

# Linking Projected Changes in Seasonal Climate Predictability and ENSO Amplitude

DILLON J. AMAYA<sup>a</sup>, NICOLA MAHER<sup>b,c,d</sup>, CLARA DESER<sup>e</sup>, MICHAEL G. JACOX<sup>a,f</sup>, MICHAEL A. ALEXANDER<sup>a</sup>,  
MATTHEW NEWMAN<sup>a</sup>, JULIANA DIAS<sup>a</sup>, AND JIALE LOU<sup>g</sup>

<sup>a</sup> NOAA/Physical Sciences Laboratory, Boulder, Colorado

<sup>b</sup> Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder,  
University of Colorado Boulder Campus, Boulder, Colorado

<sup>c</sup> Department of Atmospheric and Oceanic Sciences (ATOC), University of Colorado Boulder, Boulder, Colorado

<sup>d</sup> The Australian National University, Canberra, Australian Capital Territory, Australia

<sup>e</sup> National Science Foundation National Center for Atmospheric Research, Boulder, Colorado

<sup>f</sup> NOAA/Environmental Research Division, Southwest Fisheries Science Center, Monterey, California

<sup>g</sup> Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey

(Manuscript received 26 October 2023, in final form 25 October 2024, accepted 22 November 2024)

**ABSTRACT:** Recent studies have shown that potential predictability and actual forecast skill have varied throughout the historical record, primarily due to natural decadal variability. In this study, we explore whether and how potential predictability is projected to change in the future as a distinct response to anthropogenic climate change. We estimate the potential predictability of El Niño–Southern Oscillation (ENSO) as well as global surface temperature, precipitation, and upper-atmospheric circulation anomalies from 1921 to 2100, within a perfect model framework, using five coupled model large ensembles. We find that historical and projected ENSO amplitude changes generate global-scale shifts in climate predictability via ENSO-driven changes in the signal-to-noise ratio of seasonal forecasts, with a 10% change in Niño-3.4 standard deviation leading to a 14% change in globally averaged forecast skill at 12-month lead. This relationship suggests that potential predictability changes across much of the globe in the coming decades could be linked to anthropogenic climate change of ENSO. However, since current models substantially disagree on the sign and intensity of projected ENSO change, the trajectory of future global predictability changes cannot yet be determined. This problem is demonstrated by widely varying predictability changes seen across the five large ensembles, with models exhibiting a robust increase, robust decrease, or no significant change in predictability, depending upon their respective projected ENSO amplitude trends. Our results highlight the need for climate model development aimed at better capturing past forced and unforced changes to ENSO variability, which is necessary (if not sufficient) to constrain projected changes to climate predictability worldwide.

**KEYWORDS:** Climate change; Climate prediction; ENSO; Seasonal forecasting; Climate models

## 1. Introduction

Recent advances in model physics, resolution, ensemble sizes, and data assimilation schemes have led to increases in seasonal forecast skill (Barnston et al. 2012; Barnston and Tippett 2017). However, prediction systems may still be limited by the so-called “potential predictability” of different climate states. Potential predictability is often thought of as a prediction limit intrinsic to the chaotic nature of the climate system (Sardeshmukh et al. 2000), the point when a forecast’s initial uncertainty grows until any climate state randomly drawn from the climatological probability distribution could be expected to occur (Lorenz 1969). In practice, however, predictability is generally estimated from models, but model predictability may differ from nature’s predictability due to model errors both in the distribution of climate states and their evolution. Systematic state-dependent seasonal forecast errors (e.g., Beverley et al. 2023) likewise suggest that some aspects of nature’s predictability

may be missed by all model predictability studies. Recent studies have even suggested that some models make forecasts in some regions with real-world skill that exceeds the model’s own predictability, possibly a consequence of the signal-to-noise paradox (e.g., Scaife and Smith 2018; Weisheimer et al. 2024).

Potential predictability is not stationary or fixed in time (Newman and Sardeshmukh 2017; Weisheimer et al. 2022; Zhao et al. 2016). As a result, actual forecast skill has also varied substantially in the past (Derome et al. 2005; Kumar 2009; MacLeod et al. 2018; O’Reilly et al. 2017, 2019; Shi et al. 2015; Weisheimer et al. 2017, 2019), although overall secular trends in this skill may be harder to detect in the historical record (Ding et al. 2019). For example, Lou et al. (2023) and Weisheimer et al. (2022) showed that long-lead ENSO forecast skill was higher at the beginning and end of the twentieth century, with a multidecadal period of lower skill from the 1930s to 1950s. Further, Weisheimer et al. (2020) found that past seasonal predictability of extratropical atmospheric circulation patterns such as the Pacific–North American (PNA) pattern and the North Atlantic Oscillation (NAO) have also experienced pronounced decadal variations. While these past changes in prediction skill may result from varied model performance relative to historical observations (e.g., Weisheimer et al. 2022), these skill changes may also be driven by changes in the intrinsic predictability of the

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-23-0648.s1>.

Corresponding author: Dillon J. Amaya, [dillon.amaya@noaa.gov](mailto:dillon.amaya@noaa.gov)

DOI: 10.1175/JCLI-D-23-0648.1

© 2025 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

climate system itself (Becker et al. 2014; Newman and Sardeshmukh 2017).

Given these historical changes, it is reasonable to expect that potential predictability and/or actual prediction skill may similarly vary in the future, whether as a result of natural decadal variability (Weisheimer et al. 2020), a possible response to anthropogenic climate change (Zheng et al. 2022), or some combination of both. In particular, some general circulation models (GCMs) project that ENSO and its remote impacts may change in response to an increase in greenhouse gasses (e.g., Cai et al. 2021). For example, some models project significant changes in ENSO variability (Maher et al. 2023; Heede and Fedorov 2023), frequency (Berner et al. 2020), flavor (i.e., central vs eastern Pacific; Capotondi et al. 2015), and teleconnection strength/position (Gan et al. 2017; McGregor et al. 2022; O'Brien and Deser 2023; Zhou et al. 2014), although there is substantial model disagreement in the sign and intensity of these changes across models. Still, through its far-reaching teleconnections, ENSO is the single most important source of predictability on seasonal time scales for much of the globe (e.g., Barnett and Preisendorfer 1987; Jacox et al. 2019; Quan et al. 2006), suggesting that any future changes to ENSO's strength and/or its connectivity to the rest of the climate system could significantly impact the potential predictability of many socioeconomically relevant climate parameters.

It is crucial to assess how potential predictability may evolve as climate continues to change. Many previous studies have used hindcast systems to estimate potential predictability in the past (e.g., Shi et al. 2015; Weisheimer et al. 2019, 2020, 2022). However, model hindcasts are not useful for quantifying possible future changes in predictability as they are by definition retrospective and depend on past observations for their initialization. A different technique that can overcome these limitations and assess time-varying climate predictability in the past and the future is the “model-analog” approach. In the traditional analog framework, past observed climate states are found that closely match the current state and their subsequent evolution is treated as forecasts (Lorenz 1969). Alternatively, coupled GCMs allow for analogs to be drawn from lengthy climate simulations (often preindustrial control runs; Ding et al. 2018), with the model evolution of these analogs then treated as the forecast. This method increases the “library” of possible climate states to compare against the current observed state, resulting in closer analog matches and allowing for the generation of forecast ensembles. Such model-analog forecasts have been shown to be as skillful as initialized dynamical forecasts (Ding et al. 2018, 2019; Ding and Alexander 2023; Lou et al. 2023; Menary et al. 2021), with the added benefit of being more computationally efficient.

The “perfect model-analog” technique utilizes these same methods, but while the goal of a traditional model-analog is to leverage climate simulations to forecast the real world, the goal of using perfect model-analogs is to instead forecast the climate simulation itself (e.g., Ding et al. 2018; Zhang et al. 2023; Lenssen et al. 2024). This is accomplished by treating a portion of a climate simulation as “observations” and then drawing the analog forecasts from a different, independent portion of the same climate simulation. The resulting ensemble forecast is “perfect” in

that it has no unconditional or conditional biases (von Storch and Zwiers 1999). Thus, the forecast skill in a perfect model framework is a measure of the potential predictability (or equivalently, “potential skill”) in the climate system, at least to the extent it is captured by the model. Ding et al. (2018) showed that the model-analog technique applied in a perfect model sense to four different climate models provided an excellent estimate of the potential predictability of each model. Since the perfect model framework does not depend on real-world observations, it can be readily applied to past and future climate simulations to explore how these predictability limits change over time.

In this study, we explore whether and how seasonal climate predictability varies from 1921 to 2100 by applying the perfect model framework to five coupled model initial condition large ensembles (LEs) that are each forced with time-varying radiative forcing. The large number of ensemble members provided by each model LE (ranging from 30 to 100 depending on the model) allows us to generate hundreds of thousands of perfect model forecasts with which to assess any past, present, or projected changes in potential predictability. In particular, we generate 24-month forecasts of ENSO as well as global surface temperature, precipitation, and upper-atmospheric circulation. The forecasts are then verified against independent portions of the same large ensembles using anomaly correlation coefficient (ACC). We find that historical and projected ENSO amplitude changes generate global-scale shifts in climate predictability via ENSO-driven changes in the signal-to-noise (S2N) ratio of seasonal forecasts. However, given substantial model disagreement in projected ENSO amplitude changes, the future forced response in potential predictability remains unclear.

## 2. Data and methods

### *a. Climate model large ensemble simulations and observations*

We apply the perfect model framework to five coupled model initial condition LEs that span the phase 5 of the Coupled Model Intercomparison Project (CMIP5) and CMIP6 eras (Table 1). Such a comparison across models allows us to test the sensitivity of our results to intermodel uncertainty found in the climate response to increased radiative forcing. For efficiency, all model data output was first interpolated to a common  $2.5^\circ \times 2.5^\circ$  grid.

