

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

N. Lenssen^{1,2,3}, P. DiNezio¹, L. Goddard², C. Deser⁴, Y. Kushnir⁵, S. Mason²,
M. Newman⁶, Y. Okumura⁷

¹Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA

²International Research Institute for Climate and Society, Columbia University, Palisades, NY, USA

³Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA

⁴National Center for Atmospheric Research, Boulder, CO, USA

⁵Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

⁶NOAA Physical Sciences Laboratory, Boulder, CO, USA

⁷Jackson 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.

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 Niña 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 results hold regardless of the climate model used to make the predictions as 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 Niño 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,

67 the NMME has hindcasts initialized monthly over 1982-2010 and real-time forecasts ini-
68 tialized beginning in 2011 with lead times up to 11 months (408 forecasts for each CGCM
69 verified in (Barnston et al., 2019)) and the CMIP6 Decadal Climate Prediction Project
70 (DCPP) has hindcasts initialized yearly over 1960-2018 with lead time up to 10 years
71 (59 forecasts for each CGCM verified in (Dunstone et al., 2020)). When evaluating such
72 datasets, it is necessary to evaluate the skill of a forecast system over all hindcasts to
73 maximize sample size in the statistical estimates of forecast skill. However, pooling all
74 forecasts, particularly by ENSO state at initialization, has the potential to obfuscate the
75 underlying sources of long-lead ENSO skill if predictability is state-dependent.

76 In this study, we investigate the model and initial state dependence of multi-year
77 ENSO prediction skill. We explore initial ENSO states in terms of phase (El Niño, neu-
78 tral, La Niña) and intensity (strong, weak) providing multi-year skill. To this aim, we
79 construct and analyze a massive multi-model ensemble of model analog climate hind-
80 casts to identify initial states that lead to multi-year predictive skill. The model ana-
81 log method (Ding et al., 2018, 2019, 2020) is used to make forecasts by first identifying
82 states in a “library” of CGCM output that best match the initial state. Then, ense-
83 mble forecasts are issued according to how each of these states evolved in the CGCM. These
84 forecasts are appropriate to use to investigate ENSO predictability as they have tropi-
85 cal Pacific skill equal to or exceeding state-of-the-art initialized dynamical forecast sys-
86 tems (Ding et al., 2018). In addition, the very low computational cost allows the gen-
87 eration of very large ensemble hindcasts based on multiple CMIP-class CGCMs with leads
88 of 3+ years. Together, these features of our technique enabled us to investigate the model
89 and state dependence of 2 year ENSO prediction skill.

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

102 **2 Data and Methods**

103 **2.1 Data**

104 We use long pre-industrial control simulations of at least 500 years in duration from
105 11 state-of-the-art CGCMs to issue model-analog forecasts and to perform the verifica-
106 tions in Sections 3 and 4. The 11 CGCMs are seven CMIP-class CGCMs and the four
107 available control runs from NMME CGCMs (Table S1). All gridded products are regrid-
108 ded to a common $2^\circ \times 2^\circ$ grid before use in any analyses. The monthly mean sea-surface
109 temperature (SST) or “tos” fields and sea-surface height (SSH) or “zos” fields are used.
110 SST and SSH anomalies are created by removing the monthly climatologies.

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

117 ENSO events are defined according to quantiles of the Oceanic Niño Index (ONI)
 118 which is the seasonal (3 month) average SST anomaly over the Niño 3.4 region (5N-5S,
 119 170W-120W). These quantiles are calculated for each season for each CGCM as well as
 120 the observations. El Niño events are defined as the upper quartile, or values above the
 121 75th percentile, of ONI. Similarly, La Niña events are defined as the lower quartile, or
 122 values below the 25th percentile of ONI. This method is useful when comparing ENSO-
 123 state prediction across different CGCMs as it reduces the bias from different CGCM ENSO
 124 mean states and variabilities (Gonzalez & Goddard, 2016).

