Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL080598

Diagnosing Secular Variations in Retrospective ENSO Seasonal Forecast Skill Using CMIP5 Model-Analogs

Hui Ding[1,2], Matthew Newman[1,2], Michael A. Alexander[2], and Andrew T. Wittenberg[3]

[1]CIRES, University of Colorado Boulder, Boulder, CO, USA, [2]NOAA Earth Systems Research Laboratory, Boulder, CO, USA, [3]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Key Points:
- Seasonal tropical Indo-Pacific hindcasts for 1961-2015 are made from 28 different CMIP5 models, with some as skillful as operational CGCMs
- ENSO hindcast skill has no trend, but it is higher in strong-ENSO epochs (1980s and 1990s) and lower in weak-ENSO epochs (1970s and 2000s)
- Including the effects of projected external radiative forcings improves SST hindcast skill but not within the ENSO region

Supporting Information:
- Supporting Information S1

Correspondence to:
H. Ding,
hui.ding@noaa.gov

Citation:
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*, 1721–1730. https://doi.org/10.1029/2018GL080598

Received 24 SEP 2018
Accepted 24 JAN 2019
Accepted article online 13 FEB 2019
Published online 14 FEB 2019

Abstract Retrospective tropical Indo-Pacific forecasts for 1961–2015 are made using 28 models from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) plus four models from the North American Multi-Model Ensemble (NMME), using a model-analog technique. Forecast ensembles are extracted from preexisting model simulations, by finding those states that initially best match an observed anomaly and tracking their subsequent evolution, requiring no additional model integrations. Model-analog forecasts from the 10 "best" CMIP5 models have skill for sea surface temperature and precipitation comparable to that of both the NMME model-analog forecast ensemble and (since 1982) traditional assimilation-initialized NMME hindcasts. The El Niño–Southern Oscillation (ENSO) forecast skill has no trend over the 55-year period, and its decadal variations appear largely random, although the skill does improve during epochs of increased ENSO activity. Including the CMIP5-projected effects of external radiative forcings improves the tropical sea surface temperature skill of the model-analog forecasts but not within the ENSO region.

Plain Language Summary Seasonal forecasts are made by starting a climate model from an initial estimate of the latest global three-dimensional ocean, atmosphere, and land conditions and then using supercomputers to run the model's equations forward in time. These extensive calculations are only feasible at a few national operational centers and large research institutions. However, many similar models are also used for long simulations of the Earth's preindustrial climate, which are made freely available for climate change studies. We investigated whether seasonal forecasts might be drawn from the information already existing within these simulations, instead of by making new model computations. Within each simulation, we determine the best matches, or "model-analogs," to current observed tropical Indo-Pacific ocean surface conditions. How these analogs evolve over the next several months within each long simulation is then its seasonal forecast. We find that this much less expensive model-analog technique is as skillful as the more traditional forecasting method. We then employed it to make forecasts during 1961–2015, using 28 existing climate simulations, significantly expanding previous forecast efforts. This study suggests that with little additional effort, sufficiently realistic and long existing climate model simulations can provide the basis for skillful seasonal forecasts. That is, anyone can be a climate forecaster.

1. Introduction

Seasonal forecast skill of El Niño–Southern Oscillation (ENSO) has varied substantially over the last few decades (Barnston et al., 2012). Relatively low skill existed for much of the early 2000s, leading some researchers to suggest that the inherent predictability of ENSO itself had decreased, possibly due to a fundamentally different tropical mean state (Barnston et al., 2012; Zhao et al., 2016) partly driven by anthropogenic climate change (Yeh et al., 2009) that perhaps even altered the nature of ENSO itself (Lee & McPhaden, 2010). Other studies, however, suggested that such changes in base state and predictability might be a residual, rather than cause, of randomly clustered ENSO events arising due to the unpredictable variations of weather noise forcing of oceanic ENSO precursors (Flügel et al., 2004; Karamperidou et al., 2014; Kumar et al., 2015; Lee et al., 2016; Newman & Sardeshmukh, 2017; Newman et al., 2011; Ogata et al., 2013; Vecchi et al., 2006; Watanabe et al., 2012; Watanabe & Wittenberg, 2012; Wittenberg, 2009; Wittenberg et al., 2014). The examination of this question requires a lengthy record of retrospective forecasts, or "hindcasts," of ENSO from earlier periods. Unfortunately, hindcast data sets made by coupled general circulation models (CGCMs) are generally available only after about 1982 (e.g., Barnston et al., 2012; Kirtman et al., 2014) apart from just a few models (Huang et al., 2017; Weisheimer

et al., 2009), so it is not entirely clear if the early 2000s were unique and/or if there is a long-term trend in potential ENSO prediction skill.