The models used in our analysis include the Community Earth System Model, version 1.2, LE (CESM1-LE; 40 members; Kay et al. 2015), CESM, version 2, LE (CESM2-LE; 100 members; Rodgers et al. 2021), the Geophysical Fluid Dynamics Laboratory Seamless System for Prediction and Earth System Research (GFDL-SPEAR) medium-resolution simulation (30 members; Delworth et al. 2020), the GFDL Earth System Model with MOM, version 4 component (GFDL-ESM2M; 30 members; Burger et al. 2020, 2022), and the Max Planck Institute Grand Ensemble (MPI-GE; 100 members; Maher et al. 2019). The analysis period is 1921–2100, during which each model uses one of the specified external forcing scenarios: 1) historical + representative concentration pathway 8.5 (RCP8.5), 2) historical + shared socioeconomic pathway 3-7.0 (SSP3-7.0), or 3) historical + SSP5-8.5.

TABLE 1. Observational and model datasets used in this study. First column: radiation forcing scenario used by each model. The number of ensemble members available in each model is in parentheses. Second column: DJF averaged trend ( $^{\circ}\text{C decade}^{-1}$ ) of Niño-3.4 standard deviation ( $\sigma_{3,4}$ ) in 30-yr running windows (i.e., Fig. 1) for the period 1950–2023. For climate models, the ensemble mean trend is reported along with the percent of the respective ensemble members that agree with the sign of ensemble mean trend in parenthesis. For observations, the trend is reported along with a 95% confidence interval on the slope of the linear fit. Third column: As in the second column, but for the period 1950–2100. Fourth column: Dataset references.

Dataset	Forcing (ensemble size)	$\sigma_{3,4}$ trend ( $^{\circ}\text{C decade}^{-1}$ ) 1950–2023	$\sigma_{3,4}$ trend ( $^{\circ}\text{C decade}^{-1}$ ) 1950–2100	References
CESM1-LE	HIST+RCP8.5 (40)	0.04 (90%)	0.02 (88%)	<a href="#">Kay et al. (2015)</a>
CESM2-LE	HIST+SSP3-7.0 (100)	0.03 (77%)	0.00 (68%)	<a href="#">Rodgers et al. (2021)</a>
GFDL-SPEAR	HIST+SSP5-8.5 (30)	0.02 (73%)	0.03 (100%)	<a href="#">Delworth et al. (2020)</a>
GFDL-ESM2M	HIST+RCP8.5 (30)	0.02 (70%)	−0.02 (93%)	<a href="#">Burger et al. (2020, 2022)</a>
MPI-GE	HIST+RCP8.5 (100)	0.00 (54%)	0.00 (66%)	<a href="#">Maher et al. (2019)</a>
ERSSTv5		0.03 ± 0.01		<a href="#">Huang et al. (2017)</a>
HadISST		0.07 ± 0.01		<a href="#">Rayner et al. (2003)</a>
COBE		0.07 ± 0.01		<a href="#">Ishii et al. (2005)</a>
COBE2		0.06 ± 0.01		<a href="#">Hirahara et al. (2014)</a>

Within a given model, each ensemble member starts from a different initial condition. Over time, the ensemble members diverge due to the chaotic nature of the coupled climate system. As a result, once the memory of the initial condition fades, each ensemble member can be treated as an independent sample of the climate that has its own unique sequence of internal variability superimposed on a common forced response. We compare a portion of our model results to monthly mean data from four different observational sea surface temperature (SST) datasets: the National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed SST, version 5 (ERSSTv5; [Huang et al. 2017](#)), Hadley Centre Sea Ice and SST (HadISST; [Rayner et al. 2003](#)), and Centennial In Situ Observation-Based Estimates (COBE), version 1 (COBE1; [Ishii et al. 2005](#)) and version 2 (COBE2; [Hirahara et al. 2014](#)). We primarily compare these observational products during the period 1921–2023. Differences in data streams and gap-filling methods lead to differences between each observational estimate during the early part of the twentieth century, although these are less than the differences between model ensembles in the corresponding period. However, the observations are in better agreement during a more recent period (e.g., 1980–2023).

### b. Perfect model-analog framework

In each LE, perfect model forecasts are generated and evaluated for different 30-yr periods spaced every 10 years from 1921 to 2100 (e.g., 1921–50, 1931–60...2071–2100). The forecasts are produced within each of these 30-yr periods separately using the following method. For a given model and 30-yr period, the following steps are carried out.

- 1) We extract monthly mean SSTs from each ensemble member for the 30-yr period of interest.
- 2) We then remove the long-term monthly mean SSTs at each grid point based on the contemporaneous climatology calculated using all ensemble members (i.e., anomalies in 1921–50 are relative to a 1921–50 climatology).
- 3) We further remove the ensemble mean SST anomaly (SSTA) (i.e., the model-specific externally forced signal)

at each grid point from each of the model’s individual ensemble members.

- 4) We arbitrarily treat the first ensemble member as the “truth” or observations. Because each ensemble member is independent from one another, a data library of possible analog matches to the observations can then be constructed for each calendar month using the remaining ensemble members. For example, the data library for January in CESM1-LE consists of 39 ensemble members  $\times$  28 years = 1092 samples. Note that it is only 28 years because we aim to generate 24-month forecasts, so any possible analog matches in the final 2 years would extend beyond our 30-yr window of interest. Thus, the final 2 years in each 30-yr window are excluded from our data libraries.
- 5) For a given month, we choose analogs by minimizing the distance between the climate state in the “observed” ensemble member and those found in the corresponding monthly data library (i.e., by comparing an observed January to the January data library). The distance between climate states is estimated by calculating the total root-mean-square (RMS) difference between the observed SSTAs from 60°S to 60°N and at all longitudes and those from every possible match in the data library. Note that we do not area weight the RMS difference calculation used in our analysis (see following section for more details). The distances are then ranked in descending order. The 10 closest states from the data library and their subsequent 24-month evolution are chosen as the forecast ensemble for that month.
- 6) We repeat (4) and (5) by treating each other model ensemble member as observations and constructing the monthly data library using the remaining ensemble members.

This procedure generates a 10-member forecast for every month and every ensemble member in a given model LE. For example, applying this perfect model framework to CESM1-LE for a given 30-yr period generates 40 (ensemble members)  $\times$  12 (calendar months)  $\times$  28 (years) = 13 400 10-member, 24-month forecasts with which we can estimate seasonal climate predictability. Although we use SSTAs to identify analogs, we are not

limited only to SSTA forecasts for analysis. Once the nearest climate states are selected, the evolution of any model variable can be treated as a forecast and subsequently verified against the corresponding variable from observations (e.g., Ding et al. 2019). In this way, we assess the forecast anomalies of the following monthly mean variables from each model, with the CMIP standard variable name shown in parenthesis: SST (tos), 2-m air temperature over land (tas), precipitation (pr), and the 500-mb (1 mb = 1 hPa) streamfunction, which was calculated using the U/V wind components at 500 mb (ua, va). Note that the 500-mb streamfunction was chosen instead of a more traditional parameter like 500-mb geopotential height because not all of the LEs provided geopotential height as model outputs. As previously mentioned for SST, anomalies for all other variables are derived by removing both the long-term monthly means of the contemporaneous 30-yr period and each model's respective ensemble mean.

### c. Perfect model-analog sensitivities

There are several arbitrary choices that must be made when adapting the perfect model-analog technique for LEs. Here, we briefly discuss these decisions and how they might influence our results or conclusions. 1) We remove a given model's ensemble mean from each of its members in order to isolate the internal component of each parameter, while still accommodating forced changes in climate variability (O'Brien and Deser 2023). Doing so allows us to focus on possible forced changes in the predictability of climate variations, as opposed to the more trivial exercise of predicting the forced trend. 2) Ding et al. (2018) showed that for data libraries of several hundreds of years, analog forecast ensembles of 10–20 members produced the most accurate forecasts. This is because larger forecast ensembles include increasingly poor analog matches, resulting in lower skill over the length of the forecast. We choose the top 10 analogs for our forecast ensembles for computational efficiency; however, our results and conclusions are not qualitatively impacted when increasing the forecast ensemble size to the top 15 or 20 matches. 3) We do not area weight the RMS difference calculation so as not to overweight the tropics when drawing analogs. We find that this choice increases the overall forecast skill in the midlatitudes without overly decreasing it in the tropics. We select analogs based on SSTAs from 60°S to 60°N and at all longitudes for similar reasons (i.e., to improve the representation of the extratropics when selecting analogs). Our results and conclusions are not qualitatively impacted by these decisions.

### d. Potential predictability metrics and signal-to-noise

To assess lead-dependent potential skill in each model, we calculate  $N_e$  estimates of the ACC between each ensemble mean forecast and the corresponding observations, where  $N_e$  is the number of ensemble members in a given LE (i.e., the number of observed time series used to generate analogs). For example, there are 40 estimates of the ACC for 1921–50 when evaluating CESM1-LE. We repeat this procedure for each 30-yr period separately, and we report the ensemble mean ACC in our results. We test the significance of the

ensemble mean ACC using a 95% confidence interval based on the two-sample  $t$  test. We further determine the robustness of the change in ACC between 30-yr periods by indicating where 90% of a given model's ensemble members agree on the sign of the change. Note that all of our results are consistent with each other when using probabilistic skill metrics (e.g., Brier skill score or reliability). However, for brevity, we only show skill based on ACC.

We assess the lead-dependent S2N ratio in our forecasts following Sardeshmukh et al. (2000). For each model ensemble member  $N_e$ , the S2N ratio at lead  $l$  is

$$S2N(N_e, l) = \left( \frac{\sum_{i=1}^n \bar{x}_f^2}{\frac{1}{K} \sum_{i=1}^m x_f^2} \right)^{1/2}, \quad (1)$$

where  $x'_f = x_f - \bar{x}_f$  is the deviation of each individual forecast member ( $x_f$ ) from the ensemble mean forecast ( $\bar{x}_f$ ) for each monthly initialization ( $i$ ). The numerator in Eq. (1) represents the signal and is calculated as the square of  $\bar{x}_f$  summed over all monthly initializations in a given 30-yr period ( $n$ ). The denominator represents the noise and is calculated as the square of  $x'_f$  summed over  $m$ , where  $m$  is the total number of monthly initializations from all forecast ensemble members  $K$  (in our analysis  $K = 10$ ). Therefore, for a given 30-yr period,  $n = 12$  (calendar months)  $\times 28$  (years) = 336 and  $m = 3360$ . As with ACC, we calculate  $N_e$  estimates of the S2N ratio for each LE (one for each ensemble member) and report the ensemble mean values in our results. A higher S2N ratio indicates that there is a larger ensemble mean anomaly and/or less spread among the forecast ensemble, which results in a more skillful forecast in the perfect model framework (Sardeshmukh et al. 2000).

