Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL104097

Key Points:

- Sea surface temperature anomalies are predictable in El Niño Southern Oscillation (ENSO) regions at leads of two or even 3 years
- The maximum forecast lead time of ENSO events is obtained when forecasts are initialized in July
- Some, including single-year, ENSO events are predictable at leads of up to 30 and 24 months for the cold and warm phases, respectively

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Ding,
hui.ding@noaa.gov

Citation:

Ding, H., & Alexander, M. A. (2023). Multi-year predictability of global sea surface temperature using model-analogs. *Geophysical Research Letters*, *50*, e2023GL104097. https://doi.org/10.1029/2023GL104097

Received 12 APR 2023
Accepted 5 SEP 2023

Multi-Year Predictability of Global Sea Surface Temperature Using Model-Analogs

Hui Ding[1,2] and Michael A. Alexander[2]

[1]CIRES, University of Colorado, Boulder, CO, USA, [2]NOAA Physical Sciences Laboratory, Boulder, CO, USA

Abstract The multi-year predictability of global sea surface temperature (SST) is examined by applying a model-analog method to four control simulations to make forecasts at leads of 1–36 months over 1961–2015. The forecasts are found to have skill for annual mean SST at Year 2 (i.e., leads of 13–24 months) or even Year 3 (i.e., leads of 25–36 months) in the tropical Pacific, the North and South Pacific, the southwest Indian Ocean, and the northwestern tropical Atlantic. The seasonality in forecast skill suggests that July is the best time to initialize multi-year forecasts. The evolution of forecast skill for tropical Pacific SSTs indicates that the predictability of some El Niño Southern Oscillation events is nearly the same at leads of 6 and 18 or 24 months. Further, these El Niño and La Niña events can be predicted at leads of up to 24 and 30 months, respectively.

Plain Language Summary Past studies showed that some La Niña events occur in two consecutive years, and the tropical Pacific sea surface temperature (SST) anomalies in the second year of these events can be predicted 2-years ahead. Here, we examine El Niño Southern Oscillation prediction skill out to 3 years regardless of its initial state. In addition, we evaluate multi-year forecast skill of global SSTs, which may be independent from El Niño and La Niña events. Forecasts are made using a analog method applied to pre-existing model simulations, which has been proven to be useful in making climate forecasts. Forecast skill is evaluated by performing retrospective forecasts of historical observations over 1961–2015 at lead times of 1–36 months. The forecasts are found to be skillful for predicting the evolution of some El Niño and La Niña events at lead months of 24 and 30 months, respectively. Furthermore, SSTs in the southwest Indian Ocean, the North and South Pacific, and the northwest tropical Atlantic are predictable two or even 3 years ahead. In this study, the effects of anthropogenic climate change on SST are also considered, and the inclusion of the effect greatly improves forecast skill in many ocean basins, but not in the central and eastern tropical Pacific.

1. Introduction

The El Niño Southern Oscillation (ENSO) is the dominant mode of interannual climate variability and predicting its occurrence and evolution has been the focus of climate forecasts for nearly four decades (e.g., Latif et al., 1998; Zebiak & Cane, 1987). Seasonal ENSO forecasts based on global coupled models exhibit skill at lead times of 6–12 months (Barnston et al., 2012; Becker et al., 2022; Kirtman et al., 2014). Outside the tropical Pacific, climate predictability can also result from local air-sea interaction (e.g., Saji et al., 1999; Xie et al., 2002) and ocean dynamics (e.g., Alexander & Deser, 1995) in addition to ENSO teleconnections (Alexander et al., 2002; Trenberth et al., 1998).

Multi-year climate predictability, however, has only been examined by a few studies using coupled general circulation models (CGCM; DiNezio et al., 2017; Ham et al., 2019; Luo et al., 2008; Sharmila et al., 2022; Weisheimer et al., 2022; Wu et al., 2021; Yeager et al., 2022). In this study, multi-year forecasts are regarded as forecasts at leads of 1–3 years to distinguish them from decadal predictions (e.g., Yeager et al., 2022). Theory suggests that ENSO may be predictable 2 years in advance (e.g., Neelin et al., 1998). Luo et al. (2008) first demonstrated skillful forecasts of ENSO events at leads of 2 years over the years 1982–2004. The duration of ENSO events is asymmetrical because some La Niña events last for two continuous years (Okumura & Deser, 2010). Wu et al. (2019) further showed that one out of three (two) El Niño (La Niña) events persist for two consecutive years. DiNezio et al. (2017) found that the second peak of 2-year La Niña events are predictable at lead time of 2 years using the Decadal Prediction Large Ensemble (DPLE; Yeager et al., 2018 and Text S1 in Supporting Information S1) conducted by Community Earth System Model, version 1 (CESM1). Ham et al. (2019) demonstrated that ENSO hindcast skill at 18-month lead using a machine learning method. Wu et al. (2021) further showed that multi-year El Niño and La Niña events are well captured by a set of hindcast experiments at leads of 13 and 25 months,

respectively. Weisheimer et al. (2022) found that ENSO hindcast skill is much higher since 1960s than the previous four decades. However, these studies often only focus on ENSO, leaving multi-year predictability in other ocean basins largely unexplored (Yeager et al., 2022).

In this study, we examine the overall multi-year sea surface temperature (SST) predictability by applying the model-analog method (Ding et al., 2018; hereafter D18) to four NMME control simulations (e.g., Kirtman et al., 2014; Table S1 in Supporting Information S1) to make global SST hindcasts at leads of 1–36 months. The model-analog method finds a set of states from a long uninitialized control CGCM simulation by matching their sea surface height (SSH) and SST anomalies to observed global SSH and SST anomalies (SSTA). The matched states are called model-analogs by D18, and their subsequent evolutions in the control simulation provides an ensemble of forecasts. The method reproduces and occasionally exceeds traditional data assimilation-initialized forecast skill when being applied to the same CGCMs (D18). We also evaluate the impact of the external greenhouse gas-forced trends on the overall forecast skill following Ding et al. (2019; hereafter D19), but using a larger domain and examining longer lead times.