Global climate forecasts are typically made by initializing coupled climate models with full-field estimates of the global climate state determined using a sophisticated data assimilation system, perturbing this state several times to create an ensemble whose spread samples initial uncertainty, and finally integrating each ensemble member forward several months, yielding the forecast ensemble. This procedure, repeated monthly, requires considerable computational resources and therefore is undertaken only by large operational forecasting centers. Generating large hindcast data sets is even more resource intensive and additionally requires an extensive atmospheric and oceanic reanalysis effort.

Ding et al. (2018; hereafter D18) recently demonstrated an alternative forecast method in which predictions could be drawn from a large library of states previously simulated by a CGCM. Specifically, for each forecast month from 1982 to 2009, a set of states taken directly from a long uninitialized control CGCM simulation were chosen by matching their sea surface temperature (SST) and sea surface height (SSH) components to observed initial tropical Indo-Pacific SST and SSH anomalies. The subsequent evolution of these states (called "model-analogs" by D18) within the control simulation then provided forecast ensembles at various lead times. D18 showed that tropical Indo-Pacific SST forecast skill of four model ensembles taken from the North American Multi-Model Ensemble (NMME; Kirtman et al., 2014), in which seasonal forecasts are initialized with observations using the standard operational approach, was generally matched (and occasionally exceeded) by this model-analog method applied to the corresponding long control runs from each NMME model. This immediately suggests that *any* CGCM with a sufficiently long uninitialized control simulation could also produce seasonal forecasts, using only initial observations of SST and SSH, without otherwise building a comprehensive assimilation-initialized forecast system around the model. That is, the model-analog technique makes it possible to generate forecasts from a great many more models than only the operational models.

In this paper, we use the model-analog technique to significantly expand the preexisting tropical Indo-Pacific hindcast data set, covering the years 1961–2015 and using not only NMME models (Table S1) but also 28 CGCMs available through the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Table S2 in the supporting information). CMIP5 climate model simulations have been widely employed in numerous ENSO studies (e.g., Bellenger et al., 2014; Chen et al., 2017). The accessibility of CMIP5 output makes it particularly attractive for conducting climate forecasts using the model-analog method, even though CMIP5 models have generally not been used to make ensemble forecasts at their respective development centers. Using a larger and more diverse set of models and a longer period of record allows for more robust determination of both the spatial and seasonal dependence of tropical Indo-Pacific skill and long-term variations in skill over the past 55 years.

2. Data and Methods

2.1. Model-Analogs

Following D18, we choose analogs at each time t by minimizing a distance between a target state vector $\mathbf{x}(t)$ and each library state vector $\mathbf{y}(t')$, where the target state is defined as the observed state at the initialization time, and the library consists of all states obtained from a CGCM control simulation. D18's simple distance metric was the root-mean-square (RMS) difference between two variables chosen from the full state vectors $\mathbf{x}$ and $\mathbf{y}$: spatial maps of SST and SSH anomalies (SSTA and SSHA) within the domain bounded by 30°E to 80°W and 30°S to 30°N (see the supporting information). Distances are then ranked in ascending order, and the K states closest to the target state are chosen as the model-analog ensemble members, indicated by the set $\{\mathbf{y}(t_1'),\mathbf{y}(t_2'),...,\mathbf{y}(t_k'),...,\mathbf{y}(t_K')\}$ with k the analog index and t_k' the time of this analog in the library. D18 found that for data libraries on the order of several hundred years in length, an ensemble size of K~10–20 gave the best results. The subsequent evolution of this ensemble within the control simulation, $\{\mathbf{y}(t_1'+\tau),\mathbf{y}(t_2'+\tau),...,\mathbf{y}(t_k'+\tau),...,\mathbf{y}(t_K'+\tau)\}$, is the model-analog forecast ensemble for $\mathbf{x}(t+\tau)$ at lead time τ months. See D18 for additional discussion of important details of the technique.