Projected changes in ACC and S2N are shown relative to a reference period of 1921–50. This reference was chosen as it is the earliest common period among the model LEs and because it allows us to quantify the largest possible radiatively forced change in each model. Note that while the CESM2-LE, GFDL-ESM2M, and MPI-GE were initialized in the mid-nineteenth century, CESM1-LE and GFDL-SPEAR were initialized in 1920 and 1921, respectively. Since the ensemble members in GFDL-SPEAR were initialized from different years of a preindustrial control simulation with different ocean states, they are already well separated. However, since the ensemble members in CESM1-LE were initialized with a similar ocean state, it is possible that the ocean memory of the initial condition is present in this model for the earliest 30-yr period over which we quantify predictability (i.e., 1921–50). However, we have repeated our analysis using 1991–2020 as the reference period and found that our results and conclusions are not qualitatively impacted by this choice.

## 3. Results

### a. Forced changes in ENSO amplitude

Given ENSO's dominant role in driving seasonal climate predictability, we first assess the simulated response of ENSO

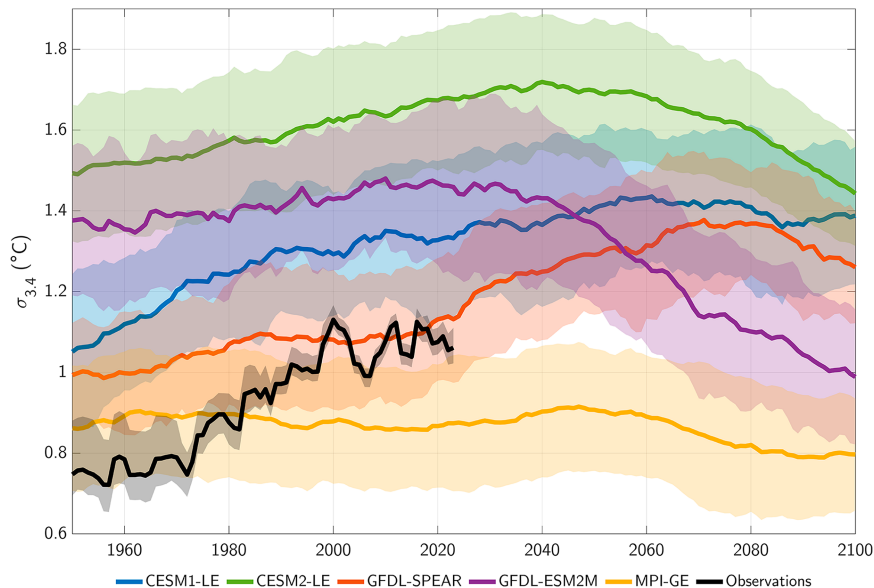


FIG. 1. Standard deviation of DJF averaged SSTA in the Niño-3.4 region in running 30-yr windows from 1921 to 2100. Years indicate end of the window (e.g., 1960 = 1931–60). Colors represent different model LEs, with thick curves for ensemble mean values and shading for one standard deviation spread across the ensemble. Black shading shows the range of observed values from 1921 to 2023 from four different observational products (ERSSTv5, HadISST, COBE, and COBE2), with the black curve representing the observed average. See Fig. S1 for individual observational estimates.

amplitude to historical and future radiative forcing in each LE, measured using a 30-yr moving window of December–February (DJF) Niño-3.4 (SSTA averaged 5°S–5°N, 170°–120°W) standard deviation from 1921 to 2060 (Fig. 1). In general, while these model simulations capture many key aspects of ENSO and its evolution (Maher et al. 2023; Planton et al. 2021), they struggle to capture ENSO amplitude (Fig. 1 black line), which they mostly overestimate apart from the MPI-GE in the mid-twentieth century and GFDL-SPEAR in the later twentieth century. Still, most model ensemble means show an increase in Niño-3.4 amplitude over the latter half of the twentieth century, which appears consistent with observations. However, SST reconstructions as well as surface pressure observations also suggest that ENSO amplitude had a mid-twentieth century minimum, with higher values early in the century (Fig. S1 in the online supplemental material; Newman et al. 2018; Lou et al. 2023). Additionally, observed ENSO amplitude may not have significantly increased over the past two decades, whereas the longer model simulations mostly show ensemble mean trends that monotonically increase over the entire twentieth and into the twenty-first century (Fig. S1). Moreover, despite noticeable changes in ENSO amplitude in most of the model’s ensemble means, there are large uncertainties in the sign of the trend across the respective LE’s members during the observational period. Indeed, only in CESM1-LE do at least 90% of the ensemble members agree on the sign of the ensemble mean trend from 1950 to 2023 (Table 1), suggesting that it may be difficult to identify forced changes in ENSO amplitude during the historical period in the other LEs. Thus,

while observations do show a robust increase in ENSO amplitude from 1950 to 2023 (Table 1), the large model uncertainties and overall variability biases make it difficult to conclude whether or not these observed trends are consistent with any of the LE’s forced response in ENSO amplitude during the historical period.

Given general (if not uniform) model agreement concerning the direction of historical changes in ENSO amplitude, it is interesting that their projections of future ENSO changes are notably different. The ensemble mean Niño-3.4 amplitude in CESM2-LE rises consistently through 2040 before decreasing consistently through 2100, while in CESM1-LE, it rises through 2060 before a milder decrease thereafter. In contrast, in the GFDL-SPEAR ensemble, ENSO undergoes a relatively mild increase from 1921 to 2020, after which it increases until about 2080 before decreasing slightly. In GFDL-ESM2M, the ensemble mean Niño-3.4 standard deviation is relatively stable until about 2040, after which it sharply decreases through the end of the century. The exception is the MPI-GE ensemble, which has little change over the historical period and even declines somewhat through 2100.

Projected trends in ENSO amplitude are more robust among some of the LEs (Table 1). All ensemble members agree with the sign of the ensemble mean trend in GFDL-SPEAR (up from 73% during the historical period) and 93% (up from 70%) of the members agree in GFDL-ESM2M. The inter-member agreement in the MPI-GE also increases (from 54% to 66%), but the ensemble mean trend remains weak and is likely not robust. The intermember agreement in CESM1-LE and

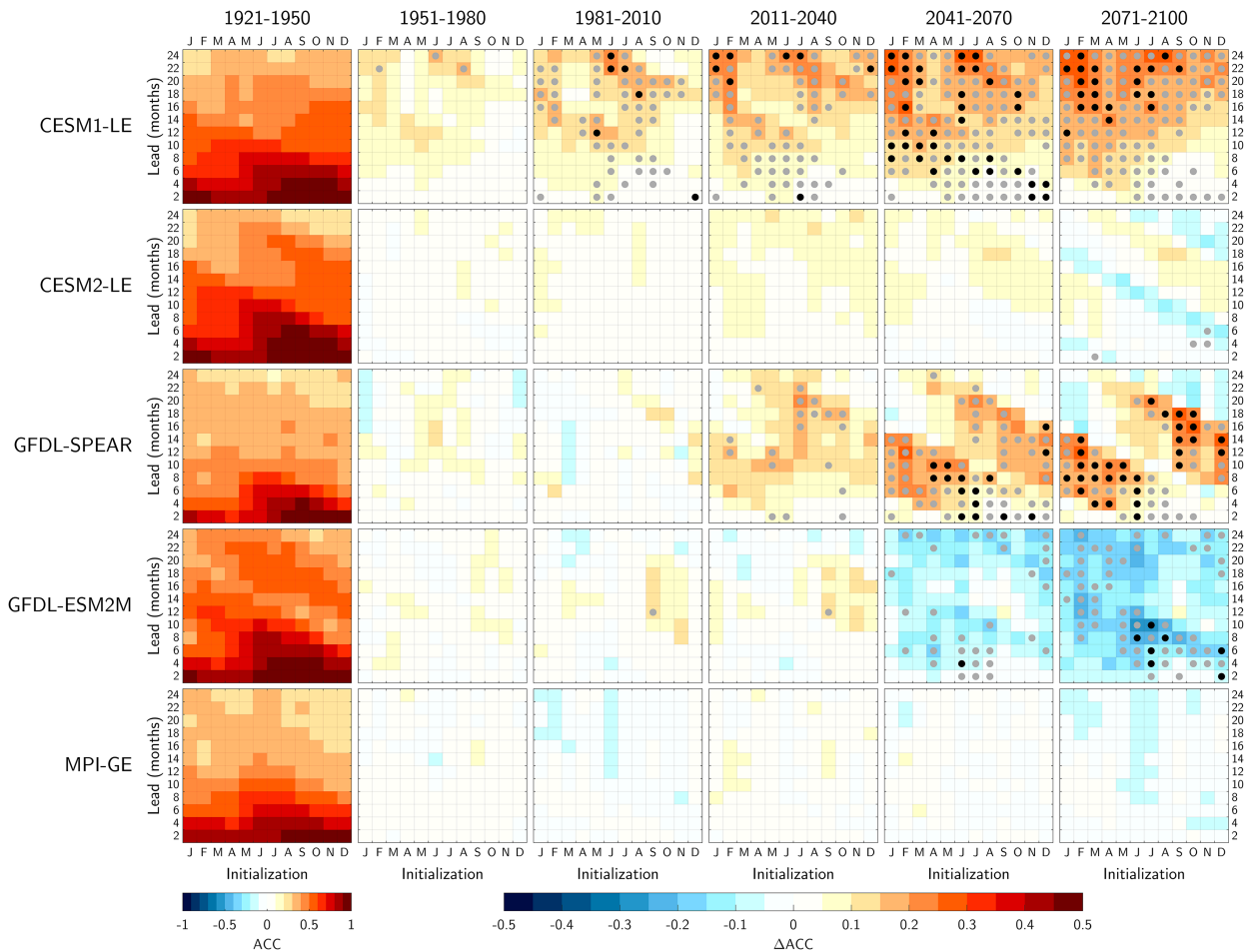


FIG. 2. (first column) Ensemble mean Niño-3.4 potential forecast skill (ACC) as a function of initialization month ( $x$  axis) and lead time ( $y$  axis) for each model LE. (second column)–(sixth column) Difference in Niño-3.4 skill between the base period 1921–50 and different 30-yr periods. For example, the (second column) shows the difference in skill between the periods 1951–80 and 1921–50. Gray (black) stipples indicate that 80% (90%) of the respective model ensemble agrees on the sign of the change.

