Strong El Niño events lead to robust multi-year ENSO predictability

N. Lenssen[1,2,3], P. DiNezio[1], L. Goddard[2], C. Deser[4], Y. Kushnir[5], S. Mason[2], M. Newman[6], Y. Okumura[7]

[1]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA
[2]International Research Institute for Climate and Society, Columbia University, Palisades, NY, USA
[3]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA
[4]National Center for Atmospheric Research, Boulder, CO, USA
[5]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA
[6]NOAA Physical Sciences Laboratory, Boulder, CO, USA
[7]Jackson School of Geosciences, University of Texas, Austin, TX, USA

Key Points:

- ENSO is predictable for 2+ years following strong El Niño events.
- Forecasts initialized during Weak El Niño, Neutral, and La Niña states are not skillfull at leads greater than 12 months.
- There is a potential long-lead forecast of opportunity out of the expected strong 2023-2024 El Niño event.

Corresponding author: Nathan Lenssen, `lenssen@mines.edu`

Abstract

The El Niño-Southern Oscillation (ENSO) phenomenon – the dominant source of climate variability on seasonal to multi-year timescales – is predictable a few seasons in advance. Forecast skill at longer multi-year timescales has been found in a few models and forecast systems, but the robustness of this predictability across models has not been firmly established owing to the cost of running dynamical model predictions at longer lead times. In this study, we use a massive collection of multi-model hindcasts performed using model analogs to show that multi-year ENSO predictability is robust across models and arises predominantly due to skillful prediction of multi-year La Nina events following strong El Niño events.

Plain Language Summary

In this study, we demonstrate that ENSO is predictable at least two years in advance when forecasts are made during strong El Niño events, such as the current El Niño expected to peak in winter 2023-2024. That is, strong El Niños provide forecasts of opportunity in which we have high confidence in multi-year predictions of ENSO. The opposite is also shown; forecasts initialized during other ENSO states (weak El Niño, Neutral, and La Niña) do not have predictive skill past 12 months. These result hold regardless of the climate model used to make the predictions a shown using 1,000s of years of retrospective climate forecasts made with 11 different state-of-the-art climate models.

1 Introduction

There is immense societal benefit from skillful multi-year climate forecasts as many human systems make decisions on this timescale (Nissan et al., 2019). The El Niño/Southern Oscillation (ENSO) – the dominant mode of climate variability at multi-year time scales – influences global weather via atmospheric teleconnections (Lenssen et al., 2020; Mason & Goddard, 2001; Ropelewski & Halpert, 1986), and has well-known predictability at lead times of 9 or less months (Barnston et al., 2019; Tippett et al., 2019; L'Heureux et al., 2020; Becker et al., 2022). Numerous forecast systems have shown small, but significant predictive skill at lead times beyond 9 months with dynamical ((Gonzalez & Goddard, 2016; Dunstone et al., 2020) and statistical (Ding & Alexander, 2023; Ham et al., 2019; Wang et al., 2023) methods, but the sources of this skill are not firmly established.

The long-lead predictability of ENSO could arise from particular sequences of ENSO events. For instance, persistent La Niña states lasting 2 or more years appear highly predictable, particularly after a strong El Niño event (DiNezio, Deser, Okumura, & Karspeck, 2017; DiNezio, Deser, Karspeck, et al., 2017; Wu et al., 2019; Wu, Okumura, Deser, & DiNezio, 2021). Conversely, El Nino states lasting multiple years might be predictable based on the onset season (Wu et al., 2019; Wu, Okumura, & DiNezio, 2021; Wu, Okumura, Deser, & DiNezio, 2021). These studies provided major advances connecting dynamical theories of ENSO to determine potential predictable multi-year sequences. However, these studies used hindcasts performed with a single coupled general circulation model (CGCM) and contain a limited number of events for retrospective validation. Evidence for multi-year predictability from other CGCMs is sparse and not systematically explored (Dunstone et al., 2020; Lou et al., 2023). Therefore, a robust assessment of skill across a multi-model ensemble is needed.

Small hindcast sample sizes are a ubiquitous limitation in ENSO-prediction research. Hindcast experiments are run over tens of years of initializations, containing only a dozen or so ENSO events. Furthermore, seasonal hindcast experiments have not historically included predictions past 12 month leads. These hindcast experiments are limited by computational costs of initialized CGCMs and/or short observational data records needed for initialization and verification (Barnston et al., 2019; Tippett et al., 2019). For instance, the NMME has hindcasts initialized monthly over 1982-2010 and real-time forecasts initialized beginning in 2011 with lead times up to 11 months (408 forecasts for each CGCM verified in (Barnston et al., 2019)) and the CMIP6 Decadal Climate Prediction Project (DCPP) has hindcasts initialized yearly over 1960-2018 with lead time up to 10 years (59 forecasts for each CGCM verified in (Dunstone et al., 2020)). When evaluating such datasets, it is necessary to evaluate the skill of a forecast system over all hindcasts to maximize sample size in the statistical estimates of forecast skill. However, pooling all forecasts, particularly by ENSO state at initialization, has the potential to obfuscate the underlying sources of long-lead ENSO skill if predictability is state-dependent.

In this study, we investigate the model and initial state dependence of multi-year ENSO prediction skill. We explore initial ENSO states in terms of phase (El Niño, neutral, La Niña) and intensity (strong, weak) providing multi-year skill. To this aim, we construct and analyze a massive multi-model ensemble of model analog climate hindcasts to identify initial states that lead to multi-year predictive skill. The model analog method (Ding et al., 2018, 2019, 2020) is used to make forecasts by first identifying states in a "library" of CGCM output that best match the initial state. Then, ensemble forecasts are issued according to how each of these states evolved in the CGCM. These forecasts are appropriate to use to investigate ENSO predictability as they have tropical Pacific skill equal to or exceeding state-of-the-art initialized dynamical forecast systems (Ding et al., 2018). In addition, the very low computational cost allows the generation of very large ensemble hindcasts based on multiple CMIP-class CGCMs with leads of 3+ years. Together, these features of our technique enabled us to investigate the model and state dependence of 2 year ENSO prediction skill.

Section 2 outlines the data and methods used int this study. In Section 3, we investigate the state-dependence of year 2 ENSO skill in perfect model hindcasts, which provide an upper bound for predictability. Then in Section 4, we investigate the state-dependence in cross-model hindcasts; we use many CGCMs as library states to predict a long control run of a single model with model analog forecasts. Finally in Section 5, we turn to the real world and use model analog forecasts to predict ENSO over the 109 year record from 1901-2009. In each of these analyses, we show that ENSO skill is highly dependent on the state at initialization as well as the target state. Nearly all of the skill at leads greater than 12 months is due to prediction out of El Niño, consistent with known multi-year patterns of ENSO such as the tendency for La Niña to follow El Niño. This state-dependency is shown through the skill of probabilistic forecasts of DJF ENSO state at leads up to 36 months.

2 Data and Methods

2.1 Data

We use long pre-industrial control simulations of at least 500 years in duration from 11 state-of-the-art CGCMs to issue model-analog forecasts and to perform the verifications in Sections 3 and 4. The 11 CGCMs are seven CMIP-class CGCMs and the four available control runs from NMME CGCMs (Table S1). All gridded products are regridded to a common $2° \times 2°$ grid before use in any analyses. The monthly mean sea-surface temperature (SST) or "tos" fields and sea-surface height (SSH) or "zos" fields are used. SST and SSH anomalies are created by removing the monthly climatologies.