The model-analog technique can be used to make forecasts of not only the subset of variables used to define the analogs (e.g., tropical Indo-Pacific SSHA and SSTA) but also any other model quantity associated with

the library states and their subsequent evolution within the control simulation. Therefore, in this study we extend the technique by making model-analog forecasts for anomalous precipitation as well.

2.2. Model and Observational Data Sets

The library data sets are monthly mean data from preindustrial control simulations of the 28 different CMIP5 models listed in Table S2, which have varying simulation lengths (≥200 years). Monthly anomalies are calculated for each model by subtracting its monthly climatology, determined over the full length of each model simulation. This effectively removes mean biases relative to observations.

Observations used to determine the initial model-analog states were obtained from monthly mean SSTs in the Hadley Sea Ice and Sea Surface Temperature v1.1 data set (Rayner et al., 2003) and SSHs from the European Centre for Medium-Range Weather Forecasts (ECMWF) Ocean Reanalysis System 4 data set (Balmaseda et al., 2013), for 1961–2015. Anomalies were determined by removing the monthly mean 1961–2015 climatology. These were also used for hindcast verification, along with precipitation anomalies derived from the Global Precipitation Climatology Project version 2.3 precipitation data set (Adler et al., 2003) but for the years 1979–2015.

All model and observed data were interpolated onto a common 2° longitude by 2° latitude grid prior to analysis.

2.3. Hindcast Determination and Skill Metrics

We applied the model-analog technique to each CMIP5 simulation to make retrospective forecasts (hindcasts) from tropical Indo-Pacific SST and SSH initial conditions during 1961–2015 at leads of 1–12 months. From each simulation, a 10-member analog ensemble ($K = 10$) was drawn; results were similar for $K = 15$. All ensemble members were equally weighted, since D18 found no improvement for unequal weighting strategies. Each initial observed state (SSHA and SSTA) was detrended using the method described below in section 3 before analogs were determined.

The SST and precipitation hindcasts were evaluated with two deterministic skill measures: anomaly correlation (AC) of the ensemble mean forecast anomaly with the verification anomaly and RMS error skill score (Barnston et al., 2015) defined as $1 - \varepsilon$, where ε is the RMS error of the ensemble-mean forecast standardized by the observed climatological standard deviation. Probabilistic skill was assessed using the ranked probability skill score (RPSS; Weigel et al., 2007) applied to the observed and forecast tercile distributions. The patterns of the metrics are generally similar, so only AC and RPSS are shown here (RMS metrics are in the supporting information). Additionally, each hindcast's ENSO skill was assessed with the pattern correlation between predicted and observed SSTA fields within an "ENSO" region bounded by 20°S to 20°N, 170°E to 80°W.

3. Accounting for Externally Forced Trends

In nature, SSH and SST anomalies arise from both externally forced (greenhouse gas, aerosols, and other radiative forcing) 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. D18 did not include secular trend components within their initial observed states, which may explain why their model-analog forecasts were less skillful than the corresponding assimilation-initialized forecasts in regions where external radiative forcing may already have affected SST, such as over the tropical Indian Ocean and northwestern and southwestern tropical Pacific (e.g., Deser et al., 2010; Knutson et al., 2014; Solomon & Newman, 2012).

Here we aim to improve this aspect of the model-analog technique by using the time-evolving multimodel ensemble mean of the historical and Representative Concentration Pathway 4.5 (RCP4.5) CMIP5 simulations to estimate the externally forced secular trend. We then use that estimate to identify and remove the trend component from the observed SSH and SST anomalies prior to searching for model-analogs in the fixed-forcing control simulations and finally add the predicted trend component back to the final forecast (see the supporting information for details). In essence, the CMIP5 ensemble mean predicts the externally forced component, and the model-analog technique predicts the internal climate anomaly. We used all 45 CMIP5 historical simulations to estimate the externally forced signal, although using just those 28 models corresponding to our

model-analog ensemble yielded essentially the same results. For each initial state at time t, the trend component was redetermined from the 1961–2015 period except for data from the interval (t, t + 5 years), which was withheld to ensure that the trend component of each hindcast was independent of the verification data.