CESM2-LE decreases slightly from 90% to 88% and from 77% to 68%, respectively. The high (>80%) intermember agreement in CESM1-LE, GFDL-SPEAR, and GFDL-ESM2M suggests that the different projected ENSO amplitude changes seen in these models are likely a robust (and model-specific) response to future radiative forcing. In contrast, the ENSO amplitude trends in the CESM2-LE and MPI-GE remain difficult to separate from internal climate variations.

Projecting future changes in predictability might appear difficult when the trends in simulated future ENSO amplitude disagree, with two models showing a robust increase (CESM1-LE and GFDL-SPEAR), one showing a robust decrease (GFDL-ESM2M), and two showing little overall change (CESM2-LE and MPI-GE). Moreover, it remains problematic whether these models can simulate past changes to ENSO variability. We might then ask whether any conclusion can be drawn regarding future changes in global predictability. However, as we will show, this model disagreement can still be leveraged to increase the understanding of the relationship between ENSO and changes in global climate predictability over time.

### b. Forced changes in ENSO predictability

Keeping in mind the issues we have identified in the previous section, we next explore whether estimates of ENSO predictability during the historical and future periods are similarly varied across the models. For the period 1921–50, perfect model-analog forecasts (hereafter referred to as “forecasts”) of the Niño-3.4 index in each LE show high (ACCs exceeding 0.8) potential skill (hereafter referred to as “skill”) at leads of less than ~6 months and for forecasts initialized/verified in boreal fall and winter (Fig. 2; left column). Skill remains elevated for forecasts verified in boreal fall/winter (regardless of initialization month or lead time), consistent with previous studies (e.g., Jacox et al. 2019).

Treating 1921–50 as a baseline, there is little change in Niño-3.4 skill in any of the models for the adjacent 30-yr period (1951–80). However, by the period 1981–2010, CESM1-LE shows a robust increase in Niño-3.4 predictability at short leads for May–September initializations and at longer leads for much of the year. This suggests that forced changes in

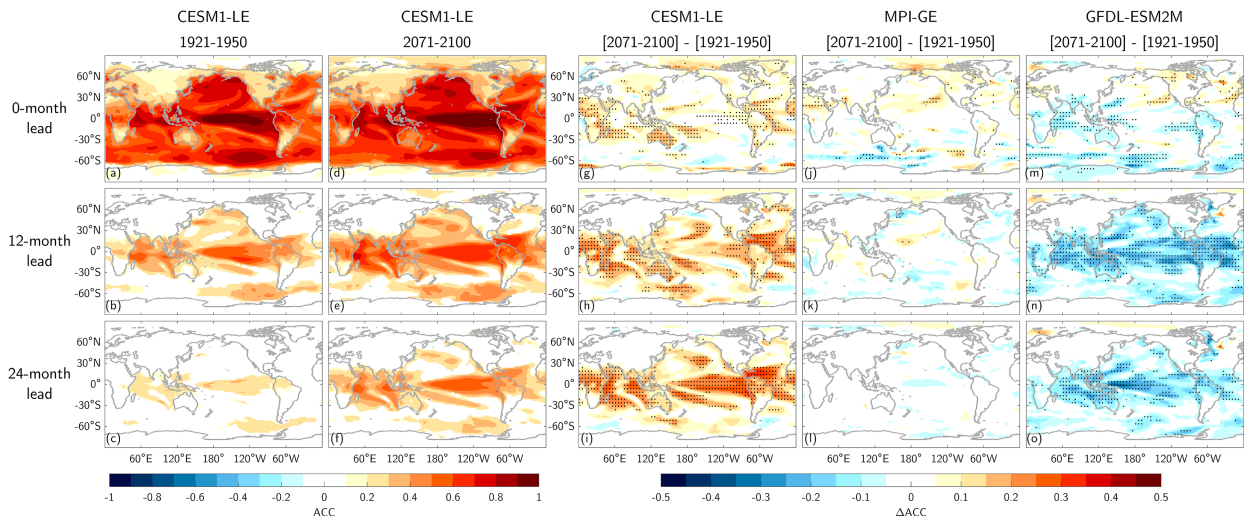


FIG. 3. Surface temperature potential predictability. (a)–(c) Ensemble mean skill of surface temperature anomalies in CESM1-LE as measured by ACC calculated across all months in the period 1921–50. (d)–(f) As in (a)–(c), but for the period 2071–2100. (g)–(o) Change in ACC between past and future periods for (g)–(i) CESM1-LE, (j)–(l) MPI-GE, and (m)–(o) GFDL-ESM2M. Skill values in (a)–(f) are only shown when 95% significant. Stipples in (g)–(o) indicate where 90% of each respective ensemble agrees on the sign of the change.

CESM1-LE ENSO predictability begin to emerge above the internal noise inherent to each ensemble member during this period. In 2011–40, CESM1-LE Niño-3.4 skill continues to increase, while GFDL-SPEAR begins to show some robust increases in predictability. Forecast skill in CESM2-LE also increases slightly during this period, but there is not widespread agreement among its ensemble members on the sign of this change. We see the largest period-to-period changes in Niño-3.4 skill between 2011–40 and 2041–70 (Fig. 2; fifth column). For example, forced changes to ENSO forecast skill in GFDL-SPEAR fully emerge during this period, with the diagonal bands of increased predictability associated with forecasts that verify in boreal summer to winter. In GFDL-ESM2M, robust decreases in predictability begin to emerge, but without a clear pattern. Finally, by the period 2071–2100, CESM1-LE and GFDL-SPEAR largely maintain the increases in ENSO predictability observed in the previous epoch, while forced decreases in Niño-3.4 forecast skill are now fully evident in GFDL-ESM2M.

There is clear model diversity in the simulated change in ENSO predictability over time, even among different versions of the same model (i.e., CESM1-LE vs CESM2-LE and GFDL-SPEAR vs GFDL-ESM2M). Regardless, the sign and timing of ENSO predictability changes in each of the LEs is consistent with their respective time-varying ENSO amplitudes (Fig. 1). For example, there are no robust changes in Niño-3.4 forecast skill in GFDL-SPEAR until the period 2011–40, which closely corresponds to the timing of the strongest increasing trend in this model’s ENSO amplitude (comparing third row of Fig. 2 to orange line in Fig. 1). Similarly, ENSO predictability in GFDL-ESM2M remains relatively stable until the period 2041–70, at which point both the forecast skill and GFDL-ESM2M’s ENSO amplitude start to sharply decrease (comparing fourth row of Fig. 2 to purple line in Fig. 1). Even the ensemble mean Niño-3.4 skill in CESM2-LE shows

hints of a close link to its weakly time-varying ENSO amplitude, with a slight increase in skill/amplitude through 2040 followed by a decrease through the end of the century, though these predictability changes are not robust across the CESM2 ensemble. The similarities between the timing of ENSO amplitude and ENSO predictability changes in the LEs suggest that there may be a causal relationship between the two. We will explore this possibility in greater detail in section 3d.

### c. Forced changes in global climate predictability

Due to ENSO’s far-reaching teleconnections, it is possible that the relationship between ENSO amplitude and climate predictability in each of the models extends beyond the Niño-3.4 region. To investigate, we calculate the projected change in the global predictability of surface temperature, precipitation, and upper-atmospheric circulation (Figs. 3–5). For brevity, we primarily focus on predictability changes in the CESM1-LE, MPI-GE, and GFDL-ESM2M in this analysis as these LEs span the model uncertainty in terms of projected ENSO amplitude change (i.e., a robust increase, no change, and a robust decrease, respectively); however, the results for all models are shown in Figs. S3–S17.

Forecasts of SSTA for 1921–50 in CESM1-LE show skill at 0-month lead (i.e., the “initialization” or the month that was used to determine the analogs) for most of the globe (globally averaged ACC = 0.62), with the tropical Pacific exhibiting the highest skill (ACCs > 0.9; Fig. 3a). There is also significant skill of near-surface air temperature anomalies (SATA) over land at 0-month lead in most regions. However, SATA skill is generally weaker than for SSTA (globally averaged ACC = 0.48), especially in midlatitudes. Overall, the high 0-month lead SSTA skill gives us confidence that the perfect model framework is reliably drawing analogs that closely correspond to the observed climate states at each initialization.

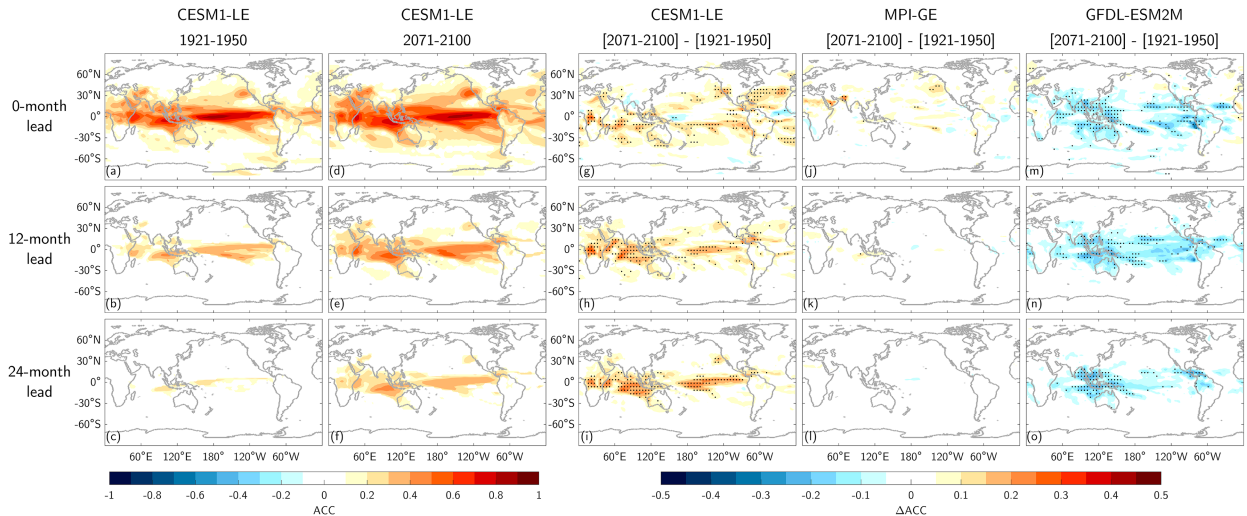


FIG. 4. As in Fig. 3, but for precipitation predictability.