The CERA-20C reanalysis is used as SST and SSH observations used to conduct observational hindcast experiment in Section 4, following (Lou et al., 2023). A reanalysis product is used to extend the record to span 1901-2009 as complete Indo-Pacific observations of SSH do not exist prior to the satellite era. As with the model output, observed SST and SSH fields are first regridded to the common 2x2 grid and then converted to anomalies prior to analysis by removing the monthly climatologies.

ENSO events are defined according to quantiles of the Oceanic Niño Index (ONI) which is the seasonal (3 month) average SST anomaly over the Niño 3.4 region (5N-5S, 170W-120W). These quantiles are calculated for each season for each CGCM as well as the observations. El Niño events are defined as the upper quartile, or values above the 75th percentile, of ONI. Similarly, La Niña events are defined as the lower quartile, or values below the 25th percentile of ONI. This method is useful when comparing ENSO-state prediction across different CGCMs as it reduces the bias from different CGCM ENSO mean states and variabilities (Gonzalez & Goddard, 2016).

2.2 Model Analog Forecasts

In general, we make model analog forecasts in a two step process. (1) We find the best analogs for the initial state by searching through a library of CGCM output. (2) We issue forecasts according to how the best analogs found in (1) evolved. We follow the full method as documented in Ding et al. (2018). In perfect model analog hindcasts (Section 2), we exclude the initial state from the library of possible analogs. In cross-model and observational hindcasts (Sections 3 and 4), we use each entire CGCM piControl run as the library for best analog states.

Best analogs are found by finding the best matches of SST and SSH fields between the initial state and all states within the same month in the CGCM library. The initial and library fields are compared over the entire tropical Indo-Pacific basin (30S–30N, 30E–80W). For each time step in the library, we calculate the root mean square (RMS) distance from the initial SST and SSH fields to the corresponding library fields. Here, all fields are normalized to have unit variance to allow adding the distances between the initial and library SST and SSH fields as well as accounting for biases in variability between datasets. These distances are ranked in ascending order and the evolution of the 15 states closest to the initial field are used to create an ensemble forecast.

For a given initial state, the ensemble forecast plume is determined by the evolution of the Niño3.4 index in the closest 15 analogs. We issue probabilistic forecasts of ENSO events at each lead as the proportion of these 15 analogs that predict El Niño, Neutral, and La Niña conditions where we define ENSO events using the quantile method described above. We choose the closest 15 analogs for our forecast as this number provides high forecast skill for a wide range of library sizes (Ding et al., 2018).

2.3 Forecast Verification

The probabilistic skill of the ENSO state forecasts is determined using RPSS, a standard skill metric for probabilistic skill (Jolliffe & Stephenson, 2012; Mason, 2018). RPSS is a measure of both a forecast's resolution, or whether the outcome differs given different forecasts, as well as a forecast's reliability, or how well the forecasted probabilities match the observed rate of events (Mason, 2018). The RPSS is a skill score comparing the Ranked Probability Score (RPS; (Epstein, 1969; Murphy, 1971)) of the forecast of interest to a climatological forecast. It is defined in such a way that an RPSS of 1.0 indicates a perfect forecast, an RPSS indicates that a forecast is equivalent to climatology, and a negative forecast indicates a forecast that is less skillful than the climatological rate of ENSO events.

3 Perfect Model Hindcast Experiment

We first investigate the perfect model skill, or the skill of a model predicting its own dynamics. That is, we use the same CGCM as both the target states as well as the library, omitting the state we are trying to predict as a possible analog. Perfect model skill is generally an upper bound of skill for the ENSO system. When predicting the state of ENSO in December-February (DJF), the peak season of ENSO, all models have positive ranked probability skill scores (RPSS) at leads of up to 12 months (Figure 1a). RPSS is a measure of probabilistic forecast skill where a value of zero indicates a forecast is on par with a forecast of climatological probabilities and positive values indicate that the model analog forecasts outperform climatological forecasts. All but two of the eleven models in the study have positive RPSS out to at least 24 months, indicating that a range of CGCMs with varied ENSO dynamics exhibit "perfect model" multi-year ENSO predictability (Figure 1a). These findings agree with theoretical calculations of ENSO predictability of around 3 years (Newman & Sardeshmukh, 2017), the skill of initialized dynamical forecasts (DiNezio, Deser, Okumura, & Karspeck, 2017; Dunstone et al., 2020; Wittenberg et al., 2014), and multi-model long lead skill of model analog forecasts (Lou et al., 2023).

A major goal of this study is to determine if specific states are causing the majority of skill in forecasts at leads greater than 12 months. As discussed, this type of information can not be determined by verification metrics performed over all initialization and target states as has been traditionally done with limited hindcast experiments. Here, we determine the probabilistic skill of DJF-target ENSO forecasts stratified by the state at initialization. The initial state bins are: strong El Niño (greater than 95th percentile of Niño3.4), weak El Niño (between 75% and 95% percentile of Niño3.4), and no El Niño which includes both neutral and La Niña states (below 75% percentile of Niño3.4). Expanding the set of initial state bins to include weak La Niña and strong La Niña does not alter year-2 skill in any CGCM (not shown).

For perfect-model forecasts from CESM1.1, by far the greatest year-2 skill comes from forecasts initialized out of strong El Niño events as seen by the large difference between the strong El Niño line and the no El Niño RPSS skill at leads greater than 12 months (Figure 1b). The strong El Niño skill between leads 12-24 months is expected due to the strong tendency for La Niña to occur after strong El Niño events. The strong El Niño skill seen at leads 24-36 months is due to the high predictability of two year La Niña events following large El Niño events that has been previously shown in CESM1 (DiNezio, Deser, Okumura, & Karspeck, 2017). The same dramatic increase in year 2+ skill does not occur from forecasts initialized during weak El Niño events, as shown between the negligible difference between the weak El Niño and no El Niño skill (Figure 1b). This analysis was performed with all CGCMs in the study, leading to qualitatively similar results (not shown).

Forecasts initialized during strong El Nino events (Niño 3.4 $>$ 95th percentile) have the greatest year-2 skill across all CGCMs used to generate hindcast experiments except CanESM5 (Figure 1c). This additional year-2 skill from strong El Niño initial states is seen in the difference between the strong El Niño RPSS and the no El Niño RPSS (Figure 1c) as this accounts for any differences in the total skill of the CGCMs. As with CESM1.1, CGCMs generally do not see much additional skill from weak El Niño initial states when compared with no El Niño (Figure 1d).

We have robustly shown that ENSO is most predictable at leads of 12+ months for perfect model analog forecasts when initialized during a strong El Niño event. This result agrees with theory that there is a strong dynamical tendency for La Niña to follow El Niño events (DiNezio, Deser, Okumura, & Karspeck, 2017; Suarez & Schopf, 1988). In addition, active ENSO states are more reliably predictable than ENSO-neutral states leading to greater probabilistic skill (Jin et al., 2008; Mason et al., 2021).