We evaluated this technique by applying it to the NMME model-analog hindcasts (Table S1) previously used in D18 and then comparing to the NMME assimilation-initialized hindcasts. Using the CMIP5 historical run multimodel ensemble mean to represent the externally forced component significantly improves model-analog SST forecast skill over the tropical Indian Ocean, the western tropical Pacific, and the South Pacific. However, including the trend component has almost no impact on skill within the ENSO region (Figure S1), where SST trends are small and not statistically significant (Solomon & Newman, 2012) apart from Niño4 (Newman et al., 2018). That is, for the 1982–2009 period, accounting for the externally forced component now allows the NMME model-analog ensemble-mean forecast skill to match (or exceed) NMME assimilation-initialized hindcast skill throughout the tropics and well into the subtropics (Figure S2). RMS skill score (Figure S3) and RPSS (Figure S4) show similar results. Consequently, we next extend this model-analog technique to the CMIP5 models and also expand the analysis to a longer hindcast epoch.

4. ENSO Skill, 1961–2015

We first assess the skill of the CMIP5 model-analog hindcasts of internal variability alone (i.e., prior to including the projected externally forced component), as D18 did for the NMME models. Figure 1 shows the ensemble-mean AC skill from every model's 6-month lead SST forecasts, compared to the skill from two different multimodel ensemble means: one using all 28 models and another based on a selected 10-model subset (see below). All CMIP5 models generate model-analog hindcasts with skill in the central and eastern equatorial Pacific, with correlations ranging between 0.6 and 0.7. Some models are also skillful in the tropical Indian Ocean and the northwestern tropical Pacific, with correlations of 0.4–0.5. In these regions, SST anomalies are strongly influenced by ENSO via the atmospheric bridge (Alexander et al., 2002), although ENSO-independent variability also occurs, especially in the tropical Indian Ocean (e.g., Saji et al., 1999; Webster et al., 1999).

In general, the grand 28-model ensemble mean (hereafter the 28-model mean) is more skillful than most individual models within most parts of the tropical Indo-Pacific Ocean (Figure 1), consistent with previous studies showing that combining forecasts from multiple models often yields better skill compared to single-model forecasts (e.g., DelSole et al., 2014; Kirtman et al., 2014). We determined a subset of 10 CMIP5 models (marked by stars in Table S2) whose ensemble mean yielded the maximum 6-month Niño3.4 skill (see Figure S5 in the supporting information for details). The resulting "CMIP5 10-model mean" is more skillful than the 28-model mean and any individual model.

Next, we assess the multimodel ensemble-mean skill with the externally forced trend component included, for the years 1961–2015 (Figure 2). The CMIP5 10-model mean has SST forecast skill comparable to that of the NMME model-analogs, for both the ensemble mean (cf. Figures 2a and 2c) and ensemble distribution (cf. Figures 2e and 2g). The NMME (Figure 2b) and CMIP5 10-model mean (Figure 2d) model-analogs also have similar ensemble-mean skill for 6-month lead precipitation hindcasts during the shorter 1979–2015 period, with the notable exception of the far eastern equatorial Pacific where the CMIP5 10-model mean is clearly worse. This difference is much less for RPSS, however (cf. Figures 2f and 2h). Both are considerably better than the 28-model mean precipitation skill (not shown), especially in the central/western tropical Pacific, an important region for teleconnections to the extratropics (Barsugli & Sardeshmukh, 2002). Since there is little trend in precipitation, the externally forced component has little effect on its skill (see also Figure S1b). Note that no atmospheric information is used to constrain the model-analog hindcasts, so the precipitation skill is only due to knowledge of the initial ocean state.

The skill of the NMME model-analogs also exceeds that of the assimilation-initialized hindcasts for the 1982–2009 period, for SST (a and c of Figures S2–S4) and even more for precipitation (b and d of Figures S2–S4), especially east of 110°W between 5°S and 5°N (see also D18). The SSTA skill reported in Figures S2–S4 is also slightly higher than in D18, which appears to be due to using SSTA from the Hadley Sea Ice and Sea Surface Temperature v1.1 data set for verification rather than the National Oceanic and Atmospheric Administration optimum interpolation SST V2 data set (Reynolds et al., 2002), especially since precipitation skill is almost unaffected (not shown).