2. Data and Methods

2.1. Model-Analog Technique

We choose analogs at each month t over the years 1961–2015 by calculating a root-mean-square (RMS) distance between the target state and each library state, where the target state is defined as the observed state at the initialization time, and the library state is obtained from a CGCM control simulation (D18). SST and SSH anomalies between 60°S and 60°N are used to compute the RMS distance to define analogs (for details see D18). Distances are then ranked in ascending order, and the K states closest to the target state are chosen as the model-analog ensemble members. In this study, we set $K = 15$ for each model in Table S1 of Supporting Information S1 unless otherwise stated. The subsequent model evolution of this ensemble is the model-analog forecast ensemble.

2.2. Model and Observational Data Sets

The library data sets consist of monthly mean data from control simulations conducted using four different models (listed in Table S1 of Supporting Information S1), which have demonstrated seasonal forecast skill from traditional assimilation-initialized hindcasts (D18). Model-analog hindcasts based on these models exhibit comparable SST forecast skill as initialized hindcasts using the same models, which compelled us to examine their multi-year predictability. For each simulation, monthly anomalies are calculated by subtracting its own monthly climatology determined using the full length of data (Table S1 in Supporting Information S1). Additionally, the CESM1 large ensemble (CESM-LE; Text S3 in Supporting Information S1) provide an additional library for model-analog forecasts, which are compared against the DPLE (Yeager et al., 2018).

SSTs from the monthly mean Hadley Sea Ice and Sea Surface Temperature v1.1 data set (HadISST; Rayner et al., 2003) and SSHs from the ECMWF Ocean Reanalysis System 4 data set (Balmaseda et al., 2013) were used to determine initial observed states over the years 1960–2015, a common period for the data sets. Observed anomalies were calculated by removing the monthly mean 1960–2015 climatology. All model and observed data were interpolated onto a common 2° longitude by 2° latitude grid.

2.3. Accounting for Externally-Forced Trends

The observed SSH and SST anomalies include both externally-forced (greenhouse gas, aerosols, and other radiative forcing) components and internal climate variability (e.g., Solomon et al., 2011). However, fixed-climate control simulations (e.g., with preindustrial or late twentieth century forcings) retain only the model's internal climate variability. Therefore, the model-analog technique predicts the internal climate anomaly. Following D19, the time-evolving multi-model ensemble mean of 45 historical and RCP4.5 CMIP5 simulations was adopted to estimate the externally-forced component, which was then used to remove the trend component from the observed SSH and SST anomalies. Readers can refer to Text S4 in Supporting Information S1 and D19 for more details.

3. Results

We first assess the skill of the multi-model ensemble-mean model-analog hindcasts alone, without the projected externally-forced component. The ensemble-mean hindcast skill for annual mean SST at leads of 13–24 months

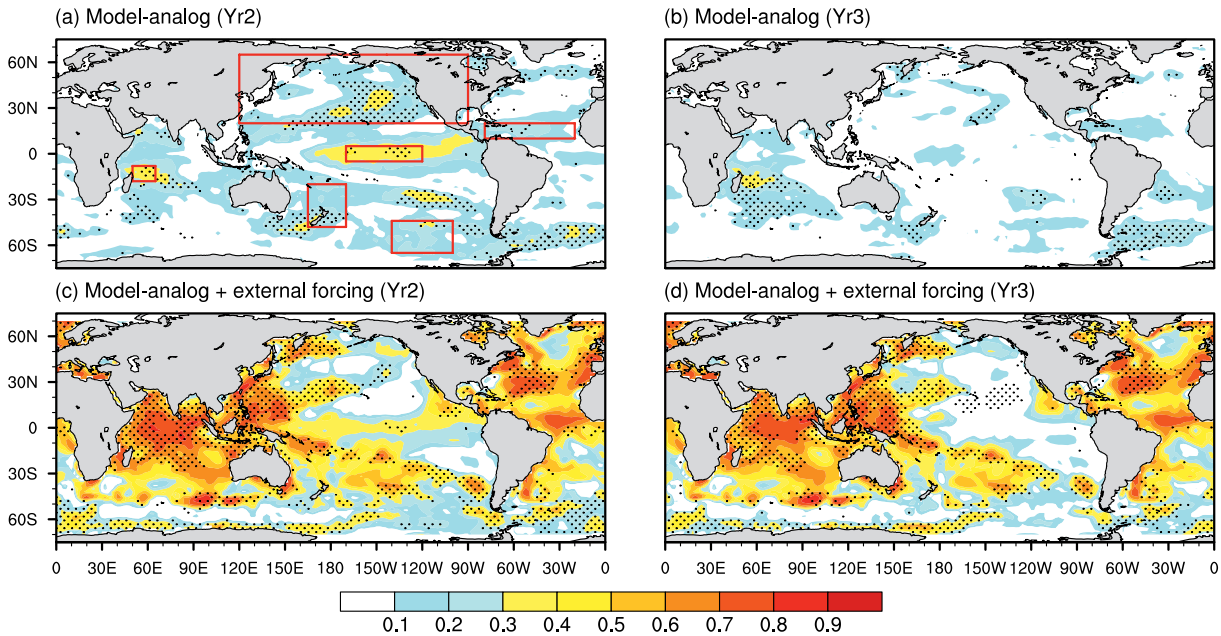

Figure 1. Skill of "Year 2" and "Year 3" hindcasts verified against observations, 1961–2015. Skill is measured by the local anomaly correlation between annual mean hindcast and observed sea surface temperature anomalies, determined for (a) "Year 2" (Months 13–24) and (b) "Year 3" (Months 25–36). (c and d) Same as (a) and (b) but for hindcasts *with* the CMIP5-projected response to external radiative forcing considered. The red boxes in (a) indicates the regions for defining the five indices shown in Figure 2. Dots indicate where the anomaly correlation coefficient is significant at 95% confidence, obtained by considering the effective degrees of freedom in a Student's *t*-test.