125 2.2 Model Analog Forecasts

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

133 Best analogs are found by finding the best matches of SST and SSH fields between
 134 the initial state and all states within the same month in the CGCM library. The initial
 135 and library fields are compared over the entire tropical Indo-Pacific basin (30S–30N, 30E–80W).
 136 For each time step in the library, we calculate the root mean square (RMS) distance from
 137 the initial SST and SSH fields to the corresponding library fields. Here, all fields are nor-
 138 malized to have unit variance to allow adding the distances between the initial and li-
 139 brary SST and SSH fields as well as accounting for biases in variability between datasets.
 140 These distances are ranked in ascending order and the evolution of the 15 states clos-
 141 est to the initial field are used to create an ensemble forecast.

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

148 2.3 Forecast Verification

149 The probabilistic skill of the ENSO state forecasts is determined using RPSS, a stan-
 150 dard skill metric for probabilistic skill (Jolliffe & Stephenson, 2012; Mason, 2018). RPSS
 151 is a measure of both a forecast’s resolution, or whether the outcome differs given differ-
 152 ent forecasts, as well as a forecast’s reliability, or how well the forecasted probabilities
 153 match the observed rate of events (Mason, 2018). The RPSS is a skill score comparing
 154 the Ranked Probability Score (RPS; (Epstein, 1969; Murphy, 1971)) of the forecast of
 155 interest to a climatological forecast. It is defined in such a way that an RPSS of 1.0 in-
 156 dicates a perfect forecast, an RPSS indicates that a forecast is equivalent to climatol-
 157 ogy, and a negative forecast indicates a forecast that is less skillful than the climatolog-
 158 ical rate of ENSO events.

159 3 Perfect Model Hindcast Experiment

160 We first investigate the perfect model skill, or the skill of a model predicting its own
 161 dynamics. That is, we use the same CGCM as both the target states as well as the li-
 162 brary, omitting the state we are trying to predict as a possible analog. Perfect model skill
 163 is generally an upper bound of skill for the ENSO system. When predicting the state
 164 of ENSO in December-February (DJF), the peak season of ENSO, all models have pos-

165 itive ranked probability skill scores (RPSS) at leads of up to 12 months (Figure 1a). RPSS
166 is a measure of probabilistic forecast skill where a value of zero indicates a forecast is on
167 par with a forecast of climatological probabilities and positive values indicate that the
168 model analog forecasts outperform climatological forecasts. All but two of the eleven mod-
169 els in the study have positive RPSS out to at least 24 months, indicating that a range
170 of CGCMs with varied ENSO dynamics exhibit “perfect model” multi-year ENSO pre-
171 dictability (Figure 1a). These findings agree with theoretical calculations of ENSO pre-
172 dictability of around 3 years (Newman & Sardeshmukh, 2017), the skill of initialized dy-
173 namical forecasts (DiNezio, Deser, Okumura, & Karspeck, 2017; Dunstone et al., 2020;
174 Wittenberg et al., 2014), and multi-model long lead skill of model analog forecasts (Lou
175 et al., 2023).

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

186 For perfect-model forecasts from CESM1.1, by far the greatest year-2 skill comes
187 from forecasts initialized out of strong El Niño events as seen by the large difference be-
188 tween the strong El Niño line and the no El Niño RPSS skill at leads greater than 12
189 months (Figure 1b). The strong El Niño skill between leads 12-24 months is expected
190 due to the strong tendency for La Niña to occur after strong El Niño events. The strong
191 El Niño skill seen at leads 24-36 months is due to the high predictability of two year La
192 Niña events following large El Niño events that has been previously shown in CESM1
193 (DiNezio, Deser, Okumura, & Karspeck, 2017). The same dramatic increase in year 2+
194 skill does not occur from forecasts initialized during weak El Niño events, as shown be-
195 tween the negligible difference between the weak El Niño and no El Niño skill (Figure
196 1b). This analysis was performed with all CGCMs in the study, leading to qualitatively
197 similar results (not shown).

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

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

211 With this greater predictability out of strong El Niño, it is natural to ask if the year
212 2+ skill is indeed due to greater predictability of subsequent La Niña events of one- or
213 two-year duration. To test this, we take each of the initial states used in Figure 1b and
214 decompose the forecast skill according to the true ENSO state upon verification. Results
215 with two of the CGCMs with greatest multiyear skill, GISS-E2.1G and CESM1.1, are

216 shown as illustrative examples (Figure 2), but similar results are found for all 11 CGCMs
217 in the study (not shown).

