Robustness of the nonlinear climate response to ENSO's extreme phases. *J. Climate*, **14**, 1277–1293, https://doi.org/10.1175/1520-0442(2001)014<1277:ROTNCR>2.0.CO;2.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–829, https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2.

Johnson, N. C., and Y. Kosaka, 2016: The impact of eastern equatorial Pacific convection on the diversity of boreal winter El Niño teleconnection patterns. *Climate Dyn.*, **47**, 3737–3765, https://doi.org/10.1007/s00382-016-3039-1.

Jong, B.-T., M. Ting, and R. Seager, 2016: El Niño's impact on California precipitation: Seasonality, regionality, and El Niño intensity. *Environ. Res. Lett.*, **11**, 054021, https://doi.org/10.1088/1748-9326/11/5/054021.

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.

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

Larkin, N. K., and D. E. Harrison, 2005: Global seasonal temperature and precipitation anomalies during El Niño autumn and winter. *Geophys. Res. Lett.*, **32**, L16705, https://doi.org/10.1029/2005GL022860.

Lee, S.-K., H. Lopez, E.-S. Chung, P. DiNezio, S.-W. Yeh, and A. T. Wittenberg, 2018: On the fragile relationship between El Niño and California rainfall. *Geophys. Res. Lett.*, **45**, 907–915, https://doi.org/10.1002/2017GL076197.

L'Heureux, M. L., M. K. Tippett, and A. G. Barnston, 2015: Characterizing ENSO coupled variability and its impact on North American seasonal precipitation and temperature. *J. Climate*, **28**, 4231–4245, https://doi.org/10.1175/JCLI-D-14-00508.1.

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384, https://doi.org/10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2.

Rohde, R., and Coauthors, 2013: A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinf. Geostat. Overview*, **1**, https://doi.org/10.4172/2327-4581.1000101.

Ropelewski, C .F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, https://doi.org/10.1007/s00704-013-0860-x.

Shukla, J., and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606, https://doi.org/10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2.

Siler, N., Y. Kosaka, S.-P. Xie, and X. Li, 2017: Tropical ocean contributions to California's surprisingly dry El Niño of 2015/16. *J. Climate*, **30**, 10 067–10 079, https://doi.org/10.1175/JCLI-D-17-0177.1.

Takahashi, K., and A. G. Martínez, 2018: The very strong coastal El Niño in 1925 in the far-eastern Pacific. *Climate Dyn.*, https://doi.org/10.1007/s00382-017-3702-1, in press.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tippett, M. K., A. G. Barnston, and S. Li, 2012: Performance of recent multimodel ENSO forecasts. *J. Appl. Meteor. Climatol.*, **51**, 637–654, https://doi.org/10.1175/JAMC-D-11-093.1.

Yu, B., X. Zhang, H. Lin, and J.-Y. Yu, 2015: Comparison of wintertime North American climate impacts associated with multiple ENSO indices. *Atmos.–Ocean*, **53**, 426–445, https://doi.org/10.1080/07055900.2015.1079697.

Yu, J.-Y., Y. Zou, S. T. Kim, and T. Lee, 2012: The changing impact of El Niño on US winter temperatures. *Geophys. Res. Lett.*, **39**, L15702, https://doi.org/10.1029/2012GL052483.