(i.e., Yr2 hindcast skill), as measured by anomaly correlation coefficient (ACC) is shown in Figure 1a. The hindcasts are skillful in the central and eastern equatorial Pacific, with correlations ranging between 0.3 and 0.4. This suggests at least 2-year predictability of ENSO events, consistent with previous studies (e.g., Wu et al., 2021). The relatively low ACC skill of 0.3–0.4 is for forecasts initialized from all months while forecast skill strongly depends on seasonality (see below). In addition, the ACC skill of 0.3–0.4 is also noted in the southwest Indian Ocean, the North and South Pacific (Figure 1a). At Yr3 (Figure 1b), ACC values ≥0.3 occur in the southwest Indian Ocean. The hindcast skill for SSH at Yr2 and Yr3 (Figure S1 in Supporting Information S1) closely resemble that for SST. In particular, the ACC skill for SSH reaches ~0.5 in the Southwest Indian Ocean between 20° and 10°S. These results indicate that the model-analog method still generates skillful forecasts for SST when it is defined using the global ocean compared to previous studies based only on the tropical Indo-Pacific sector (D18; D19).

We now evaluate how external forcing (i.e., greenhouse gas, aerosols or volcanic aerosols) impacts multi-year hindcast skill by including the externally-forced trend component, obtained from the CMIP5 historical run multi-model ensemble mean (see the supplementary material), in Figures 1c and 1d. Including the externally-forced component, as shown by comparing the upper and lower rows in Figure 1, greatly improves model-analog SST forecast skill over the tropical Indian Ocean, the western tropical Pacific, South Pacific and large areas in the North Atlantic (AC of 0.6–0.8), but not within the ENSO region. The inclusion of the effect of greenhouse gas forcing also increases forecast skill for SSH in large areas in the extratropical Indian and Pacific Oceans and nearly the entire Atlantic Ocean (Figure S1 in Supporting Information S1).

We also compare the skill of model-analog hindcasts to those calculated from the DPLE. Figure S2 in Supporting Information S1 shows the hindcast skill at leads of Yr1, Yr2, and Yr3 from the DPLE and the CESM-LE based model-analog ensemble-mean hindcasts with and without external forcing. The DPLE and the model-analog hindcasts are initialized from November and October, respectively, for a fair comparison. The model-analog hindcasts alone (right column) reproduce or even slightly exceed the DPLE forecast skill for the tropical Pacific SST for the three leads, but the addition of external forcing (middle column) enables the model-analog hindcasts to better match the DPLE forecast skill in the Indian Ocean, the South and North Pacific, and the tropical and North Atlantic. This is consistent with the impact of greenhouse gas forcing on forecast skill in Figure 1. In the northern