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

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

232 4 Cross-Model Hindcast Experiment

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

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

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

259 As with the perfect model hindcasts, we decompose the cross-model RPSS by state
260 at initialization. Again, we see that most of the year-2 skill comes from predictions out
261 of strong El Niño events (Figures 3b,d). When using GISS-E2.1G as the hindcast tar-
262 get, all but three CGCMs show better 12-18 month skill and all CGCMs show better 18-
263 24 month out of strong El Niño events than other initial states (Figure 3b). When pre-
264 dicting the more realistic CESM1.1 ENSO, all CGCMs have much more skill when ini-
265 tialized during strong El Niño events when compared with no El Niño events (Figure 3e).

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

268 **5 Observational Hindcast Experiment**

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

279 We expect substantial sampling uncertainty in quantifying skill over the 109-year
280 hindcast due to the limited sample size in the verification statistics as well as the known
281 multidecadal variability in ENSO predictability (Wittenberg, 2009; Wittenberg et al.,
282 2014; Lou et al., 2023). To make fair comparisons between the observational hindcasts
283 here and the cross-model hindcasts in Section 3, we quantify this sampling uncertainty
284 in the observational hindcast. We use a bootstrapping approach in which we create and
285 verify 200 hindcasts using analogs from each CGCM over random 109-year periods of
286 the 1,800 year CESM1.1 piControl. The 95% likely skill from the subsampled 109 year
287 cross-model CESM1.1 hindcasts and the range of the observational hindcast skill over-
288 lap for all leads but 4 months (Figure 4b). Thus, we cannot reject the hypothesis that
289 DJF skill is lower when predicting the observed ENSO system than when predicting the
290 CESM1.1 ENSO system.

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

304 **6 Summary and Discussion**

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

309 These findings are important for both climate predictability research and for cli-
310 mate service applications using seasonal to multi-year predictions. Research into ENSO
311 and climate predictability generally focuses on metrics of skill aggregated over all fore-
312 casts, a required assumption given the small hindcasts available. As such, multiple stud-
313 ies have claimed that ENSO can be predicted skillfully into the second year (Dunstone
314 et al., 2020; Gonzalez & Goddard, 2016; Ham et al., 2019; Wang et al., 2023). Our find-

ings 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

All code used to process the data, run the experiments, and verify forecasts can be found at <https://github.com/nlenssen/LongLeadENSO/>. All raw data is publicly available as well as on reasonable request from the authors.

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.