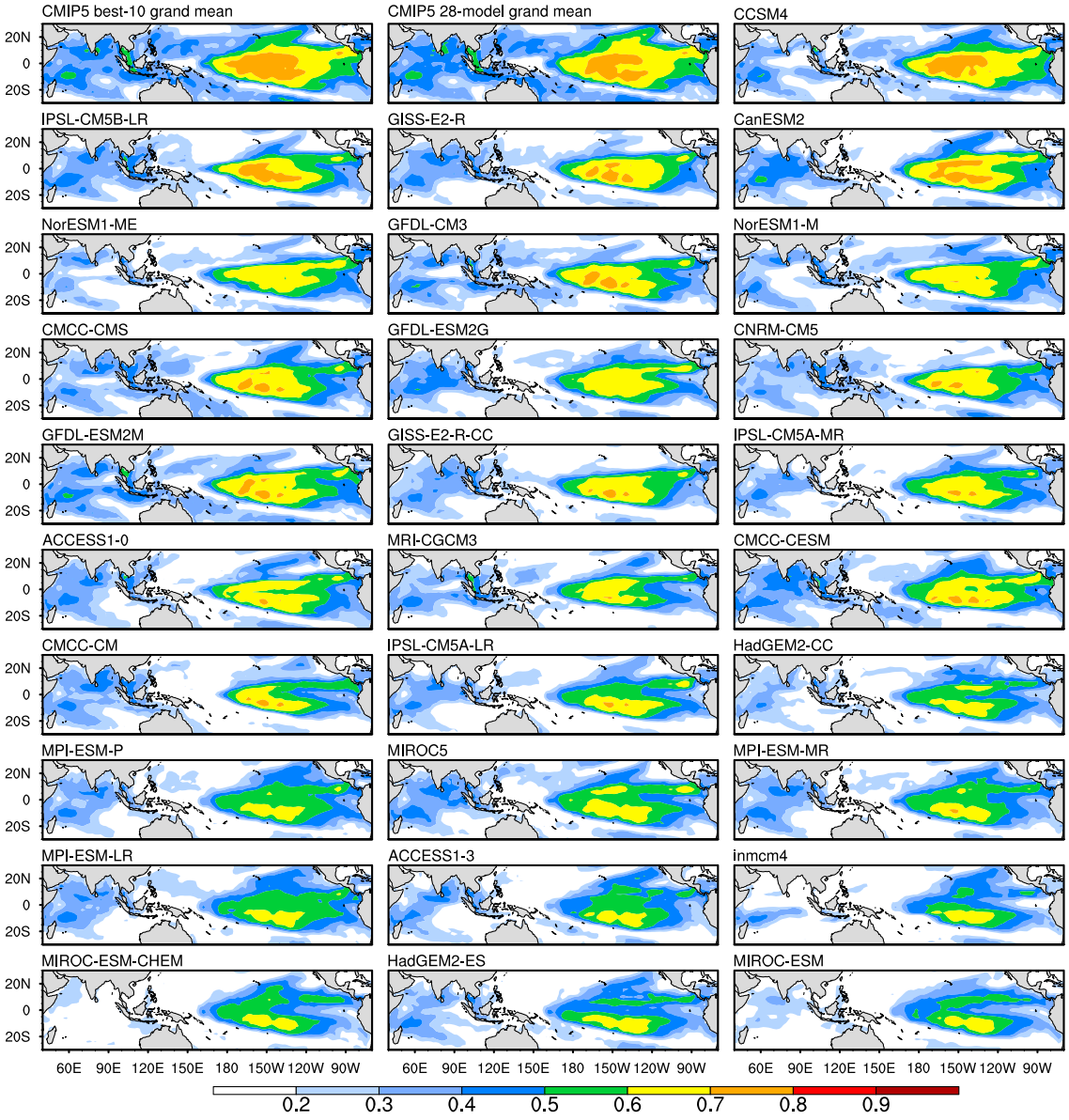

Figure 1. Model-analog hindcast local correlation skill at 6-month lead, relative to observed sea surface temperature anomalies, calculated for both "best 10-" and 28-model grand ensemble means, and individual model ensemble means. Effects of projected radiative forcings are *not* included in this figure.