With this greater predictability out of strong El Niño, it is natural to ask if the year 2+ skill is indeed due to greater predictability of subsequent La Niña events of one- or two-year duration. To test this, we take each of the initial states used in Figure 1b and decompose the forecast skill according to the true ENSO state upon verification. Results with two of the CGCMs with greatest multiyear skill, GISS-E2.1G and CESM1.1, are shown as illustrative examples (Figure 2), but similar results are found for all 11 CGCMs in the study (not shown).

All of the skill in forecasts initialized during strong El Niño events is due to very skillful forecasts of La Niña events (Figures 2a,d). This result, which holds for 11 CGCMs, provides robust support for the theory that strong El Niño events precede highly predictable single and double La Niña events (DiNezio, Deser, Okumura, & Karspeck, 2017). In addition, there is evidence of weak El Niño events leading to predictable double El Niño events (Wu, Okumura, & DiNezio, 2021) as seen by the positive El Niño skill in leads 12-18 for El Niño targets (Figures 2b,e). Finally, there is some evidence for predicting El Niño multiple years in advance from neutral states (Figures 2c,f).

Decomposing skill calculations by the state at verification is very useful to understand what states a forecast system predicts well, but is artificial as it is impossible to know the target state a priori when making real time forecasts. Thus, the analysis presented in Figure 2 can only be used to show retrospectively that certain verification states lead to greater skill, and the results in Figure 1 should be used to understand what the perfect model, or upper bound, of ENSO skill is using model analog forecasts.

4 Cross-Model Hindcast Experiment

Perfect-model prediction studies are useful to determine possible upper bounds of ENSO predictability, but do not necessarily reflect real-world predictability, especially if a CGCM does not simulate ENSO dynamics realistically. To confirm the perfect-model findings presented in Section 2, we perform two "cross-model" hindcast experiments in which we use each model to predict the full preindustrial control (piControl) runs of GISS-E2.1G and CESM1.1. Cross-model hindcasts investigate the forecast skill of model-analog forecasts in predicting a target ENSO system that is different from the library ENSO system, analogous to the case of using model-analog forecasts to predict the real-world ENSO system. By using this cross-model hindcast setup, we are able to generate thousands of years of hindcasts in a setting that better represents operational forecasts than perfect model hindcasts.

We use each of the 10 other CGCMs to issue model-analog forecasts of the 851-year GISS-E2.1G piControl as it has the greatest perfect model skill, but a highly oscillatory ENSO (Figure S1). We additionally perform hindcasts over the 1,800 year CESM1.1 piControl as it has a morerealistic ENSO, particularly in terms of the asymmetric evolution of El Niño and La Niña events (Figure S1; (Capotondi et al., 2020; DiNezio, Deser, Okumura, & Karspeck, 2017)).

The cross-model skill is generally lower than the perfect model skill, but there is still positive RPSS skill at leads of 24 months for most CGCMs in both cross-model experiments (Figures 3a,d). When predicting GISS-E2.1G, many of the CGCMs are nearly as skillful as their perfect model benchmark (Figure 3a). This is expected as GISS-E2.1G has a relatively oscillatory ENSO, leading to a more predictable system (Figure S1). When predicting the more complex and realistic ENSO in CESM1.1, the cross-model skill is lower because of this more complex and less active ENSO (Figure 3d). Note that both of these CGCMs simulate two-year La Niña events near the observed rate of around 6.8/100 years, with 7.5/100 years in GISS-E2.1G and 6.7/100 years in CESM1.1 (Table S1).

As with the perfect model hindcasts, we decompose the cross-model RPSS by state at initialization. Again, we see that most of the year-2 skill comes from predictions out of strong El Niño events (Figures 3b,d). When using GISS-E2.1G as the hindcast target, all but three CGCMs show better 12-18 month skill and all CGCMs show better 18-24 month out of strong El Niño events than other initial states (Figure 3b). When predicting the more realistic CESM1.1 ENSO, all CGCMs have much more skill when initialized during strong El Niño events when compared with no El Niño events (Figure 3e).

Following the perfect model results, initialization during weak El Niño events does not dramatically increase year-2 skill (Figures 3c,f).

5 Observational Hindcast Experiment

To demonstrate that the above results hold for the real-world ENSO system, we create model-analog hindcasts using a library from each CGCM to predict a 109-year record of the real-world ENSO system from 1901-2009 (Laloyaux et al., 2018). These observational hindcasts show that model-analog forecasts have skill at leads exceeding 12 months with some CGCM analogs, in agreement with previous studies (Figure 4a; (Liu et al., 2022; Lou et al., 2023)). In addition, the observational hindcasts show comparable, albeit slightly lower, skill in predicting the observations to their skill in predicting the full piControl of CESM1.1 (Figure 3d). This lower skill for the observations is because CGCMs generally overestimate the ENSO signal-to-noise ratio leading to overconfident forecasts of the real world system (Eade et al., 2014; Tippett et al., 2020).

We expect substantial sampling uncertainty in quantifying skill over the 109-year hindcast due to the limited sample size in the verification statistics as well as the known multidecadal variability in ENSO predictability (Wittenberg, 2009; Wittenberg et al., 2014; Lou et al., 2023). To make fair comparisons between the observational hindcasts here and the cross-model hindcasts in Section 3, we quantify this sampling uncertainty in the observational hindcast. We use a bootstrapping approach in which we create and verify 200 hindcasts using analogs from each CGCM over random 109-year periods of the 1,800 year CESM1.1 piControl. The 95% likely skill from the subsampled 109 year cross-model CESM1.1 hindcasts and the range of the observational hindcast skill overlap for all leads but 4 months (Figure 4b). Thus, we cannot reject the hypothesis that DJF skill is lower when predicting the observed ENSO system than when predicting the CESM1.1 ENSO system.

This subsampling analysis is additionally used to estimate the 95% confidence intervals of skill when stratifying by initial state on the 109-year observational record (Figure 4c). We again take random 109-year periods of the CESM1.1 piControl and determine the 95% likely range of forecast skill given the ENSO state at initialization. As expected, there is large uncertainty when verifying such few forecasts (violin plots in Figure 4c), but the majority of year-2 skill comes from predictions initialized during strong El Niño events . The strong El Niño-initialized observational hindcasts (box plots in Figure 4c) show comparable skill to the cross-model case at 12-18 month leads, but lower skill at 18-24 month leads. However, the middle 50% of CGCMs show positive RPSS at leads of 18-24 months when initialized during strong El Niño, again suggesting that there is a multi-year forecast of opportunity during strong El Niño events. On the other hand, there is no significant skill beyond 12 months in the observational hindcasts when the initial state is not a strong El Niño event (Figure 4c).

6 Summary and Discussion

There is skill in predicting ENSO at leads of 12-24 months, but it is nearly entirely due to the high long-lead predictability of the system following strong El Niño events. This finding is robust in long multi-model perfect model hindcasts, long multi-model cross-model hindcasts, and predictions over a 109-year observational reanalysis.

These findings are important for both climate predictability research and for climate service applications using seasonal to multi-year predictions. Research into ENSO and climate predictability generally focuses on metrics of skill aggregated over all forecasts, a required assumption given the small hindcasts available. As such, multiple studies have claimed that ENSO can be predicted skillfully into the second year (Dunstone et al., 2020; Gonzalez & Goddard, 2016; Ham et al., 2019; Wang et al., 2023). Our findings make clear that this second-year skill is not always present in the system; second-year skill is highly state dependent with robust multi-year skill only possible out of large El Niño events.