- 361 Becker, E. J., Kirtman, B. P., L’Heureux, M., Muñoz, Á. G., & Pegion, K. (2022).
 362 A decade of the North American Multimodel Ensemble (NMME): Research,
 363 application, and future directions. *Bulletin of the American Meteorological*
 364 *Society*, *103*(3), E973–E995.
- 365 Cai, W., Santoso, A., Wang, G., Wu, L., Collins, M., Lengaigne, M., . . . Timmer-
 366 mann, A. (2020). ENSO Response to Greenhouse Forcing. *El Niño Southern*
 367 *Oscillation in a Changing Climate*, 289–307.
- 368 Capotondi, A., Deser, C., Phillips, A., Okumura, Y., & Larson, S. (2020). ENSO
 369 and Pacific decadal variability in the Community Earth System Model version
 370 2. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2019MS002022.
- 371 DiNezio, P. N., Deser, C., Karspeck, A., Yeager, S., Okumura, Y., Danabasoglu, G.,
 372 . . . Meehl, G. A. (2017). A 2 year forecast for a 60–80% chance of la niña in
 373 2017–2018. *Geophysical Research Letters*, *44*(22), 11–624.
- 374 DiNezio, P. N., Deser, C., Okumura, Y., & Karspeck, A. (2017). Predictability of 2-
 375 year La Niña events in a coupled general circulation model. *Climate dynamics*,
 376 *49*(11), 4237–4261.
- 377 Ding, H., & Alexander, M. A. (2023). Multi-year predictability of global sea surface
 378 temperature using model-analogs. *Geophysical Research Letters (Accepted)*.
- 379 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018). Skillful cli-
 380 mate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *Journal*
 381 *of Climate*, *31*(14), 5437–5459.
- 382 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2019). Diagnosing
 383 secular variations in retrospective ENSO seasonal forecast skill using cmip5
 384 model-analogs. *Geophysical Research Letters*, *46*(3), 1721–1730.
- 385 Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2020). Relating
 386 CMIP5 model biases to seasonal forecast skill in the tropical Pacific. *Geophys-
 387 ical Research Letters*, *47*(5), e2019GL086765.
- 388 Dunstone, N., Smith, D., Yeager, S., Danabasoglu, G., et al. (2020). Skilful inter-
 389 annual climate prediction from two large initialised model ensembles. *Environ-
 390 mental Research Letters*, *15*(9), 094083.
- 391 Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robin-
 392 son, N. (2014). Do seasonal-to-decadal climate predictions underestimate
 393 the predictability of the real world? *Geophysical Research Letters*, *41*(15),
 394 5620–5628.
- 395 Epstein, E. S. (1969). A scoring system for probability forecasts of ranked cate-
 396 gories. *Journal of Applied Meteorology*, *8*(6), 985–987.
- 397 Geng, T., Jia, F., Cai, W., Wu, L., Gan, B., Jing, Z., . . . McPhaden, M. J. (2023).
 398 Increased occurrences of consecutive La Niña events under global warming.
 399 *Nature*, *619*(7971), 774–781.
- 400 Gonzalez, P. L., & Goddard, L. (2016). Long-lead ENSO predictability from CMIP5
 401 decadal hindcasts. *Climate Dynamics*, *46*(9), 3127–3147.
- 402 Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO
 403 forecasts. *Nature*, *573*(7775), 568–572.
- 404 Jin, E. K., Kinter, J. L., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B., . . . others
 405 (2008). Current status of ENSO prediction skill in coupled ocean–atmosphere
 406 models. *Climate Dynamics*, *31*, 647–664.
- 407 Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: A practitioner’s*
 408 *guide in atmospheric science* (2nd ed.). West Sussex, England: Wiley.
- 409 Lalouaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S.,
 410 Buizza, R., . . . others (2018). CERA-20C: a coupled reanalysis of the twen-
 411 tieth century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172–
 412 1195.
- 413 Lenssen, N., Goddard, L., & Mason, S. (2020). Seasonal Forecast Skill of ENSO
 414 Teleconnection Maps. *Weather and Forecasting*, *35*(6), 2387–2406.
- 415 L’Heureux, M. L., Levine, A. F., Newman, M., Ganter, C., Luo, J.-J., Tippett,