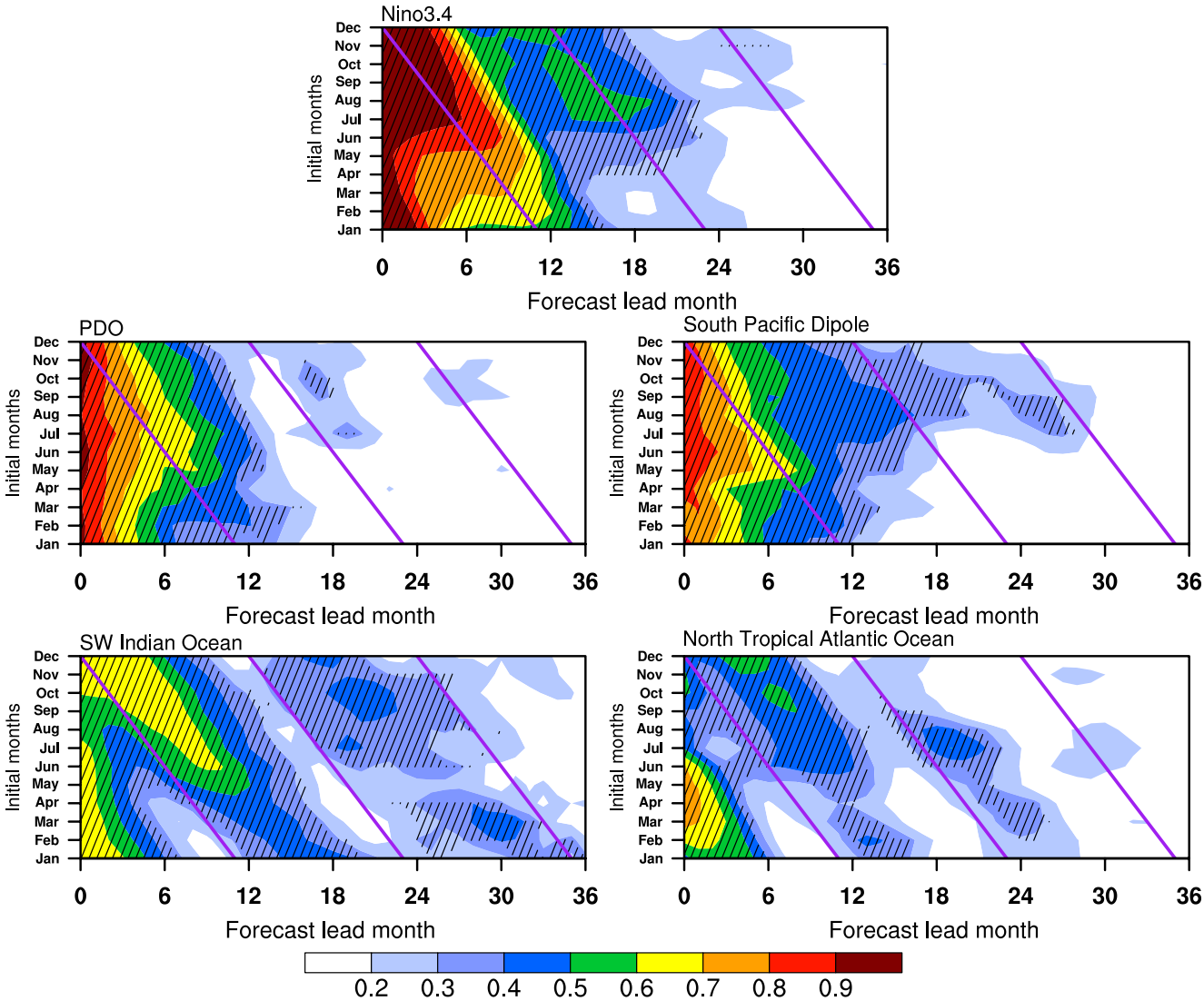

Figure 2. Anomaly correlation coefficient (ACC) skill of 3-month running mean multi-model ensemble mean hindcasts over the years 1961–2015. Shown is skill for the SSTA indices: (a) Nino3.4 (170°–120° W, 5°S–5° N), (b) PDO, defined by projecting sea surface temperature onto the first empirical orthogonal function mode in the North Pacific (20°–65°N, 120°E–90°W), (c) South Pacific Ocean Dipole ((165°E–170°W, 20°–48°S)—(140°–100°W, 44°–65°S); Guan et al., 2014), (d) SW Indian Ocean (50°–65°E, 8°–18°S), and (e) North Tropical Atlantic (79°–20°W, 10°–20°N), as a function of forecast lead (abscissa) and initial month (ordinate). For example, the skill for hindcasts initialized in August at a lead of 18 months is verified by the observed SSTA averaged in Jan/Feb/Mar 1 year and half later. Hatching indicates that ACC is significant at the 95% confidence. Purple lines indicate forecasts of December.

North Atlantic, model-analog forecast skill is much lower than the DPLE skill in Yr2 and Yr3, suggesting that model-analog predictions based on just SST and SSH may not fully represent variability driven by the Atlantic meridional overturning circulation, as well as other processes (Yeager, 2020).

From now on we will focus on predictability without including the effects of greenhouse gas forcing. Since hindcasts are initialized from each month, we can evaluate how forecast skill depends on the forecast lead and initial state for monthly SSTA in five regions (Figure 2) in which the initialized forecast skill at Yr2 is relatively high (Figure 1a). The SSTA averaged over the Nino3.4 region is well forecasted (ACC $\geq$0.6) at lead times of 1–12 months, consistent with the current operational forecast skill (Becker et al., 2022; Weisheimer et al., 2022; Yeager et al., 2022) and previous model-analog hindcast skill (D19). In addition, the forecast skill for the Nino3.4 SSTA displays the well-documented spring predictability barrier (e.g., McPhaden, 2003). At lead times of 13–18 months, ACC $\geq$0.5 is noted for forecasts initialized in July-December, suggesting that ENSO is predictable at leads beyond 1 year. Here, it is regarded as potential useful forecast skill when the ACC equals or

exceeds 0.5 (Yeager et al., 2022). It also suggests that on average, ENSO events can be predicted by forecasts initialized from the previous summer (e.g., JJA) at a lead time of 18 months, several months longer than in previous studies that initialized forecasts in November (e.g., Weisheimer et al., 2022). Lou et al. (2023) also noted the seasonality of ENSO multi-year forecast skill by applying the model-analog method to CMIP6 simulations. The seasonality of PDO forecast skill closely resembles that of the Nino3.4 SSTA, although the PDO forecast skill is much lower. The maximum ACC skill for the PDO in the second year occurs 2-months later than that for ENSO, consistent with a portion of the PDO being driven by ENSO (e.g., Newman et al., 2003) via the atmosphere bridge (Alexander et al., 2002). Likewise, the South Pacific Dipole displays a similar pattern of ACC skill as those of ENSO and the PDO. The ACC skill of the South Pacific Dipole is about 0.4 (0.30) at lead times of 12–24 (25–30) months when initialized in August. The SSTA in the Southwest Indian Ocean displays potential predictability during years 2 and 3, which may result from remote ENSO forcing and local air-sea interaction (Xie et al., 2002). The pattern of seasonality indicates that the multi-year skill mainly comes from SSTA being verified in July-September. For example, SSTA forecasts initialized in Jan-Mar that verify in July-September at lead of 30 months have ACC values of 0.3–0.4. There is also some predictability at longer lead times (13–24 months) in the northern tropical Atlantic. For example, ACC values reach 0.4 at a lead time of about 20 months when hindcasts are initialized from July, indicating that the skill comes from predicting SSTA in boreal spring. Short-term forecast skill is also greatest for predictions of SSTA in spring. These results are consistent with studies (e.g., Enfield & Mayer, 1997) that found that SST variability in the northern tropical Atlantic is strongly affected by ENSO in spring and summer. In addition, the Atlantic Meridional Mode may also contribute to climate predictability beyond 1 year (Amaya et al., 2017; Chang et al., 1997).

We now examine the evolution of ENSO predictability over the past 55 years (Figure 3). The hindcast skill is measured based on the pattern correlation of hindcasts and observed anomalies within the ENSO region for the multi-model model-analog ensemble mean. This skill metric has been employed in previous studies (e.g., Newman & Sardeshmukh, 2017), since the SSTA pattern is an important metric for ENSO (e.g., Capotondi et al., 2015). The evolution of Month 6 hindcast skill is shown in black curve (Figure 3). Consistent with D19, there is a close correspondence between the modulation of 6-month lead hindcast skill curve and ENSO activity (Figure 3). This suggests that higher (lower) hindcast skill is coincident with stronger (weaker) ENSO activity and that there is no obvious trend in the Month 6 hindcast skill over the entire period, also noted by previous studies (e.g., Kumar et al., 2015; D19).