Our results present both good and bad news for climate services or decision makers relying on climate information. A strong El Niño event presents a multi-year forecast of opportunity for ENSO. Since ENSO is the dominant driver of climate variability on multi-year timescales, we expect that multi-year predictions of climate impacts will have the greatest multi-year skill out of strong El Niño events. Such forecasts of opportunity should be investigated further. On the other hand, there is little evidence shown here for multi-year ENSO skill when initializing in a state other than a strong El Niño. Thus, climate service and humanitarian actions will likely need to rely on information other than climate forecasts when making decisions at leads past 12 months if a strong El Niño event is not ongoing.

This study has implications for future predictability of ENSO under climate change. If climate change leads to an increased chance of extreme El Niño events (Cai et al., 2020) and subsequent multi-year La Niña events (Geng et al., 2023), our findings suggest that ENSO will become more predictable at longer leads on average, in agreement with studies using model analog forecasts on future ENSO predictability (Amaya et al., 2024).

The ability to generate multi-model hindcasts over thousands of years on a laptop using model analog forecasts is an incredibly powerful tool. Large sample sizes provide the ability to decompose forecast skill by both initial and target state to determine what ENSO states led to multi-year skill. In addition, large samples make it possible to quantify the sampling uncertainty on forecasts of the observational record to determine the robustness of skill analyses over a shorter record. Model analog forecasts combined with the wealth of output from CMIP provide a tool for robustly exploring questions about climate variability, predictability, and change.

Our conclusions are particularly salient given the incipient strong El Niño expected to peak during the 2023-2024 boreal winter. Following our findings, ENSO forecasts issued this coming winter will provide actionable information about the state of ENSO through 2025.

7 Open Research

The live code-base used to process the data, run the experiments, and verify forecasts can be found at `https://github.com/nlenssen/LongLeadENSO/`. An archived code-base is available on Zenodo at `https://doi.org/10.5281/zenodo.10045616`. All raw, intermediate, and final data is archived at Zenodo at `https://doi.org/10.5281/zenodo.10045687`.

Acknowledgments

The authors thank Weston Anderson, Ángel Muñoz, Kevin Schwarzwald, and Xian Wu for their feedback and suggestions. Funding for NL, PDN, and YO through the NSF Climate & Large-Scale Dynamics Award #1756883. NL, LG, and SM were funded in part by ACToday, a Columbia World Project funded by Columbia University.

References

Amaya, D. J., Maher, N., Deser, C., Jacox, M. G., Newman, M., Alexander, M. A., ... Lou, J. (2024). Future changes in seasonal climate predictability. *Science Advances (Under Review)*.

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019).

Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, *53*(12), 7215–7234.

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.

Cai, W., Santoso, A., Wang, G., Wu, L., Collins, M., Lengaigne, M., ... Timmermann, A. (2020). ENSO Response to Greenhouse Forcing. *El Niño Southern Oscillation in a Changing Climate*, 289–307.

Capotondi, A., Deser, C., Phillips, A., Okumura, Y., & Larson, S. (2020). ENSO and Pacific decadal variability in the Community Earth System Model version 2. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2019MS002022.

DiNezio, P. N., Deser, C., Karspeck, A., Yeager, S., Okumura, Y., Danabasoglu, G., ... Meehl, G. A. (2017). A 2 year forecast for a 60–80% chance of la niña in 2017–2018. *Geophysical Research Letters*, *44*(22), 11–624.

DiNezio, P. N., Deser, C., Okumura, Y., & Karspeck, A. (2017). Predictability of 2-year La Niña events in a coupled general circulation model. *Climate dynamics*, *49*(11), 4237–4261.

Ding, H., & Alexander, M. A. (2023). Multi-year predictability of global sea surface temperature using model-analogs. *Geophsyical Research Letters (Accepted)*.

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.

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.

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2020). Relating CMIP5 model biases to seasonal forecast skill in the tropical Pacific. *Geophysical Research Letters*, *47*(5), e2019GL086765.

Dunstone, N., Smith, D., Yeager, S., Danabasoglu, G., et al. (2020). Skilful interannual climate prediction from two large initialised model ensembles. *Environmental Research Letters*, *15*(9), 094083.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*(6), 985-987.

Geng, T., Jia, F., Cai, W., Wu, L., Gan, B., Jing, Z., ... McPhaden, M. J. (2023). Increased occurrences of consecutive La Niña events under global warming. *Nature*, *619*(7971), 774–781.

Gonzalez, P. L., & Goddard, L. (2016). Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Climate Dynamics*, *46*(9), 3127–3147.

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572.

Jin, E. K., Kinter, J. L., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B., ... others (2008). Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, *31*, 647–664.

Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: A practitioner's guide in atmospheric science* (2nd ed.). West Sussex, England: Wiley.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., ... others (2018). CERA-20C: a coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172–1195.

Lenssen, N., Goddard, L., & Mason, S. (2020). Seasonal Forecast Skill of ENSO Teleconnection Maps. *Weather and Forecasting*, *35*(6), 2387–2406.
L'Heureux, M. L., Levine, A. F., Newman, M., Ganter, C., Luo, J.-J., Tippett, M. K., & Stockdale, T. N. (2020). ENSO prediction. *AGU Monograph: El Niño Southern Oscillation in a changing climate*, 227–246.
Liu, T., Song, X., Tang, Y., Shen, Z., & Tan, X. (2022). ENSO predictability over the past 137 years based on a CESM ensemble prediction system. *Journal of Climate*, *35*(2), 763–777.
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.
Mason, S. J. (2018). Guidance on verification of operational seasonal climate forecasts. *World Meteorological Organization, Commission for Climatology XIV Technical Report*.
Mason, S. J., Ferro, C. A., & Landman, W. A. (2021). Forecasts of "normal". *Quarterly Journal of the Royal Meteorological Society*, *147*(735), 1225–1236.
Mason, S. J., & Goddard, L. (2001). Probabilistic precipitation anomalies associated with ENSO. *Bulletin of the American Meteorological Society*, *82*(4), 619-638.
Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied Meteorology*, *10*(1), 155-156.
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.
Nissan, H., Goddard, L., de Perez, E. C., et al. (2019). On the use and misuse of climate change projections in international development. *Wiley Interdisciplinary Reviews: Climate Change*, *10*(3), e579.
Ropelewski, C. F., & Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review*, *114*(12), 2352-2362.
Suarez, M. J., & Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal of Atmospheric Sciences*, *45*(21), 3283–3287.
Tippett, M. K., L'Heureux, M. L., Becker, E. J., & Kumar, A. (2020). Excessive momentum and false alarms in late-spring ENSO forecasts. *Geophysical research letters*, *47*(8), e2020GL087008.
Tippett, M. K., Ranganathan, M., L'Heureux, M., Barnston, A. G., & DelSole, T. (2019). Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dynamics*, *53*, 7497–7518.
Wang, H., Hu, S., & Li, X. (2023). An interpretable deep learning ENSO forecasting model. *Ocean-Land-Atmosphere Research*, *2*, 0012.
Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*(12).
Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A., & Zeng, F. (2014). ENSO modulation: Is it decadally predictable? *Journal of Climate*, *27*(7), 2667–2681.
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.
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.
Wu, X., Okumura, Y. M., & DiNezio, P. N. (2021). Predictability of El Niño duration based on the onset timing. *Journal of Climate*, *34*(4), 1351–1366.