- 416 M. K., & Stockdale, T. N. (2020). ENSO prediction. *AGU Monograph: El*
 417 *Niño Southern Oscillation in a changing climate*, 227–246.
- 418 Liu, T., Song, X., Tang, Y., Shen, Z., & Tan, X. (2022). ENSO predictability over
 419 the past 137 years based on a CESM ensemble prediction system. *Journal of*
 420 *Climate*, 35(2), 763–777.
- 421 Lou, J., Newman, M., & Hoell, A. (2023). Multi-decadal variation of ENSO forecast
 422 skill since the late 1800s. *npj Climate and Atmospheric Science*, 6(1), 89.
- 423 Mason, S. J. (2018). Guidance on verification of operational seasonal climate fore-
 424 casts. *World Meteorological Organization, Commission for Climatology XIV*
 425 *Technical Report*.
- 426 Mason, S. J., Ferro, C. A., & Landman, W. A. (2021). Forecasts of “normal”. *Quar-*
 427 *terly Journal of the Royal Meteorological Society*, 147(735), 1225–1236.
- 428 Mason, S. J., & Goddard, L. (2001). Probabilistic precipitation anomalies associated
 429 with ENSO. *Bulletin of the American Meteorological Society*, 82(4), 619–638.
- 430 Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied*
 431 *Meteorology*, 10(1), 155–156.
- 432 Newman, M., & Sardeshmukh, P. D. (2017). Are we near the predictability limit of
 433 tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*,
 434 44(16), 8520–8529.
- 435 Nissan, H., Goddard, L., de Perez, E. C., et al. (2019). On the use and misuse of cli-
 436 mate change projections in international development. *Wiley Interdisciplinary*
 437 *Reviews: Climate Change*, 10(3), e579.
- 438 Ropelewski, C. F., & Halpert, M. S. (1986). North American precipitation and tem-
 439 perature patterns associated with the El Niño/Southern Oscillation (ENSO).
 440 *Monthly Weather Review*, 114(12), 2352–2362.
- 441 Suarez, M. J., & Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal*
 442 *of Atmospheric Sciences*, 45(21), 3283–3287.
- 443 Tippett, M. K., L’Heureux, M. L., Becker, E. J., & Kumar, A. (2020). Excessive mo-
 444 mentum and false alarms in late-spring ENSO forecasts. *Geophysical research*
 445 *letters*, 47(8), e2020GL087008.
- 446 Tippett, M. K., Ranganathan, M., L’Heureux, M., Barnston, A. G., & DelSole, T.
 447 (2019). Assessing probabilistic predictions of ENSO phase and intensity from
 448 the North American Multimodel Ensemble. *Climate Dynamics*, 53, 7497–
 449 7518.
- 450 Wang, H., Hu, S., & Li, X. (2023). An interpretable deep learning ENSO forecasting
 451 model. *Ocean-Land-Atmosphere Research*, 2, 0012.
- 452 Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simu-
 453 lations? *Geophysical Research Letters*, 36(12).
- 454 Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A., & Zeng, F. (2014).
 455 ENSO modulation: Is it decadal predictability? *Journal of Climate*, 27(7),
 456 2667–2681.
- 457 Wu, X., Okumura, Y. M., Deser, C., & DiNezio, P. N. (2021). Two-year dynami-
 458 cal predictions of ENSO event duration during 1954–2015. *Journal of Climate*,
 459 34(10), 4069–4087.
- 460 Wu, X., Okumura, Y. M., & DiNezio, P. N. (2019). What controls the duration of
 461 El Niño and La Niña events? *Journal of Climate*, 32(18), 5941–5965.
- 462 Wu, X., Okumura, Y. M., & DiNezio, P. N. (2021). Predictability of El Niño dura-
 463 tion based on the onset timing. *Journal of Climate*, 34(4), 1351–1366.