We examine the evolution of 2-year predictability for ENSO by comparing pattern correlation hindcast skill at lead times of 12, 18, and 24 months with that at 6 months (Figure 3). For clarity, only correlations greater than 0.4 are shown for the longer lead times while the entire evolution is shown in Figure S3 of Supporting Information S1. Like the 6-month forecasts, the hindcast skill at the longer lead times also depend on the level of ENSO activity. For some ENSO events, the pattern correlation hindcast skill is very high (0.7–0.8) at both short (6-month) and longer leads of 18 and even 24 months. Here, we define the ensemble mean hindcasts as skillful when the pattern correlation equals or exceeds 0.75 in the seasonal mean of December and the next January and February. The events that are successfully predicted at long leads are 1965/66 (18), 1972/73 (24), 1982/83 (18), 1986/87 (24), 1991/92 (18), 1997/98 (18), 2002/03 (24), 2009/10 (24) El Niño events and the 1984/85 (24), 1988/89 (18), 1998/99 (18), 1999/2000 (24) La Niña events, where the numbers in parentheses indicate the lead time in which the pattern correlation $\geq$0.75. The list includes 1-year ENSO events, for example, the 2009/2010 El Niño and 1988/89 La Niña events, in addition to 2-year ENSO events, for example, the 1986/87 El Niño and 1999/2000 La Niña events.

The multi-model ensemble mean forecasts of the 1999/2000 La Niña and 2009/2010 El Niño events are shown in Figure 4, for lead times of 6, 12, 18, 24, 30, and 36 months. The ensemble mean hindcast initialized in Jan 1997 underestimates the amplitude of the 1997/98 El Niño, although many individual ensemble members reach or even exceed ~2.5°C, the observed SSTA. The 1999/2000 La Niña event is successfully predicted by forecasts initialized in July 1997 at a lead time of 30 months, although the forecasts underestimate the observed magnitude. However, the observed Nino3.4 SSTA is within the spread of the forecast members, which is also the case at other lead times. The multi-model ensemble mean forecast reproduces the evolution of Nino3.4 SSTA from the initial state in July 1997 to January 2000, including the 1997/98 El Niño and the following 2-year La Niña event. However, the ensemble mean forecast does not reproduce the seasonal weakening and subsequent strengthening of negative SSTA in 1999. That the forecast lead time of nearly 30 months is obtained through initializing hindcasts from the developing phase of the previous El Niño event is consistent with Figure 2, and extends the lead

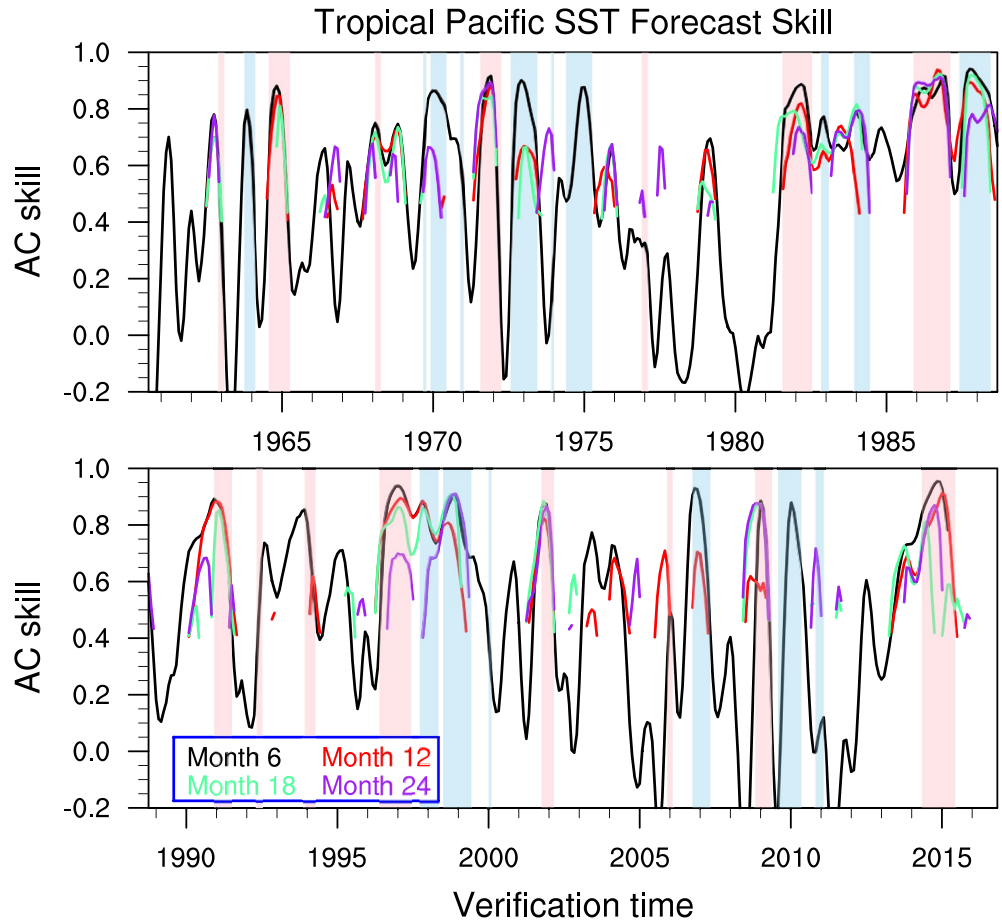

Figure 3. Evolution of model-analog ensemble-mean hindcast skill in the tropical Pacific El Niño Southern Oscillation region at lead times of 6, 12, 18, and 24 months. The measure of skill is the pattern correlation between the ensemble mean hindcasts and the corresponding observed 3-month running mean SSTA within the region bounded by 20°S–20°N, 170°E−70° W. For clarity, a 5-month running mean has been applied to the skill time series, and values below 0.4 are not shown for the three longer leads. The light red (blue) shaded epochs indicate periods of >1 standard deviation warm (<−1 standard deviation cold) Nino3.4 SSTA.

time of forecasts by 4 months relative to the forecasts by DiNezio et al. (2017) and Wu et al. (2021), who initialized forecasts near the peak of this El Niño event. Furthermore, the forecasts indicate that a La Niña event may occur with an ensemble mean SSTA of −0.5°C during the winter of 1999/2000 when the forecasts are initialized in January 1997 with a lead time of 36 months, although there is a large spread among the ensemble members. Surprisingly, the 12-month ensemble mean forecast failed to predict this La Niña event. For the 2009/2010 El Niño, the ensemble mean forecast reproduces the event with a SSTA of 1°C when initialized from January 2008 with a lead time of 24 month. For this La Niña event, the observed Nino3.4 SSTA is also within the spread of forecast members at the six lead times. The evolution of the forecasts of El Niño events (1972/73, 1986/87, 1987/88, 2002/03) and combined El Niño - La Niña events (1972/73, 1982/83, and 2015/16) are shown in Figures S4–S10 in Supporting Information S1. The hindcasts of these events exhibit various level of skill.