Figure 3 shows how equatorial forecast skill depends upon lead time for SST (left column) and precipitation (right column). SST skill for the CMIP5 10-model mean and the NMME model-analog mean is generally comparable (cf. Figures 3a and 3c). Including the projected externally forced trend component improves skill in the Indian Ocean and the western Pacific at all forecast leads, more so for the CMIP5 10-model mean than the NMME model-analog mean (not shown). In the central and eastern Pacific, where the trend has little impact on skill, the NMME and CMIP5 model-analogs have remarkably similar skill for leads up to ~10 months, except east of about 100°E. Likewise, the seasonality of Niño3.4 SSTA skill (Figure S6) is nearly identical for CMIP5 and NMME. In particular, both have quantitatively similar spring predictability barriers (e.g., Latif et al., 1998; McPhaden, 2003), with hindcasts initialized during boreal spring having generally reduced skill and hindcasts initialized in all other months showing a sharp decrease in skill with lead time when verifying during spring.

The notable difference in precipitation skill between the NMME and CMIP5 model-analog ensemble means, shown for Month 6 hindcasts, can be seen at all leads when comparing NMME (Figure 3b) and CMIP5 10-

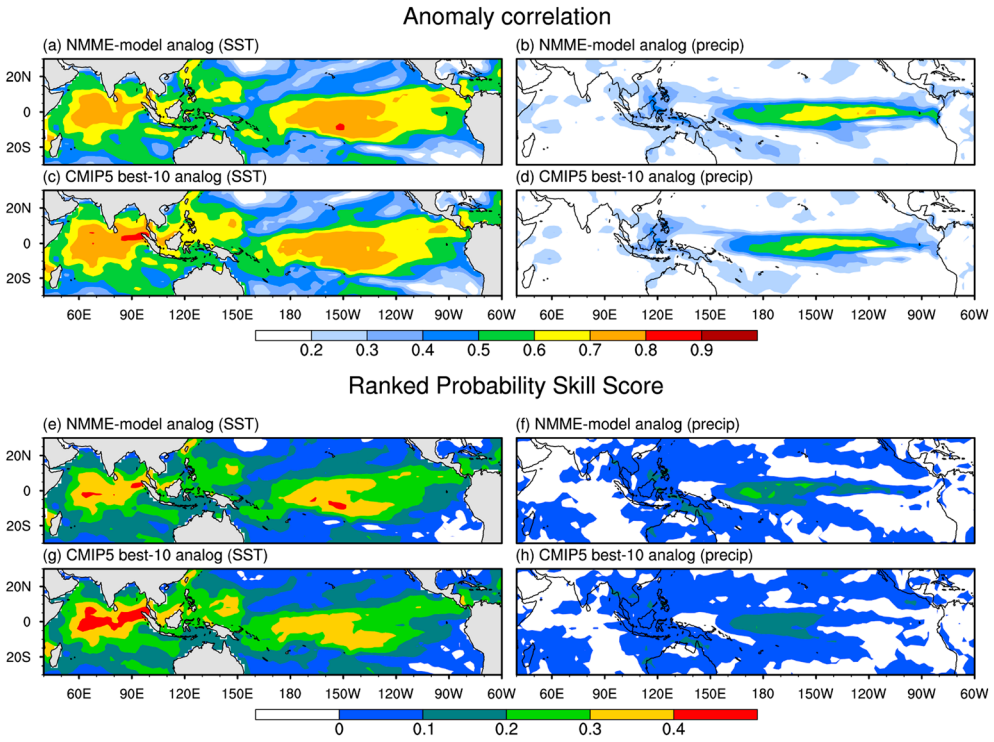

Figure 2. Six-month lead hindcast skill of observed SST (a, c, e, g), 1961–2015, and precipitation (b, d, f, h), 1979–2015, variations, *including* the CMIP5-projected response to external radiative forcing. Anomaly correlation and ranked probability skill score are shown in panels (a–d) and (e–h), respectively. Analog hindcasts are based on the NMME models (a, b, e, and f) and the "best 10" CMIP5 models (c, d, g, and h). SST = sea surface temperature; CMIP5 = fifth phase of the Coupled Model Intercomparison Project; NMME = North American Multi-Model Ensemble.