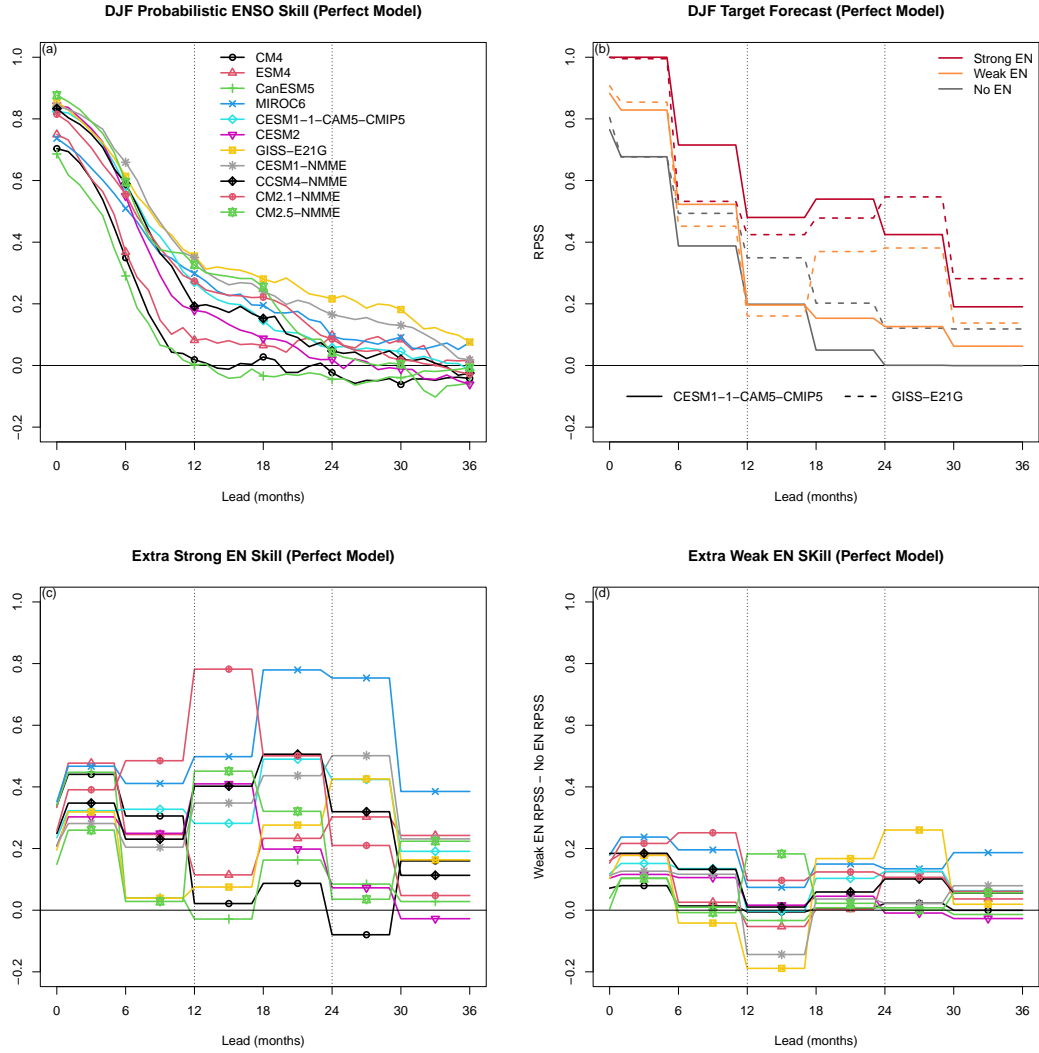


Figure 1. The model analog DJF RPSS skill for (a) perfect model hindcasts of all 11 CGCMs used in the study and (b) perfect model hindcasts stratified by ENSO state at initialization for two example CGCMs, CESM1.1 and GISS-E2.1G. The extra skill added when initializing during El Niño conditions is shown 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.

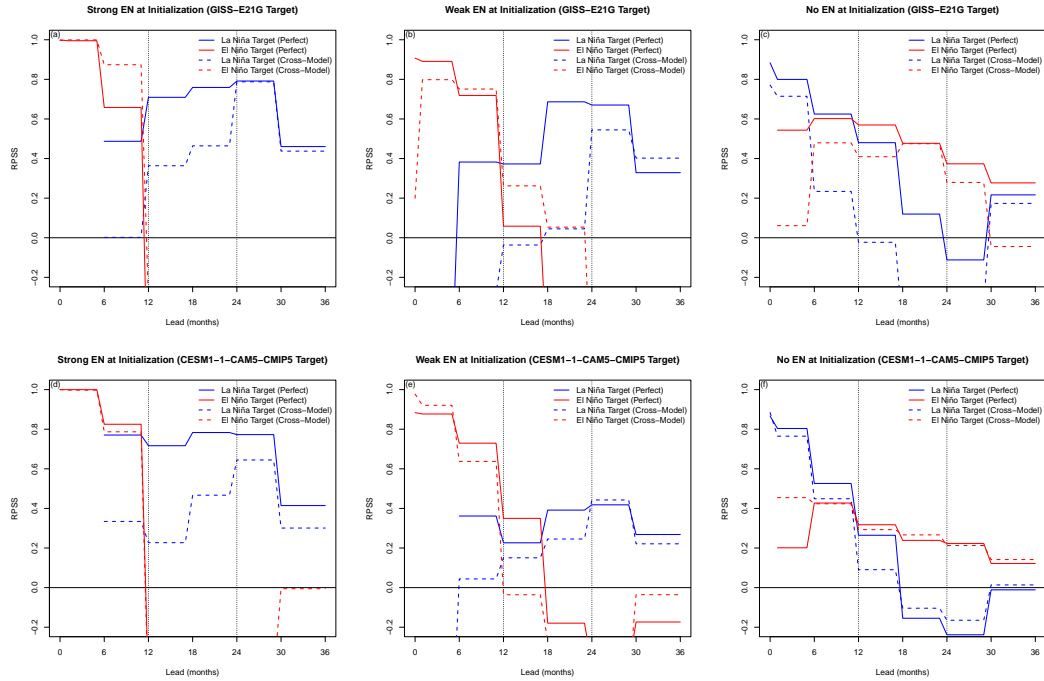


Figure 2. A second decomposition of the skill analysis in Figure 1b in which the skill is stratified by initial state in CESM1.1 and GISS-E2.1G where (a,d) show the skill of forecasts initialized during strong EN, (b,e) during weak EN, and (c,f) during no EN. The top row shows forecasts predicting the 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.