The skill of surface temperature decreases with increasing lead time (Figs. 3b,c; see Figs. S3–S7 for skill maps at additional lead times), although this reduction is more apparent for SATA than for SSTA. The SSTA ACCs at 12-month lead exceed 0.6 in the tropical Pacific, consistent with previous model-analog forecast studies (e.g., Ding et al. 2018). There is also significant SATA predictability over tropical land surfaces, as well as significant SSTA predictability throughout most of the North Pacific, the tropical Atlantic, the tropical Indian Ocean, and the Southern Ocean west of the Drake Passage. These regions are known to be influenced by large-scale ENSO teleconnections (e.g., He et al. 2020; Horel and Wallace 1981; Mo and Ghil 1987), confirming that ENSO is a key source of long-lead predictability in our forecasts. Skill further degrades out to 24-month leads (Fig. 3c). The historical skill in CESM1-LE is qualitatively representative of the other LEs (Figs. S3–S7); the other LEs each show elevated

skill in the tropics and in regions influenced by ENSO, as well as skill that degrades more slowly over the ocean than over land.

In CESM1-LE, there is a robust increase in SSTA and SATA predictability in the future at all leads, with only a few small regions of decreasing predictability (Figs. 3d–i). In particular, the 0-month lead SSTA skill increases in the western tropical Pacific as well as the Indian and Atlantic Oceans (Fig. 3i). Similarly, there is a robust increase in projected SATA predictability at 0-month lead over much of Africa, portions of eastern Asia, equatorial South America, and all of Australia. Further, the widespread ensemble agreement (black stipples) indicates that these predictability changes are a “robust” (defined here as 90% ensemble agreement on the sign of the change) part of the model-specific forced response and not due to chance. The CESM1-LE changes in SSTA/SATA predictability are starker at 12- and 24-month leads (Figs. 3h,i), with robust increases

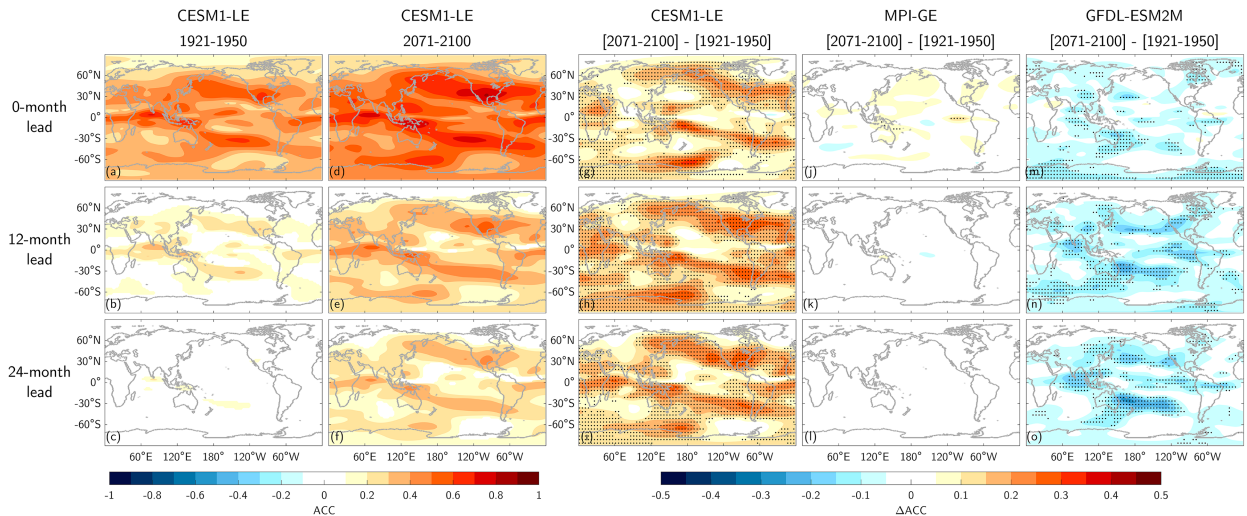


FIG. 5. As in Fig. 3, but for 500-mb streamfunction ( $\psi_{500}$ ) predictability.



in ACC throughout the global tropics in an ENSO-like pattern. There are also robust long-lead increases in SSTA skill in the North Pacific and along the U.S. West Coast and SATA skill in the American Southwest. Combined, the increase in skill in the global tropics as well as extratropical regions known to be influenced by ENSO suggests that the projected increases in ENSO amplitude simulated by CESM1-LE (Fig. 1) may indeed be manifesting on the global scales via ENSO-driven teleconnections. The existence of this relationship is further supported by the comparison of the CESM1-LE skill changes to the absolute value of the regression of CESM1-LE surface temperature anomalies on the Niño-3.4 index (Fig. S2), which we use as a measure of the strength of ENSO's teleconnections in this model. The pattern correlation is 0.67 at 12-month lead and 0.77 at 24-month lead, suggesting that the areas with the largest change in predictability closely correspond to the regions that are most sensitive to Niño-3.4 variability.

The projected changes in global surface temperature predictability among the other models are also consistent with their respective simulated shifts in ENSO amplitude (Figs. 3j–o and Figs. S3–S7). For example, the MPI-GE at 0-month lead shows some isolated regions of increasing and decreasing SSTA/SATA skill, but without a clear pattern. At longer leads, the skill change in the MPI-GE is close to zero nearly everywhere and there is little agreement among the ensemble on the sign of the change, consistent with the insignificant ENSO amplitude trend in this model. In contrast, GFDL-ESM2M shows a robust decrease in SSTA/SATA predictability for most of the globe (Figs. 3m–o) in a similar ENSO-like pattern as seen in CESM1-LE (pattern correlation =  $-0.74$  at 12-month lead), consistent with its projected decreasing trend in ENSO amplitude and an overall decrease in the predictability of ENSO's teleconnections. Indeed, the pattern correlation between the GFDL-ESM2M skill changes and this model's surface temperature/Niño-3.4 regression map (Fig. S2) is  $-0.61$  at 12-month lead and  $-0.64$  at 24-month lead.

These results hold for other climate parameters. For example, CESM1-LE shows robust increases in projected precipitation predictability at all leads in CESM1-LE (Figs. 4d–i), with centers of action that coincide with known ENSO teleconnections, such as the Indian Ocean (e.g., Gupta et al. 2023), the Caribbean (e.g., Giannini et al. 2001), and the U.S. West Coast (e.g., Cash and Burls 2019). Similar to the surface temperature skill changes, the relationship between precipitation skill changes and ENSO teleconnections is further supported by the pattern correlation with the regression map of CESM1-LE precipitation anomalies on the Niño-3.4 index ( $R = 0.70$  at 12-month lead and  $0.82$  at 24-month lead; Fig. S2). The precipitation skill changes are also consistent with forecasts of 500-mb streamfunction anomalies ( $\psi_{500}$ ) in CESM1-LE (Figs. 5d–i), which show an increase in skill in the PNA and Pacific–South American (PSA) regions (Horel and Wallace 1981; Mo and Ghil 1987) and have a pattern correlation with the  $\psi_{500}$ /Niño-3.4 regression map of  $0.88$  at both 12- and 24-month leads. As with SSTA/SATA, the MPI-GE shows only isolated regions of robust precipitation/ $\psi_{500}$  skill change at 0-month, but no significant change at longer leads (Figs. 4j–l and 5j–l). Similarly, GFDL-ESM2M shows a robust decrease in precipitation

predictability at all leads throughout the tropics and in midlatitude regions influenced by ENSO, such as the U.S. West Coast and Australia (Figs. 4m–o), which is supported by pattern correlations of  $-0.60$  at 12-month lead and  $-0.64$  at 24-month lead with its precipitation/Niño-3.4 regression map (Fig. S2). This is consistent with the decrease in GFDL-ESM2M  $\psi_{500}$  skill in the PNA and PSA regions of the North and South Pacific (Figs. 5m–o) and further supported by pattern correlations of  $-0.78$  12-month lead and  $-0.77$  at 24-month lead with its  $\psi_{500}$ /Niño-3.4 regression map (Fig. S2). Results are consistent for additional leads times and for GFDL-SPEAR and CESM2-LE (Figs. S8–S17).

As was the case with projected ENSO amplitude, the large model disagreement in the sign and intensity of projected climate predictability change suggests that we cannot make any definitive conclusions about whether predictability will be improved or degraded in the future based on these models alone. However, the consistency between each model's ENSO amplitude (Fig. 1), ENSO predictability (Fig. 2), and global climate predictability (Figs. 3–5) suggests a causal link exists between future changes in ENSO and future changes in global climate predictability.

#### d. Linking projected predictability changes to ENSO amplitude

##### 1) SIGNAL-TO-NOISE

The link between projected climate predictability and ENSO amplitude may be related to ENSO's role as the dominant internal climate mode, allowing one to detect its influence across much of the globe despite the presence of other forms of variability (e.g., weather or other climate modes). For example, if ENSO amplitude increases in the future (e.g., as projected by CESM1-LE and GFDL-SPEAR), then that may lead to an increase in the S2N ratio of ENSO and its teleconnections, which would tend to contribute to an overall more deterministic climate system and more skillful forecasts (e.g., Sardeshmukh et al. 2000). To test this hypothesis, we calculate changes in the S2N ratio [Eq. (1)] for surface temperature as a function of lead time in each of the two time periods (Fig. 6). During the period 1921–50, the S2N ratios in CESM1-LE forecasts at 0-month lead follow an ENSO-like pattern, with the highest values in the equatorial Pacific (maximum value = 1.94). Weaker (but still elevated) values are seen in the Indian Ocean, the South Pacific, the northeast Pacific along the U.S. West Coast, the North Atlantic, and over the tropical African and South American land surfaces (Fig. 6a). The S2N ratio decreases with increasing lead time (Figs. 6b,c); however, the ENSO-like pattern of elevated S2N ratio persists at 12-month lead before mostly dissipating at 24-month lead.