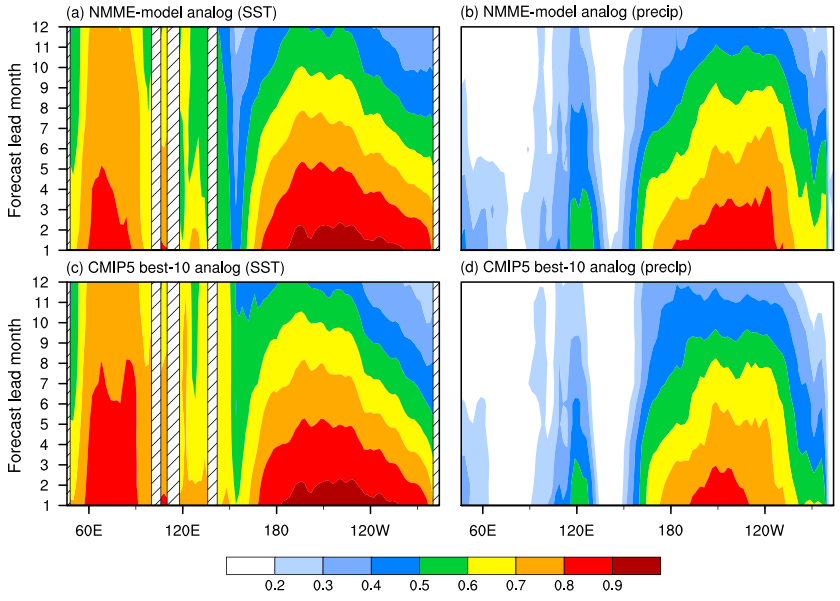

Figure 3. Equatorial model-analog hindcast local anomaly correlation skill for SST (left column), 1961–2015, and precipitation (right column), 1979–2015, averaged between 5°S and 5°N, as a function of longitude (abscissa) and lead time (ordinate). Analog hindcasts are based on (a, b) the NMME models and (c, d) the "best 10" CMIP5 models, respectively. In all model-analog hindcasts, the projected response to external radiative forcing is included. SST = sea surface temperature; NMME = North American Multi-Model Ensemble; CMIP5 = fifth phase of the Coupled Model Intercomparison Project.