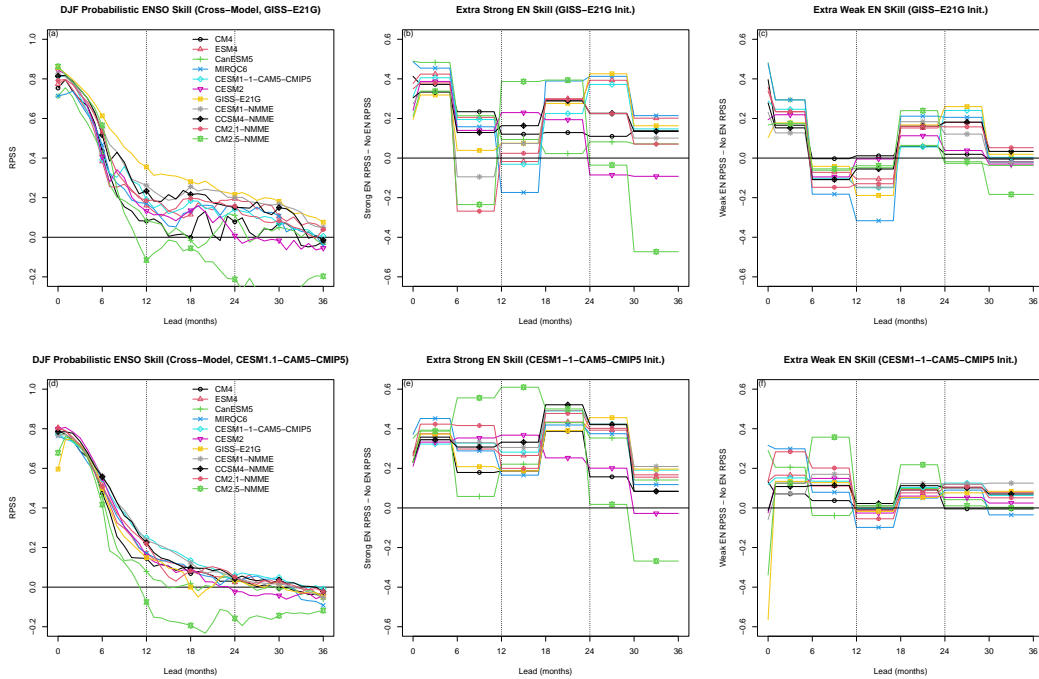


Figure 3. The model analog DJF RPSS skill of cross-model hindcasts using libraries from all 11 CGCMs to predict the piControl of (a) GISS-E2.1G and (d) CESM1.1. The remaining panels follow the analysis presented in Figures 1c,d by summarizing the extra skill in (b,e) forecasts initialized during strong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

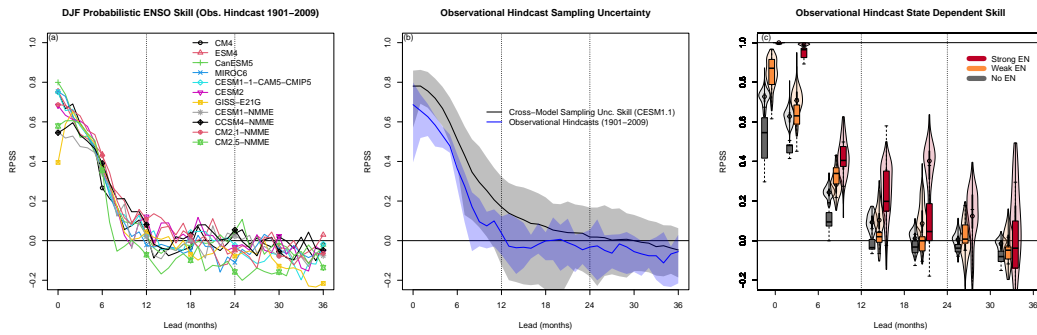


Figure 4. The model analog DJF RPSS skill of forecasts using libraries from all 11 CGCMs to predict (a) the observational record from 1901-2009 (109 years). The grey in (b) shows the 95% confidence interval due to sampling uncertainty estimated as the empirical median and 95% confidence interval of 200 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.