The patterns of projected S2N change in each of the LEs (Figs. 6g–o) are remarkably similar to the surface temperature ACC changes seen in Fig. 3, with pattern correlations between the ACC and S2N maps at 0-, 12-, and 24-month lead of  $0.89$ ,  $0.98$ , and  $0.99$  for CESM1-LE;  $0.88$ ,  $0.93$ , and  $0.84$  for the MPI-GE; and  $0.81$ ,  $0.98$ , and  $0.97$  for GFDL-ESM2M, respectively. Decomposing the S2N equation into a signal and

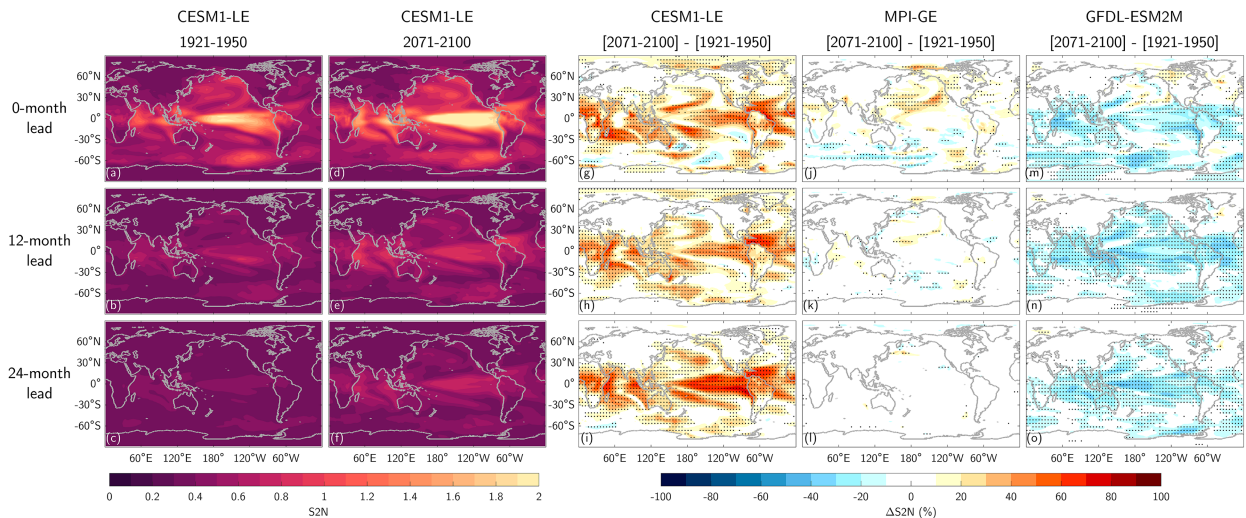


FIG. 6. S2N ratios for surface temperature anomaly forecasts. (a)–(c) Ensemble mean S2N of surface temperature forecasts in CESM1-LE calculated across all months in the period 1921–50. (d)–(f) As in (a)–(c), but for the period 2071–2100. (g)–(o) Percent change in S2N between past and future periods for (g)–(i) CESM1-LE, (j)–(l) MPI-GE, and (m)–(o) GFDL-ESM2M. Stipples in (g)–(o) indicate where 90% of a respective model’s ensemble agrees on the sign of the change.

noise component [i.e., the numerator and denominator of Eq. (1), respectively], we find that the changes in the signal are over 5 times larger than the changes in the noise for much of the globe (Figs. S18 and S19). For example, the signal change averaged 60°S–60°N at 12-month lead in CESM1-LE is 27%, compared to just a 4.7% change in the noise. In the case of CESM1-LE, this indicates that the amplitude of a typical ensemble mean forecast anomaly is larger in the future without a substantial increase in the average forecast spread (i.e., the forecast uncertainty). These results are consistent with previous studies linking ENSO amplitude to S2N and/or climate predictability (Capotondi et al. 2015; Chen et al. 2004; Gu and Philander 1997; Sardeshmukh et al. 2000; Suarez and Schopf 1988; Weisheimer et al. 2022; Zhao et al. 2016).

## 2) TIME-VARYING GLOBAL PREDICTABILITY CHANGES

From Fig. 6, it is clear that at the end of the twenty-first century, ENSO amplitude changes manifest on global scales via ENSO-driven changes in the S2N ratio. This relationship is apparent in all of the LEs and over time. In particular, globally averaged S2N ratios linearly scale with time-varying ENSO amplitude in each model (Fig. 7a), even in the MPI-GE and CESM2-LE where time-varying ENSO amplitude changes are weak. There is also a near-perfect linear relationship between globally averaged S2N and ACC (Fig. 7b), consistent with previous studies relating perfect model skill to S2N ratios (Sardeshmukh et al. 2000). As a result, we find a high correspondence between each LE’s time-evolving Niño-3.4 amplitude and their respective globally averaged ACC (Fig. 7c). For example, at 12-month lead, the globally averaged SSTA skill in CESM1-LE increases roughly linearly over time with increasing ENSO amplitude (Fig. 7c circles;  $R = 0.95$ ). In GFDL-ESM2M, there is a decrease in skill over time that closely

corresponds to this model’s decrease in ENSO amplitude (Fig. 7c triangles;  $R = 0.97$ ).

By calculating the regression coefficient between globally averaged ACC/S2N and Niño-3.4 amplitude (i.e., by fitting a line to each model’s scatter in Figs. 7a,c), we can estimate the skill/S2N changes for a given change in ENSO amplitude. Averaging the regression coefficients across models, we find that a 10% increase in Niño-3.4 amplitude is associated with a 14% and 7% increase in globally averaged forecast skill and the S2N ratio, respectively, at 12-month lead. Combined, the results shown in Fig. 7 further support our hypothesis that time-varying changes in predictability are driven by same-sign changes in global S2N ratios, which in turn are driven by each respective LE’s projected change in ENSO amplitude. The close link between ENSO amplitude, S2N ratios, and forecast skill is consistent across models (different marker types in Fig. 7), lead times, and variables (not shown).

## 4. Summary and discussion

In this study, we investigated projected changes in seasonal potential predictability across five coupled GCM LEs. Using a perfect model-analog technique, we generated hundreds of thousands of synthetic seasonal forecasts to estimate predictability changes from 1921 to 2100. We found large model disagreement in the sign and intensity of future climate predictability changes, with some models exhibiting increasing trends in predictability and others exhibiting decreasing trends or no significant change. Despite these very different model trajectories, a common relationship emerged between a model’s simulated change in ENSO amplitude and the global climate predictability within that same model. In particular, model-specific shifts in ENSO amplitude manifested on global scales via ENSO-driven changes in the S2N ratio of forecasts, thus altering the skill of

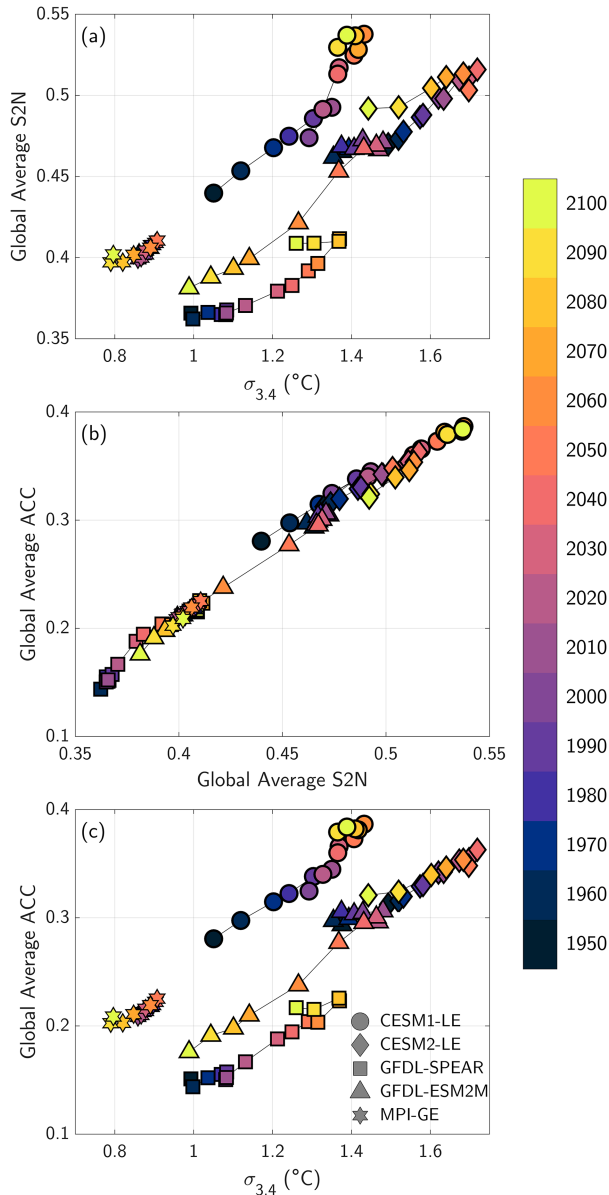


FIG. 7. (a) Global average forecast S2N at 12-month lead (y axis) vs DJF averaged Niño-3.4 standard deviation (x axis) in different 30-yr periods. (b) As in (a), but for global average ACC vs global average S2N ratio. (c) As in (a), but for global average ACC vs Niño-3.4 standard deviation. All ACC and S2N values are based on ensemble mean SSTA forecasts from each model (i.e., different shapes). Shading of each shape indicates the 30-yr window over which the forecast skill, S2N ratio, and Niño-3.4 standard deviation are calculated, with the year indicating the end of the window. For example, the shading for 1950 corresponds to 1921–50.

forecasts as the climate system became more or less deterministic. For example, forecasts from models with increasing ENSO amplitude trends (e.g., CESM1, GFDL-SPEAR, and CESM2 until  $\sim 2040$ ) were associated with a higher S2N ratio in the future, which led to an overall more deterministic climate system

and increased potential for significant forecast skill. The higher S2N ratio resulted from a larger ensemble mean forecast anomaly (i.e., signal), owing to ENSO’s role as a bigger “hammer” to the climate system. The opposite was true for models with decreasing ENSO trends (e.g., GFDL-ESM2M and CESM2 after  $\sim 2040$ ). Overall, when averaged across models, a 10% increase in ENSO amplitude was associated with a 14% and 7% increase in globally averaged correlation skill and S2N, respectively, at 12-month lead.