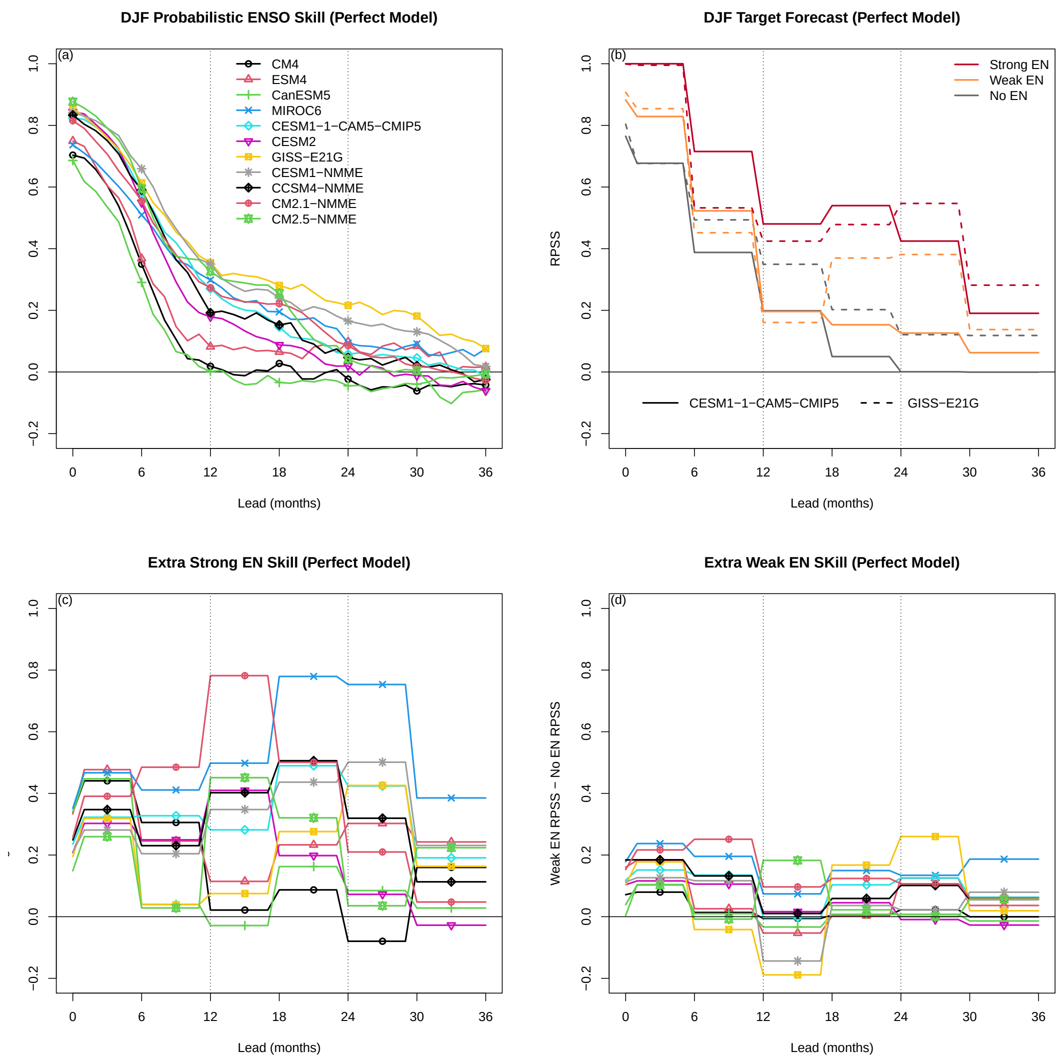


Figure 1. The model analog DJF RPSS skill for (a) perfect model hindcasts of all 11 CGCMs used in the study and (b) perfect model hindcasts stratified by ENSO state at initialization for two example CGCMs, CESM1.1 and GISS-E2.1G. The extra skill added when initializing during El Niño conditions is shown by the difference in RPSS between (c) strong EN initial states and no EN initial states and (d) weak EN initial states and no EN initial states.

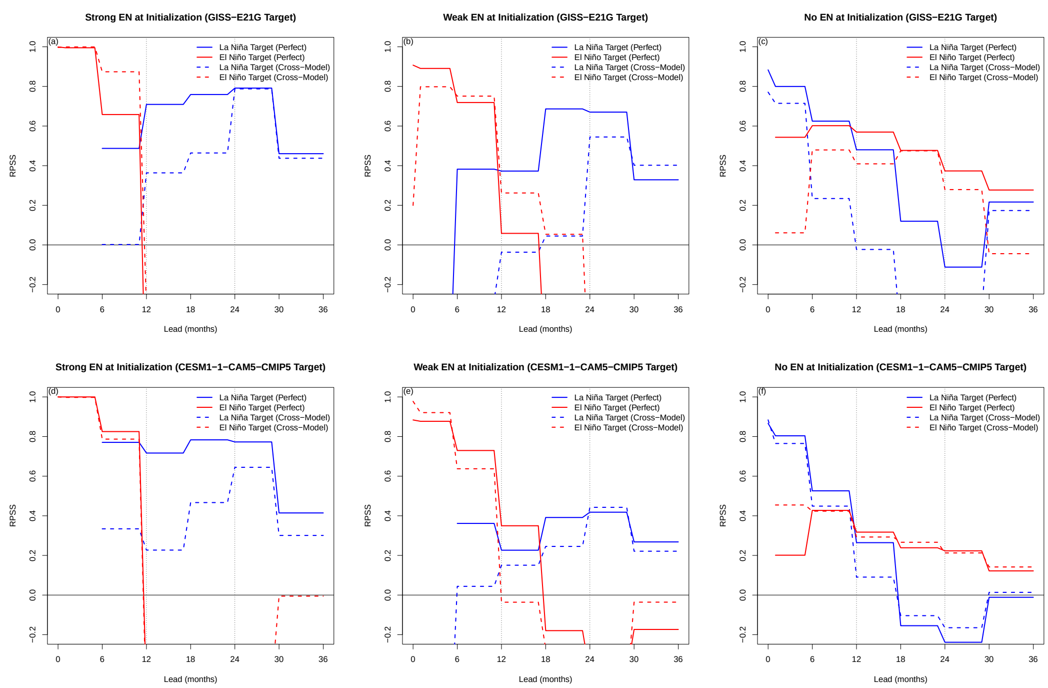


Figure 2. A second decomposition of the skill analysis in Figure 1b in which the skill is stratified by initial state in CESM1.1 and GISS-E2.1G where (a,d) show the skill of forecasts initialized during strong EN, (b,e) during weak EN, and (c,f) during no EN. The top row shows forecasts predicting the piControl of GISS-E2.1G and the bottom row shows forecasts predicting the piControl of CESM1.1. In all plots, solid lines indicate perfect model skill, and dashed lines indicate cross-model skill. That is, a dotted line of the top row indicates CESM1.1 predicting GISS-E2.1 piControl.