4. Conclusion and Discussion

In this study, we have examined the multi-year predictability of global SST, using a model-analog method (D18) applied to four pre-existing control simulations to initialize retrospective forecasts at leads of 1–36 months during 1961–2015. A difference is that initial observed SST and SSH anomalies are detrended prior to matching them to corresponding anomalies calculated from the control simulations, in the global oceans to define analogs, while in D18 and D19 these anomalies are matched only in the tropical Indo-Pacific. The analysis of the hindcasts

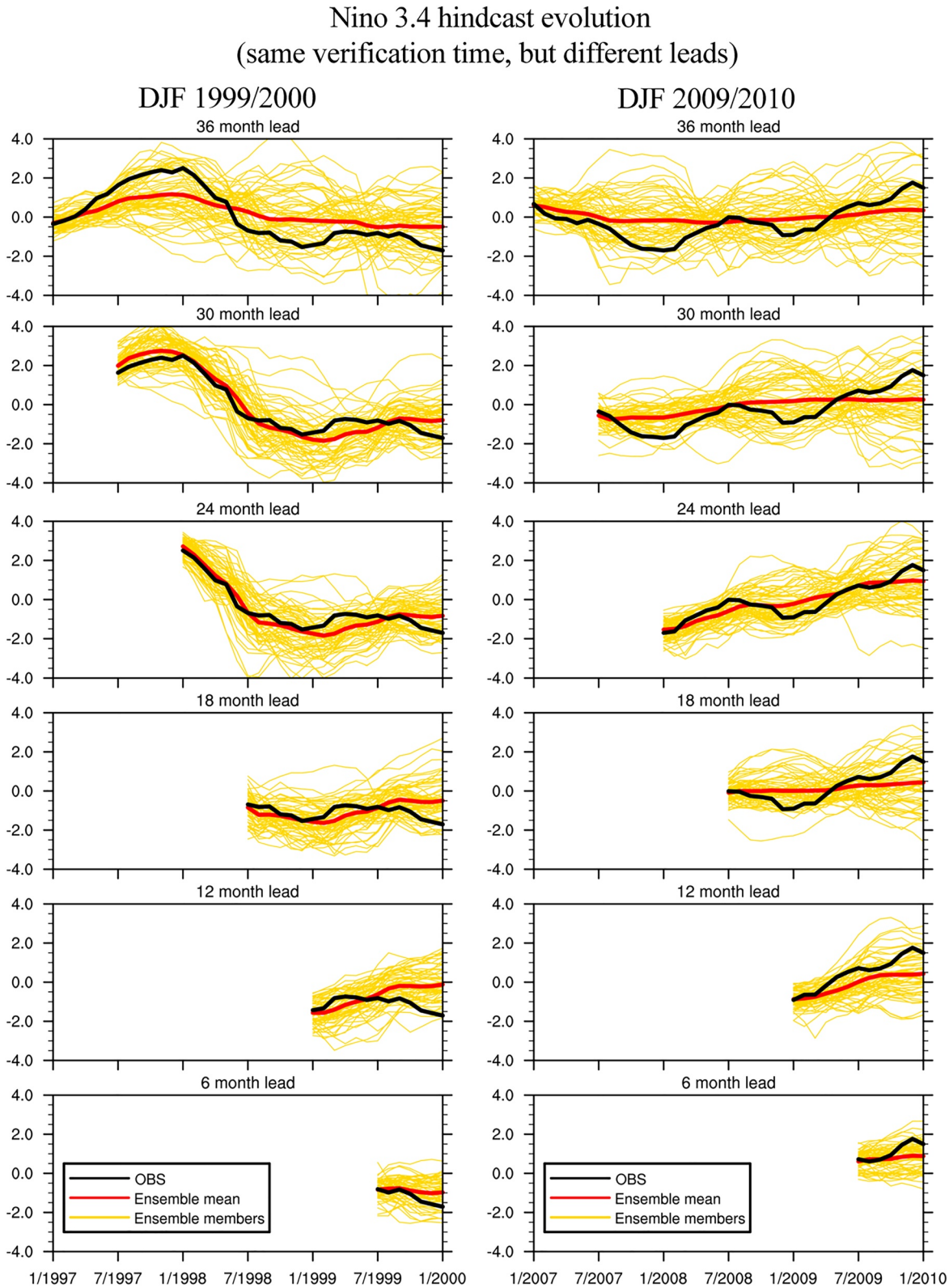

Figure 4. The model-analog hindcasts of SSTA in Nino3.4 domain at lead times of 6, 12, 18, 24, 30, and 36 months. The observation and multi-model ensemble mean forecasts are shown with thick black and red lines, respectively, while the thin yellow lines are ensemble members. All hindcasts verify against DJF 1999/2000 (left) or DJF 2009/2010 (right).

suggests predictability for SST in the equatorial Pacific, the North and South Pacific, the southwest Indian Ocean and northern tropical Atlantic at Year 2 (Months 13–24) and even at Year 3 (Months 25–36) in the southwest Indian Ocean. The predictability outside the tropical Pacific, may arise in part from ENSO's remote impact on other ocean basins via the atmospheric bridge (e.g., Alexander et al., 2002). The hindcast skill in those regions are strongly modulated by the season. In particular, the seasonality of ENSO forecast skill indicates that the ACC skill still reaches a value of 0.5 at 18-month lead when hindcasts are initialized from July, suggesting that it is the best initial time to initialize multi-year forecasts; they are often initialized from November in other recent studies (e.g., Weisheimer et al., 2022). Following D19, the effect of the externally-forced trends on forecast skill is examined (see Text S5 in Supporting Information S1).