Given the large model disagreement in projected changes to both ENSO and global predictability, as well as the notable shortcomings by these and other climate models in capturing observed historical changes in ENSO amplitude (e.g., see our Fig. 1 as well as Maher et al. 2023; Wills et al. 2022), we cannot conclude which of the projected outcomes is most likely to occur. However, the relationship between ENSO amplitude and forecast skill is robust and improves our understanding of how future variations in ENSO characteristics may lead to changes in forecast skill. This lends nicely to the concept of climate change “storylines” (Shepherd et al. 2018) to further summarize the possible outcomes. For example, ENSO amplitude has increased since 1970 (Fig. 1 and Table 1), and if this represents a forced response that will persist into the future, then our results suggest that regions strongly influenced by ENSO and its teleconnections may become more predictable as these portions of the climate system become more deterministic. However, if ENSO variability were to instead decrease in the future as suggested by several recent studies (Wengel et al. 2021; Callahan et al. 2021; Peng et al. 2024), then historical forecast skill relationships that depend on ENSO and its teleconnections may become less reliable as these regions become less deterministic.

Of course, our study is based upon “perfect model” predictability, which is sometimes a reasonable proxy for “actual” skill (i.e., the skill of real-world predictions; Wheeler et al. 2017; Newman and Sardeshmukh 2017), but this may not always be the case (e.g., Kumar et al. 2014; Weisheimer et al. 2022; Scaife and Smith 2018). Still, previous studies have shown that multidecadal variations in actual skill are closely tied to observed variations in ENSO amplitude (Weisheimer et al. 2022; Lou et al. 2023), which is consistent with our conclusions and provides some confidence that our analysis of potential skill is relevant for real-world predictability. However, our results also make plain that the ongoing issues in ENSO simulation across multiple generations of climate models inhibit our ability not just to make useful projections of future tropical Pacific climate variability but also to constrain the expected trajectory of projected global climate predictability changes. Future model development is therefore needed to improve the realism of climate models to better capture observed ENSO variability and its multidecadal changes.

While our analysis takes an important first step toward understanding future climate predictability changes, there are a number of important questions that remain. First, is there a strong seasonality to future global predictability changes? Our study focused primarily on potential skill computed across all months; however, there were some seasonal differences in ENSO predictability changes (Fig. 2). Additionally,

Maher et al. (2023) showed that ENSO amplitude changes in the LEs analyzed here are stronger in some seasons (typically boreal winter) than others (see their Fig. 4). Therefore, it is possible that ENSO's impact on future predictability may be seasonally dependent. Next, what other ENSO-related factors impact future climate predictability? Many studies have shown that ENSO frequency (e.g., Berner et al. 2020), flavor (i.e., central vs eastern Pacific; Capotondi et al. 2015), and asymmetry (i.e., the duration of El Niño vs La Niña events; Maher et al. 2023) may change in the future. Changes to these characteristics may alter ENSO's influence on the rest of the climate system and thereby climate predictability. Additionally, there may be changes in the background mean state (e.g., the strength of the east–west temperature gradient in the equatorial Pacific) that impact the overall climate response to ENSO (Cai et al. 2021). While we did not find a significant relationship between predictability in our forecasts and each LE's time-varying ENSO frequency or flavor preference (not shown), we encourage future studies to investigate these mechanisms in more detail.

Although ENSO is a dominant driver of seasonal forecast skill for much of the globe, there are likely other mechanisms that contribute to the predictability limits of different regions and variables. For example, Shi et al. (2022) showed that long-term shoaling of the mixed layer in the future may reduce the thermal inertia of the ocean, thereby decreasing ocean memory and year-to-year SST persistence, especially in the midlatitudes. Similarly, Kumar et al. (2023) found that global warming decreases soil moisture memory over North America due to an increase in potential evapotranspiration. In both cases, the reduction in climate memory increases variability at less predictable high frequencies (e.g., weather time scales) while decreasing variability at lower frequencies (e.g., seasonal and longer), thus “whitening” the power spectrum and contributing to a decrease in persistence-related predictability. However, it is still unclear to what extent these changes may be offset by the dynamical drivers of predictability change related to ENSO. More research is needed to unpack the dynamic versus thermodynamic contributions to future climate predictability change.

**Acknowledgments.** We thank Friedrich Burger and Thomas Frölicher for providing us with the GFDL-ESM2M data used in this study. We also thank two anonymous reviewers for their insightful comments and suggestions that improved the quality of this work. This research was supported in part by the Australian Research Council Discovery Early Career Researcher Award DE230100315 and the NOAA Cooperative Agreement NA22OAR4320151. The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation under Cooperative Agreement 1852977.

**Data availability statement.** Large ensemble datasets are available as follows: CESM1-LE (<https://www.cesm.ucar.edu/projects/community-projects/MMLEA/>); CESM2-LE (<https://www.cesm.ucar.edu/projects/community-projects/LENS2/>); GFDL-SPEAR ([https://www.gfdl.noaa.gov/spear\\_large\\_ensembles/](https://www.gfdl.noaa.gov/spear_large_ensembles/));

GFDL-ESM2M, provided by Friedrich Burger and Thomas Frölicher at the University of Bern; MPI-GE (<https://esgf-metagrid.cloud.dkrz.de/>); ERSSTv5, COBE, and COBE2 (<https://psl.noaa.gov/data/gridded/tables/sst.html>); and HadISST data (<https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>).

## REFERENCES

- Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850, [https://doi.org/10.1175/1520-0493\(1987\)115<1825:OALOMA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2).
- Barnston, A. G., and M. K. Tippett, 2017: Do statistical pattern corrections improve seasonal climate predictions in the North American multimodel ensemble models? *J. Climate*, **30**, 8335–8355, <https://doi.org/10.1175/JCLI-D-17-0054.1>.
- , —, M. L. L'Heureux, S. Li, and P. D. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, <https://doi.org/10.1175/BAMS-D-11-00111.1>.
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, <https://doi.org/10.1175/JCLI-D-13-00597.1>.
- Berner, J., H. M. Christensen, and P. D. Sardeshmukh, 2020: Does ENSO regularity increase in a warming climate? *J. Climate*, **33**, 1247–1259, <https://doi.org/10.1175/JCLI-D-19-0545.1>.
- Beverly, J. D., M. Newman, and A. Hoell, 2023: Rapid development of systematic ENSO-related seasonal forecast errors. *Geophys. Res. Lett.*, **50**, e2022GL102249, <https://doi.org/10.1029/2022GL102249>.
- Burger, F. A., J. G. John, and T. L. Frölicher, 2020: Increase in ocean acidity variability and extremes under increasing atmospheric CO<sub>2</sub>. *Biogeosciences*, **17**, 4633–4662, <https://doi.org/10.5194/bg-17-4633-2020>.
- , J. Terhaar, and T. L. Frölicher, 2022: Compound marine heatwaves and ocean acidity extremes. *Nat. Commun.*, **13**, 4722, <https://doi.org/10.1038/s41467-022-32120-7>.
- Cai, W., and Coauthors, 2021: Changing El Niño–Southern Oscillation in a warming climate. *Nat. Rev. Earth Environ.*, **2**, 628–644, <https://doi.org/10.1038/s43017-021-00199-z>.
- Callahan, C. W., C. Chen, M. Rugenstein, J. Bloch-Johnson, S. Yang, and E. J. Moyer, 2021: Robust decrease in El Niño/Southern Oscillation amplitude under long-term warming. *Nat. Climate Change*, **11**, 752–757, <https://doi.org/10.1038/s41558-021-01099-2>.
- 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>.
- Cash, B. A., and N. J. Burls, 2019: Predictable and unpredictable aspects of U.S. West Coast rainfall and El Niño: Understanding the 2015/16 event. *J. Climate*, **32**, 2843–2868, <https://doi.org/10.1175/JCLI-D-18-0181.1>.
- Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, **428**, 733–736, <https://doi.org/10.1038/nature02439>.
- Delworth, T. L., and Coauthors, 2020: SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001895, <https://doi.org/10.1029/2019MS001895>.