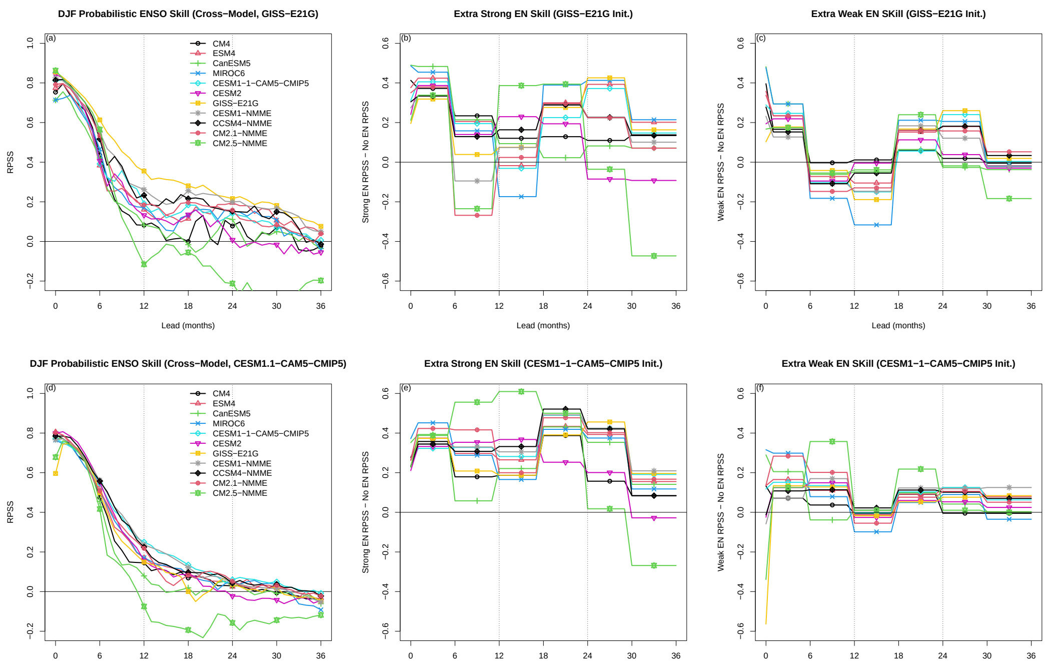


Figure 3. The model analog DJF RPSS skill of cross-model hindcasts using libraries from all 11 CGCMs to predict the piControl of (a) GISS-E2.1G and (d) CESM1.1. The remaining panels follow the analysis presented in Figures 1c,d by summarizing the extra skill in (b,e) forecasts initialized during strong EN relative to no EN and (c,f) forecasts initialized during weak EN relative to no EN.

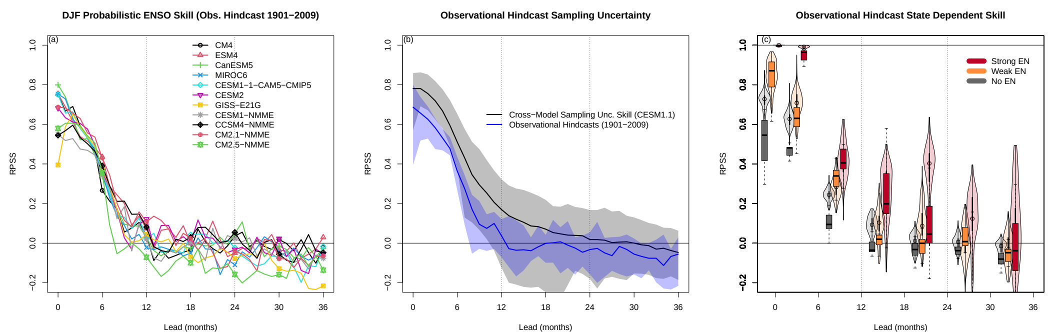


Figure 4. The model analog DJF RPSS skill of forecasts using libraries from all 11 CGCMs to predict (a) the observational record from 1901-2009 (109 years). The grey in (b) shows the 95% confidence interval due to sampling uncertainty estimated as the empirical median and 95% confidence interval of 200 simulations of all CGCMs making 109 year cross-model hindcasts of the CESM1.1 piControl. The sampling uncertainty is compared with blue curve showing the range over all 11 CGCMs of observational skill. Note that the blue range in (b) is exactly the range of the skill shown in (a). The final panel (c) is an expanded version of Figure 1b and shows the RPSS skill given the state at initialization. The violin plots with transparent colors show the sampling distribution from the resampled 109 year cross-model hindcasts of CESM1.1. The box plots with solid colors show the spread of skill for the 11 CGCMs in predicting the observational record.

Figure 1.