The pattern correlation between observed and forecasted SST anomalies in the tropical Pacific, indicates that some ENSO events can be predicted at leads of 18 months or even longer. These events also include both 1-year ENSO and 2-year El Niño events besides 2-year La Niña events (Wu et al., 2021). Our results indicate that El Niño events are also predicated at leads of 18 and even 24 months, longer than the 13-month predictability for El Niño events noted by Wu et al. (2021). For example, the 2009/10 El Niño event is reproduced by forecasts initialized in January 2008 at 24-month lead, and the pattern correlation for SSTA in the seasonal mean of December 2009 to February 2010 in the ENSO region reaches 0.88. Surprisingly, the 1999/2000 La Niña event was well predicted by the multi-model ensemble mean hindcasts initialized in July 1997 at a lead time of 30 months. The dynamics for these events warrant further examination and may help to better understand their extended predictability.

There is a close correspondence between ENSO activity and tropical Pacific SST forecast skill at long leads (i.e., 18 and 24 months), as noted for seasonal forecasts by D19. This suggests that high forecast skill tends to occur when the signal-to-noise ratio is relatively large, in particular during La Niña events (not shown). It is noted that the signal-to-noise ratio coincides largely with the ensemble mean hindcasts. This is because the inter-member forecast spread varies around a value, which approximately equals 1.0, rather than displaying large interannual variations (not shown), so that the skill is due to the ENSO signal, which increases with ENSO's amplitude, rather than the noise.

Scientific issues that are important to address for model-analog predictions include: the influence of a model's ability to simulate multi-year La Niña (Okumura & Deser, 2010) and El Niño (Wu et al., 2019) events, the statistical characteristics of higher-frequency phenomena such as westerly wind bursts (e.g., McPhaden, 1999) and interactions between basins on forecast skill (Alexander et al., 2022; Keenlyside et al., 2013). The model-analogs likely do not adequately include westerly wind bursts and high frequency forcing ("atmospheric noise") in general, which may have been especially important in El Niño-La Niña's evolution during 1997–1999.

Data Availability Statement

HadISST and ORA-S4 data are available at https://www.metoffice.gov.uk/hadobs/hadisst/ and https://www.cen.uni-hamburg.de/en/icdc/data/ocean/easy-init-ocean/ecmwf-ocean-reanalysis-system-4-oras4.html, respectively. The hindcasts and data from the four control runs are available at https://downloads.psl.noaa.gov/Projects/FAIR_paper_data/20230407_01/ and https://data1.gfdl.noaa.gov. The CESM-LE data are available at https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.cesmLE.html. The CMIP5 data output is available at https://esgf-node.llnl.gov/search/cmip5/.

Acknowledgments
This work has been funded by the NOAA Earth System Modeling program. We acknowledge the NCAR and CMIP5 climate modeling groups for producing and making available their model output.

References

Alexander, M. A., Blade, I., Newman, M., Lanzante, J., Lau, N., & Scott, J. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, *15*(16), 2205–2231. https://doi.org/10.1175/1520-0442(2002)015<2205:tabtio>2.0.co;2

Alexander, M. A., & Deser, C. (1995). A mechanism for the recurrence of wintertime midlatitude SST anomalies. *Journal of Physical Oceanography*, *25*(1), 122–137. https://doi.org/10.1175/1520-0485(1995)025<0122:amftro>2.0.co;2

Alexander, M. A., Shin, S.-I., & Battisti, D. S. (2022). The influence of the trend, basin interactions, and ocean dynamics on Tropical Ocean Prediction. *Geophysical Research Letters*, *49*(3), e2021GL096120. https://doi.org/10.1029/2021GL096120

Amaya, D. J., DeFlorio, M. J., Miller, A. J., & Xie, S.-P. (2017). WES feedback and the Atlantic meridional mode: Observations and CMIP5 comparisons. *Climate Dynamics*, *49*(9), 1665–1679. https://doi.org/10.1007/s00382-016-3411-1

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. https://doi.org/10.1002/qj.2063

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, *93*(5), 631–651. https://doi.org/10.1175/BAMS-D-11-00111.1

Becker, E. J., Kirtman, B. P., L'Heureux, M., Muñoz, Á. G., & Pegion, K. (2022). A decade of the North American Multimodel Ensemble (NMME): Research, application, and future directions. *Bulletin of the American Meteorological Society*, *103*(3), E973–E995. https://doi.org/10.1175/BAMS-D-20-0327.1

Capotondi, A., Wittenberg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, *96*(6), 921–938. https://doi.org/10.1175/BAMS-D-13-00117.1

Chang, P., Ji, L., & Li, H. (1997). A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions. *Nature*, *385*(6616), 516–518. https://doi.org/10.1038/385516a0

DiNezio, P. N., Deser, C., Karspeck, A., Yeager, S., Okumura, Y., Danabasoglu, G., et al. (2017). A 2 year forecast for a 60–80% chance of La Niña in 2017–2018. *Geophysical Research Letters*, *44*(22), 11–624. https://doi.org/10.1002/2017GL074904

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018). Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *Journal of Climate*, *31*(14), 5437–5459. https://doi.org/10.1175/JCLI-D-17-0661.1

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2019). Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophysical Research Letters*, *46*(3), 1721–1730. https://doi.org/10.1029/2018GL080598

Enfield, D. B., & Mayer, D. A. (1997). Tropical Atlantic sea surface temperature variability and its relation to El Niño-Southern Oscillation. *Journal of Geophysical Research*, *102*(C1), 929–945. https://doi.org/10.1029/96JC03296