- Derome, J., H. Lin, and G. Brunet, 2005: Seasonal forecasting with a simple general circulation model: Predictive skill in the AO and PNA. *J. Climate*, **18**, 597–609, <https://doi.org/10.1175/JCLI-3289.1>.
- Ding, H., and M. A. Alexander, 2023: Multi-year predictability of global sea surface temperature using model-analogs. *Geophys. Res. Lett.*, **50**, e2023GL104097, <https://doi.org/10.1029/2023GL104097>.
- , M. Newman, M. A. Alexander, and A. T. Wittenberg, 2018: Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *J. Climate*, **31**, 5437–5459, <https://doi.org/10.1175/JCLI-D-17-0661.1>.
- , —, —, and —, 2019: Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.*, **46**, 1721–1730, <https://doi.org/10.1029/2018GL080598>.
- Gan, B., L. Wu, F. Jia, S. Li, W. Cai, H. Nakamura, M. A. Alexander, and A. J. Miller, 2017: On the response of the Aleutian low to greenhouse warming. *J. Climate*, **30**, 3907–3925, <https://doi.org/10.1175/JCLI-D-15-0789.1>.
- Giannini, A., Y. Kushnir, and M. A. Cane, 2001: Seasonality in the impact of ENSO and the North Atlantic high on Caribbean rainfall. *Phys. Chem. Earth*, **26B**, 143–147, [https://doi.org/10.1016/S1464-1909\(00\)00231-8](https://doi.org/10.1016/S1464-1909(00)00231-8).
- Gu, D., and S. G. H. Philander, 1997: Interdecadal climate fluctuations that depend on exchanges between the tropics and extratropics. *Science*, **275**, 805–807, <https://doi.org/10.1126/science.275.5301.805>.
- Gupta, A., A. C. Pandey, and A. K. Mitra, 2023: Impact of atmospheric response of ENSO and IOD on boreal summer rainfall variability over the Indian Ocean in coupled models. *Climate Dyn.*, **61**, 4107–4124, <https://doi.org/10.1007/s00382-023-06796-6>.
- He, S., J.-Y. Yu, S. Yang, and S.-W. Fang, 2020: ENSO's impacts on the tropical Indian and Atlantic Oceans via tropical atmospheric processes: Observations versus CMIP5 simulations. *Climate Dyn.*, **54**, 4627–4640, <https://doi.org/10.1007/s00382-020-05247-w>.
- Heede, U. K., and A. V. Fedorov, 2023: Towards understanding the robust strengthening of ENSO and more frequent extreme El Niño events in CMIP6 global warming simulations. *Climate Dyn.*, **61**, 3047–3060, <https://doi.org/10.1007/s00382-023-06856-x>.
- 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., 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](https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2).
- Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, <https://doi.org/10.1175/JCLI-D-16-0836.1>.
- Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses 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>.
- Jacox, M. G., M. A. Alexander, C. A. Stock, and G. Hervieux, 2019: On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. *Climate Dyn.*, **53**, 7519–7533, <https://doi.org/10.1007/s00382-017-3608-y>.
- 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>.
- Kumar, A., 2009: Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Mon. Wea. Rev.*, **137**, 2622–2631, <https://doi.org/10.1175/2009MWR2814.1>.
- , P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, <https://doi.org/10.1175/MWR-D-13-00287.1>.
- Kumar, S., C. F. Dewes, M. Newman, and Y. Duan, 2023: Robust changes in North America's hydroclimate variability and predictability. *Earth's Future*, **11**, e2022EF003239, <https://doi.org/10.1029/2022EF003239>.
- Lenssen, N., P. DiNezio, L. Goddard, C. Deser, Y. Kushnir, S. J. Mason, M. Newman, and Y. Okumura, 2024: Strong El Niño events lead to robust multi-year ENSO predictability. *Geophys. Res. Lett.*, **51**, e2023GL106988, <https://doi.org/10.1029/2023GL106988>.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Lou, J., M. Newman, and A. Hoell, 2023: Multi-decadal variation of ENSO forecast skill since the late 1800s. *npj Climate Atmos. Sci.*, **6**, 89, <https://doi.org/10.1038/s41612-023-00417-z>.
- MacLeod, D., C. O'Reilly, T. Palmer, and A. Weisheimer, 2018: Flow dependent ensemble spread in seasonal forecasts of the boreal winter extratropics. *Atmos. Sci. Lett.*, **19**, e815, <https://doi.org/10.1002/asl.815>.
- 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>.
- , and Coauthors, 2023: The future of the El Niño–Southern Oscillation: Using large ensembles to illuminate time-varying responses and inter-model differences. *Earth Syst. Dyn.*, **14**, 413–431, <https://doi.org/10.5194/esd-14-413-2023>.
- McGregor, S., C. Cassou, Y. Kosaka, and A. S. Phillips, 2022: Projected ENSO teleconnection changes in CMIP6. *Geophys. Res. Lett.*, **49**, e2021GL097511, <https://doi.org/10.1029/2021GL097511>.
- Menary, M. B., J. Mignot, and J. Robson, 2021: Skillful decadal predictions of subpolar North Atlantic SSTs using CMIP model-analogs. *Environ. Res. Lett.*, **16**, 064090, <https://doi.org/10.1088/1748-9326/ac06fb>.
- Mo, K. C., and M. Ghil, 1987: Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.*, **44**, 877–902, [https://doi.org/10.1175/1520-0469\(1987\)044<0877:SADOPA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1987)044<0877:SADOPA>2.0.CO;2).
- Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.*, **44**, 8520–8529, <https://doi.org/10.1002/2017GL074088>.
- , A. T. Wittenberg, L. Cheng, G. P. Compo, and C. A. Smith, 2018: The extreme 2015/16 El Niño, in the context of historical climate variability and change. *Bull. Amer. Meteor. Soc.*, **99**, S16–S20, <https://doi.org/10.1175/BAMS-D-17-0116.1>.
- O'Brien, J. P., and C. Deser, 2023: Quantifying and understanding forced changes to unforced modes of atmospheric circulation variability over the North Pacific in a coupled model large ensemble. *J. Climate*, **36**, 19–37, <https://doi.org/10.1175/JCLI-D-22-0101.1>.

- O'Reilly, C. H., J. Heatley, D. MacLeod, A. Weisheimer, T. N. Palmer, N. Schaller, and T. Woollings, 2017: Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century. *Geophys. Res. Lett.*, **44**, 5729–5738, <https://doi.org/10.1002/2017GL073736>.
- , T. Woollings, L. Zanna, and A. Weisheimer, 2019: An interdecadal shift of the extratropical teleconnection from the tropical Pacific during boreal summer. *Geophys. Res. Lett.*, **46**, 13 379–13 388, <https://doi.org/10.1029/2019GL084079>.
- Peng, Q., S.-P. Xie, and C. Deser, 2024: Collapsed upwelling projected to weaken ENSO under sustained warming beyond the twenty-first century. *Nat. Climate Change*, **14**, 815–822, <https://doi.org/10.1038/s41558-024-02061-8>.
- Planton, Y. Y., and Coauthors, 2021: Evaluating climate models with the CLIVAR 2020 ENSO metrics package. *Bull. Amer. Meteor. Soc.*, **102**, E193–E217, <https://doi.org/10.1175/BAMS-D-19-0337.1>.
- Quan, X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of U.S. seasonal forecast skill. *J. Climate*, **19**, 3279–3293, <https://doi.org/10.1175/JCLI3789.1>.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. 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>.
- Rodgers, K. B., and Coauthors, 2021: Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.*, **12**, 1393–1411, <https://doi.org/10.5194/esd-12-1393-2021>.
- Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate*, **13**, 4268–4286, [https://doi.org/10.1175/1520-0442\(2000\)013<4268:COPAWE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4268:COPAWE>2.0.CO;2).
- Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, <https://doi.org/10.1038/s41612-018-0038-4>.
- Shepherd, T. G., and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, **151**, 555–571, <https://doi.org/10.1007/s10584-018-2317-9>.
- Shi, H., and Coauthors, 2022: Global decline in ocean memory over the 21st century. *Sci. Adv.*, **8**, eabm3468, <https://doi.org/10.1126/sciadv.abm3468>.
- Shi, W., N. Schaller, D. MacLeod, T. N. Palmer, and A. Weisheimer, 2015: Impact of hindcast length on estimates of seasonal climate predictability. *Geophys. Res. Lett.*, **42**, 1554–1559, <https://doi.org/10.1002/2014GL062829>.
- Suarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. *J. Atmos. Sci.*, **45**, 3283–3287, [https://doi.org/10.1175/1520-0469\(1988\)045<3283:ADAOFE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2).
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Weisheimer, A., N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer, 2017: Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quart. J. Roy. Meteor. Soc.*, **143**, 917–926, <https://doi.org/10.1002/qj.2976>.
- , D. Decremmer, D. MacLeod, C. O'Reilly, T. N. Stockdale, S. Johnson, and T. N. Palmer, 2019: How confident are predictability estimates of the winter North Atlantic Oscillation? *Quart. J. Roy. Meteor. Soc.*, **145**, 140–159, <https://doi.org/10.1002/qj.3446>.
- , D. J. Belfort, D. MacLeod, T. Palmer, C. O'Reilly, and K. Strømmen, 2020: Seasonal forecasts of the twentieth century. *Bull. Amer. Meteor. Soc.*, **101**, E1413–E1426, <https://doi.org/10.1175/BAMS-D-19-0019.1>.
- , M. A. Balmaseda, T. N. Stockdale, M. Mayer, S. Sharmila, H. Hendon, and O. Alves, 2022: Variability of ENSO forecast skill in 2-year global reforecasts over the 20th century. *Geophys. Res. Lett.*, **49**, e2022GL097885, <https://doi.org/10.1029/2022GL097885>.
- , and Coauthors, 2024: The signal-to-noise paradox in climate forecasts: Revisiting our understanding and identifying future priorities. *Bull. Amer. Meteor. Soc.*, **105**, E651–E659, <https://doi.org/10.1175/BAMS-D-24-0019.1>.
- Wengel, C., S.-S. Lee, M. F. Stuecker, A. Timmermann, J.-E. Chu, and F. Schloesser, 2021: Future high-resolution El Niño/Southern Oscillation dynamics. *Nat. Climate Change*, **11**, 758–765, <https://doi.org/10.1038/s41558-021-01132-4>.
- Wheeler, M. C., H. Zhu, A. H. Sobel, D. Hudson, and F. Vitart, 2017: Seamless precipitation prediction skill comparison between two global models. *Quart. J. Roy. Meteor. Soc.*, **143**, 374–383, <https://doi.org/10.1002/qj.2928>.
- Wills, R. C. J., Y. Dong, C. Proistosescu, K. C. Armour, and D. S. Battisti, 2022: Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophys. Res. Lett.*, **49**, e2022GL100011, <https://doi.org/10.1029/2022GL100011>.
- Zhang, L., T. L. Delworth, X. Yang, Y. Morioka, F. Zeng, and F. Lu, 2023: Skillful decadal prediction skill over the Southern Ocean based on GFDL SPEAR Model-Analogs. *Environ. Res. Commun.*, **5**, 021002, <https://doi.org/10.1088/2515-7620/acb90e>.
- Zhao, M., H. H. Hendon, O. Alves, G. Liu, and G. Wang, 2016: Weakened eastern Pacific El Niño predictability in the early twenty-first century. *J. Climate*, **29**, 6805–6822, <https://doi.org/10.1175/JCLI-D-15-0876.1>.
- Zheng, Y., M. Rugenstein, P. Pieper, G. Beobide-Arsuaga, and J. Baehr, 2022: El Niño–Southern Oscillation (ENSO) predictability in equilibrated warmer climates. *Earth Syst. Dyn.*, **13**, 1611–1623, <https://doi.org/10.5194/esd-13-1611-2022>.
- Zhou, Z.-Q., S.-P. Xie, X.-T. Zheng, Q. Liu, and H. Wang, 2014: Global warming-induced changes in El Niño teleconnections over the North Pacific and North America. *J. Climate*, **27**, 9050–9064, <https://doi.org/10.1175/JCLI-D-14-00254.1>.