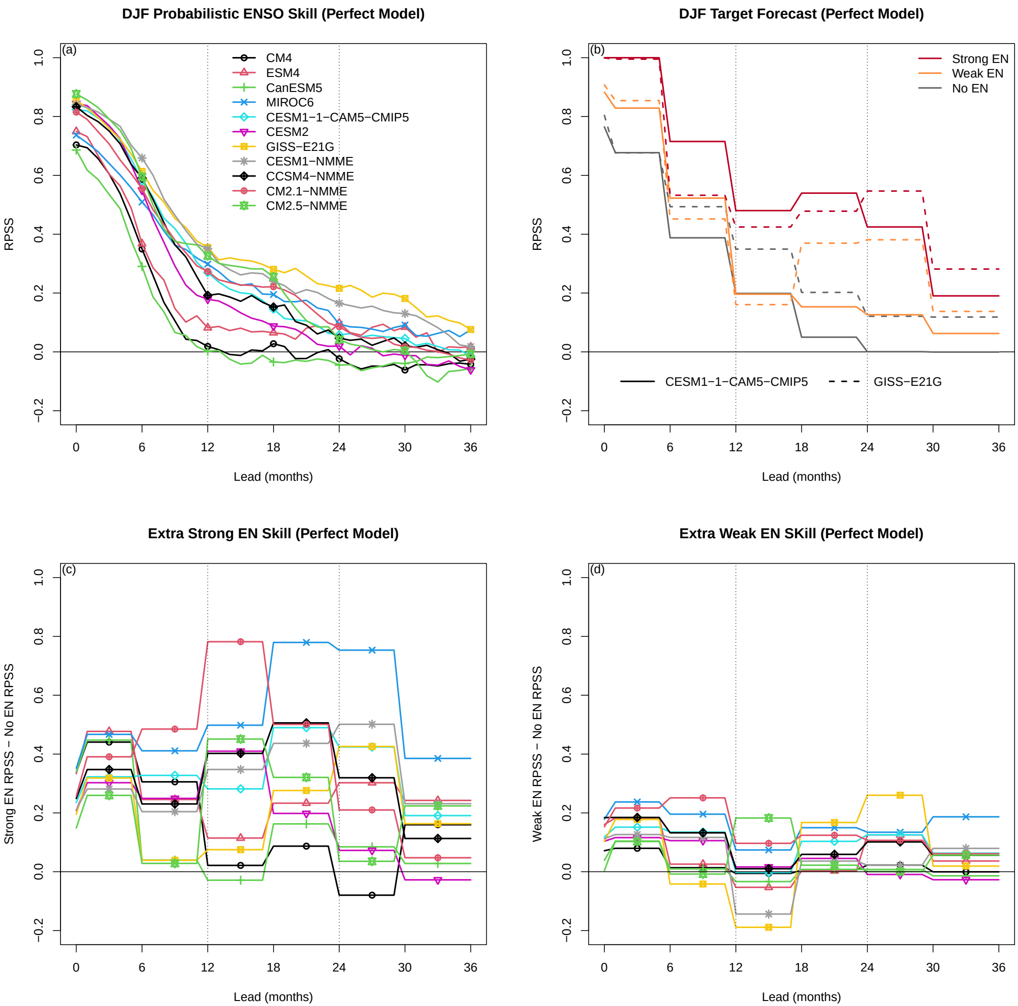

DJF Probabilistic ENSO Skill (Perfect Model)
(a)
CM4
ESM4
CanESM5
MIROC6
CESM1−1−CAM5−CMIP5
CESM2
GISS−E21G
CESM1−NMME
CCSM4−NMME
CM2.1−NMME
CM2.5−NMME
RPSS
Lead (months)
DJF Target Forecast (Perfect Model)
(b)
Strong EN
Weak EN
No EN
CESM1−1−CAM5−CMIP5
GISS−E21G
Extra Strong EN Skill (Perfect Model)
(c)
Strong EN RPSS − No EN RPSS
Extra Weak EN SKill (Perfect Model)
(d)
Weak EN RPSS − No EN RPSS

Figure 2.

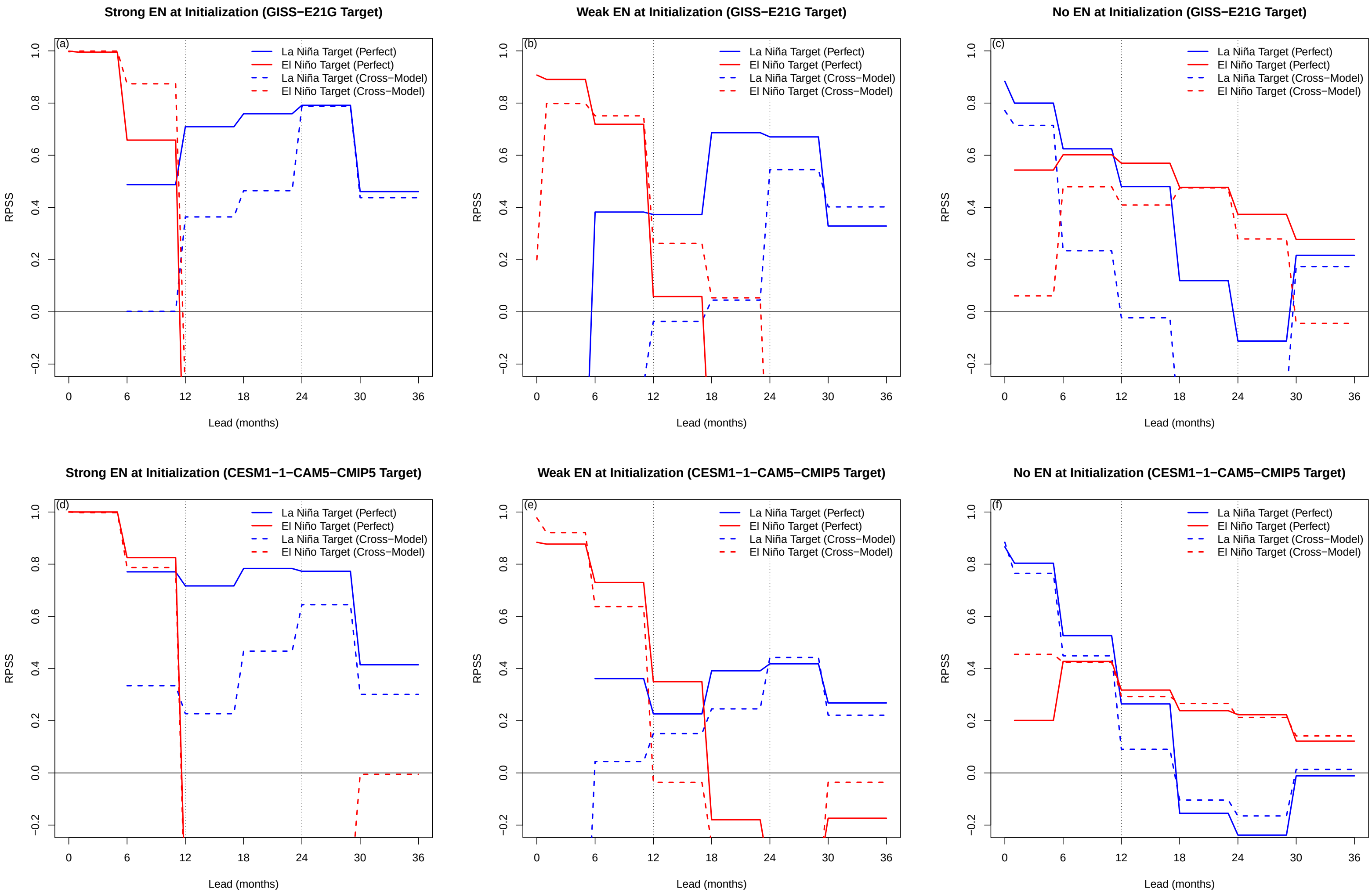

Strong EN at Initialization (GISS−E21G Target)
(a)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)
Weak EN at Initialization (GISS−E21G Target)
(b)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)
No EN at Initialization (GISS−E21G Target)
(c)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)
Strong EN at Initialization (CESM1−1−CAM5−CMIP5 Target)
(d)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)
Weak EN at Initialization (CESM1−1−CAM5−CMIP5 Target)
(e)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)
No EN at Initialization (CESM1−1−CAM5−CMIP5 Target)
(f)
La Niña Target (Perfect)
El Niño Target (Perfect)
La Niña Target (Cross−Model)
El Niño Target (Cross−Model)
RPSS
Lead (months)

Figure 3.