Guan, Y., Zhu, J., Huang, B., Hu, Z.-Z., & Kinter, III, J. L. (2014). South Pacific Ocean dipole: A predictable mode on multiseasonal time scales. *Journal of Climate*, *27*(4), 1648–1658. https://doi.org/10.1175/JCLI-D-13-00293.1

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Keenlyside, N. S., Ding, H., & Latif, M. (2013). Potential of equatorial Atlantic variability to enhance El Niño prediction. *Geophysical Research Letters*, *40*(10), 2278–2283. https://doi.org/10.1002/grl.50362

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American multimodel ensemble: Phase-1 seasonal- to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kumar, A., Chen, M., Xue, Y., & Behringer, D. (2015). An analysis of the temporal evolution of ENSO prediction skill in the context of the equatorial Pacific Ocean observing system. *Monthly Weather Review*, *143*(8), 3204–3213. https://doi.org/10.1175/MWR-D-15-0035.1

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., et al. (1998). A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, *103*(C7), 14375–14393. https://doi.org/10.1029/97JC03413

Lou, J., Newman, M., & Hoell, A. (2023). Multi-decadal variation of ENSO forecast skill since the late 1800s. *npj Climate and Atmospheric Science*, *6*(1), 89. https://doi.org/10.1038/s41612-023-00417-z

Luo, J.-J., Masson, S., Behera, S. K., & Yamagata, T. (2008). Extended ENSO predictions using a fully coupled ocean-atmosphere model. *Journal of Climate*, *21*(1), 84–93. https://doi.org/10.1175/2007JCLI1412.1

McPhaden, M. J. (1999). Genesis and evolution of the 1997–98 El Niño. *Science*, *283*(5404), 950–954. https://doi.org/10.1126/science.283.5404.950

McPhaden, M. J. (2003). Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical Research Letters*, *30*(9), 1480. https://doi.org/10.1029/2003GL016872

Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., & Zebiak, S. E. (1998). ENSO theory. *Journal of Geophysical Research*, *103*(C7), 14261–14290. https://doi.org/10.1029/97JC03424

Newman, M., Compo, G. P., & Alexander, M. A. (2003). ENSO-forced variability of the Pacific decadal oscillation. *Journal of Climate*, *16*(23), 3853–3857. https://doi.org/10.1175/1520-0442(2003)016⟨3853:EVOTPD⟩2.0.CO;2

Newman, M., & Sardeshmukh, P. D. (2017). Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*, *44*(16), 8520–8529. https://doi.org/10.1002/2017GL074088

Okumura, Y. M., & Deser, C. (2010). Asymmetry in the duration of El Niño and La Niña. *Journal of Climate*, *23*(21), 5826–5843. https://doi.org/10.1175/2010JCLI3592.1

Rayner, N., Parker, D. E., Horton, E., Folland, C., Alexander, L., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Saji, N., Goswami, B. N., Vinayachandran, P., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, *401*(6751), 360–363. https://doi.org/10.1038/43854

Sharmila, S., Hendon, H., Alves, O., Weisheimer, A., & Balmaseda, M. (2022). Contrasting El Niño-La Niña predictability and prediction skill in 2-year reforecasts of the 20th century. *Journal of Climate*, *36*(5), 1269–1285. https://doi.org/10.1175/JCLI-D-22-0028.1

Solomon, A., Goddard, L., Kumar, A., Carton, J., Deser, C., Fukumori, I., et al. (2011). Distinguishing the roles of natural and anthropogenically forced decadal climate variability: Implications for prediction. *Bulletin of the American Meteorological Society*, *92*(2), 141–156. https://doi.org/10.1175/2010BAMS2962.1

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., & Ro- pelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research*, *103*(C7), 14291–14324. https://doi.org/10.1029/97JC01444

Weisheimer, A., Balmaseda, M. A., Stockdale, T. N., Mayer, M., Sharmila, S., Hen- don, H., & Alves, O. (2022). Variability of ENSO forecast skill in 2-year global reforecasts over the 20th century. *Geophysical Research Letters*, *49*(10), e2022GL097885. https://doi.org/10.1029/2022GL097885

Wu, X., Okumura, Y. M., Deser, C., & DiNezio, P. N. (2021). Two-year dynamical predictions of ENSO event duration during 1954–2015. *Journal of Climate*, *34*(10), 4069–4087. https://doi.org/10.1175/JCLI-D-20-0619.1

Wu, X., Okumura, Y. M., & DiNezio, P. N. (2019). What controls the duration of El Niño and La Niña events? *Journal of Climate*, *32*(18), 5941–5965. https://doi.org/10.1175/JCLI-D-18-0681.1

Xie, S.-P., Annamalai, H., Schott, F. A., & McCreary, J. P. (2002). Structure and mechanisms of South Indian Ocean climate variability. *Journal of Climate*, *15*(8), 864–878. https://doi.org/10.1175/1520-0442(2002)015⟨0864:SAMOSI⟩2.0.CO;2

Yeager, S. (2020). The abyssal origins of North Atlantic decadal predictability. *Climate Dynamics*, *55*(7–8), 2253–2271. https://doi.org/10.1007/s00382-020-05382-4

Yeager, S. G., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., et al. (2018). Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*, *99*(9), 1867–1886. https://doi.org/10.1175/BAMS-D-17-0098.1

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., et al. (2022). The Seasonal-to-Multiyear Large ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geoscientific Model Development*, *15*(16), 6451–6493. https://doi.org/10.5194/gmd-15-6451-2022

Zebiak, S. E., & Cane, M. A. (1987). A model El Niño–Southern Oscillation. *Monthly Weather Review*, *115*(10), 2262–2278. https://doi.org/10.1175/1520-0493(1987)115<2262:ameno>2.0.co;2

References From the Supporting Information

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., et al. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climate Dynamics*, *109*(1–2), 33–57. https://doi.org/10.1007/s10584-011-0149-y

Shin, S.-I., & Newman, M. (2021). Seasonal predictability of global and North American coastal sea surface temperature and height anomalies. *Geophysical Research Letters*, *48*(10), e2020GL091886. https://doi.org/10.1029/2020GL091886