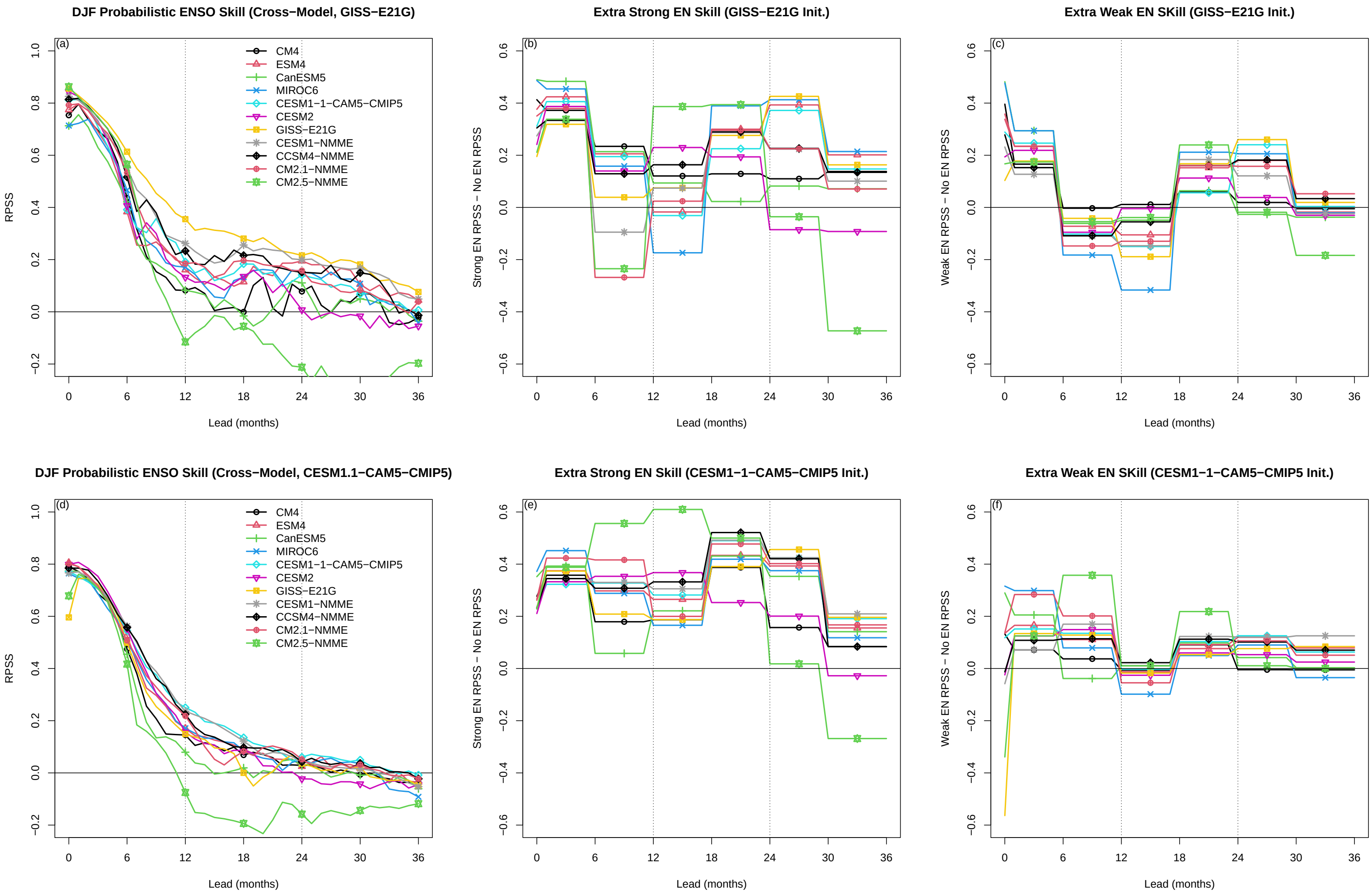
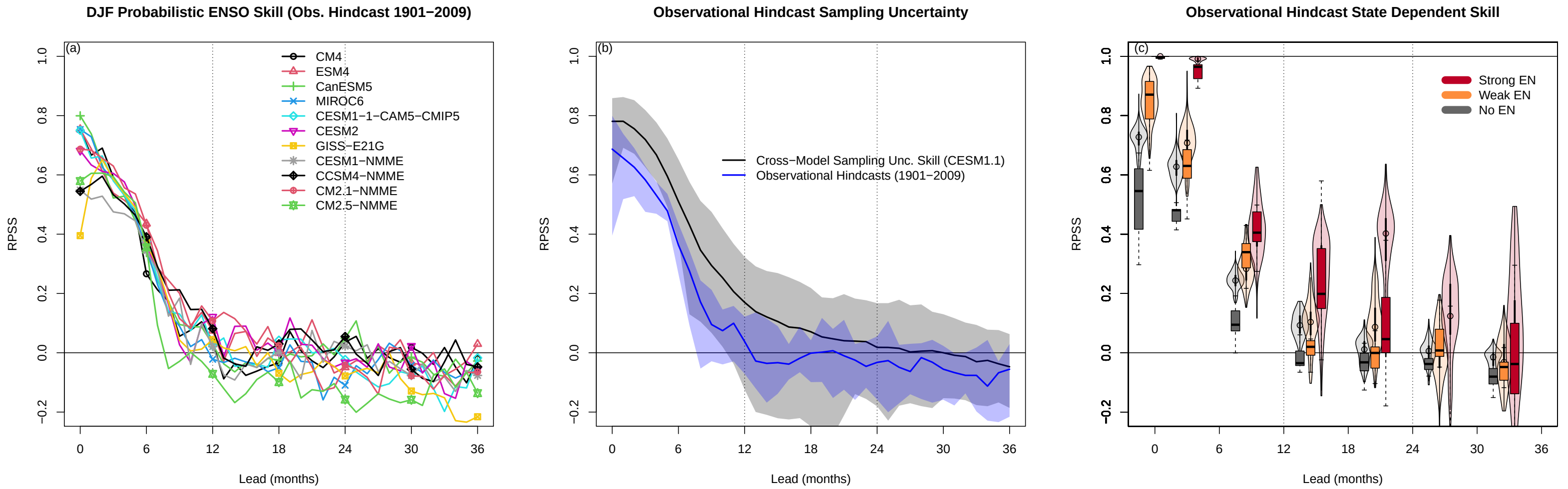

DJF Probabilistic ENSO Skill (Cross−Model, GISS−E21G)
(a)
CM4
ESM4
CanESM5
MIROC6
CESM1−1−CAM5−CMIP5
CESM2
GISS−E21G
CESM1−NMME
CCSM4−NMME
CM2.1−NMME
CM2.5−NMME
RPSS
Lead (months)
Extra Strong EN Skill (GISS−E21G Init.)
(b)
Strong EN RPSS − No EN RPSS
Lead (months)
Extra Weak EN SKill (GISS−E21G Init.)
(c)
Weak EN RPSS − No EN RPSS
Lead (months)
DJF Probabilistic ENSO Skill (Cross−Model, CESM1.1−CAM5−CMIP5)
(d)
CM4
ESM4
CanESM5
MIROC6
CESM1−1−CAM5−CMIP5
CESM2
GISS−E21G
CESM1−NMME
CCSM4−NMME
CM2.1−NMME
CM2.5−NMME
RPSS
Lead (months)
Extra Strong EN Skill (CESM1−1−CAM5−CMIP5 Init.)
(e)
Strong EN RPSS − No EN RPSS
Lead (months)
Extra Weak EN SKill (CESM1−1−CAM5−CMIP5 Init.)
(f)
Weak EN RPSS − No EN RPSS
Lead (months)

Figure 4.



DJF Probabilistic ENSO Skill (Obs. Hindcast 1901−2009)
(a)
CM4
ESM4
CanESM5
MIROC6
CESM1−1−CAM5−CMIP5
CESM2
GISS−E21G
CESM1−NMME
CCSM4−NMME
CM2.1−NMME
CM2.5−NMME
RPSS
Lead (months)
Observational Hindcast Sampling Uncertainty
(b)
Cross−Model Sampling Unc. Skill (CESM1.1)
Observational Hindcasts (1901−2009)
RPSS
Lead (months)
Observational Hindcast State Dependent Skill
(c)
Strong EN
Weak EN
No EN
RPSS
Lead (months)