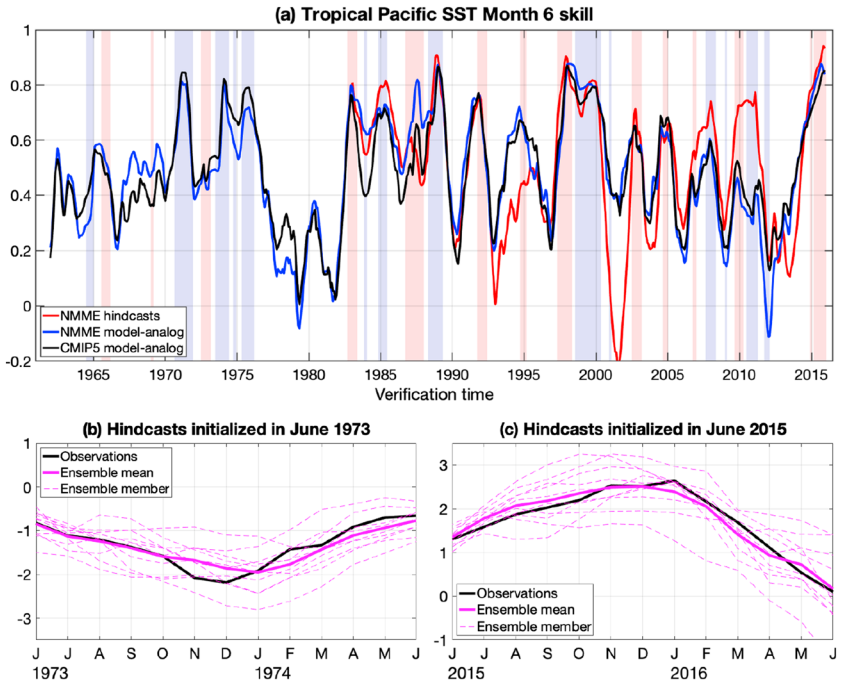

Figure 4. (a) Temporal variations of 6-month lead hindcast skill in the El Niño–Southern Oscillation region, for forecasts initialized monthly during 1961–2015. The measure of skill used is the spatial pattern correlation between the multimodel grand ensemble mean forecast and the corresponding observed monthly SST anomalies within the region bounded by 20°S to 20°N, 170°E to 70°W, smoothed with a 13-month running mean to emphasize annual variations. All anomalies are determined relative to a 1961–2015 monthly climatology. In all model-analog hindcasts, the projected response to external radiative forcing is included. The light red (blue) shaded epochs indicate periods of >1 sigma warm (< −1 sigma cold) Niño3.4 seasonal SST anomalies. (b, c) Evolution of predicted anomalous Niño3.4 index from the CCSM4 model-analog hindcast ensemble, initialized in (b) June 1973 and (c) June 2015. Solid magenta lines show the ensemble means, dashed magenta lines show individual ensemble members, and solid black lines show the observations. These figures represent forecast leads of 1–12 months, where July is Month 1 and June of the following year is Month 12. SST = sea surface temperature; NMME = North American Multi-Model Ensemble; CMIP5 = fifth phase of the Coupled Model Intercomparison Project.

model mean (Figure 3d) equatorial precipitation skill. This difference is evident throughout most of the eastern Pacific, with smaller differences near the dateline. Skill is even less for the 28-model mean (not shown).

The 6-month lead SSTA prediction skill over the past 55 years is shown in Figure 4a, based on the pattern correlation of hindcast and observed anomalies within the ENSO region (section 2.3), for both the CMIP5 10-model mean and the NMME model-analog ensemble mean. Also shown is skill of the operational eight-model NMME grand ensemble mean (Kirtman et al., 2014) since 1982. The evolution of skill is smoothed using a 13-month running mean to emphasize the year-to-year variations. All the skill curves display similar modulation, with a marked correspondence between the modulation of skill and ENSO activity such that higher (lower) skill is coincident with stronger (weaker) ENSO activity, also noted by previous studies (e.g., Kumar et al., 2015; Newman & Sardeshmukh, 2017). The early 2000s period of reduced skill is evident, but now it is also clear that this period is not unique, since skill was even lower during 1976–1982. Conversely, the high skill evident in the late 1980s and 1990s appears to be matched in the early 1970s. Overall, both skill time series are consistent with red noise with an ~1-year decorrelation time scale; that is, on average skill variations are almost independent from year to year. Additionally, there is no significant trend in skill over the entire period.

The predicted and observed evolution of Niño3.4 SST anomalies for two high-skill events is also shown in Figure 4, using model-analog ensembles initialized in either June 1973 (Figure 4b) or June 2015 (Figure 4c) and their subsequent 1- to 12-month lead forecasts. These model-analogs are taken from the CMIP5 CCSM4 preindustrial simulation, which has the highest 6-month correlation skill in Nino3.4 SST

(Figure S5), but results are similar for other skillful models. The observed time evolution of both ENSO events is clearly captured by the model-analog hindcasts. Note that the initial spread of each ensemble is not particularly small, partly because the model-analogs primarily capture the dominant spatial patterns (e.g., Figures S7 and S8). However, the spread also does not increase substantially, and the observed evolution generally stays well within the envelope of the ensemble members. See D18 for more discussion of this point.

5. Summary and Discussion

We have shown that, with the inclusion of the projected externally forced climate response from the CMIP5 historical ensemble mean, the NMME model-analog ensemble-mean and the NMME assimilation-initialized ensemble-mean hindcasts have almost identical skill for both SSTA and precipitation anomalies throughout the tropical Indo-Pacific. This suggests that we have likewise determined the skill of the CMIP5 models, as well as of the CMIP5 multimodel ensemble mean, without running any of them as traditional assimilation-initialized forecast models. Most of the CMIP5 models generate skillful forecasts, although the skill varies considerably from model to model. Likewise, the skill of a smaller subset of the models (10, determined by the single metric used here) is higher than the skill of the full 28-model ensemble. The 10-model mean SST hindcast skill is quite close to the NMME model-analog grand ensemble mean skill in the equatorial Pacific. This is surprising, since the CMIP5 models were finalized about 8 years ago, while the NMME model development since then aimed to better reproduce observed aspects of ENSO variability for the purpose of making operational forecasts. The larger gap in precipitation skill in the eastern tropical Pacific, however, may have implications for how the CMIP5 models are capturing tropical air-sea coupling.

The model-analog technique essentially measures how well each CMIP5 model reproduced the observed evolution of ENSO, including its precipitation response. This work therefore has great practical significance given that the CMIP5 project was designed to support the Intergovernmental Panel on Climate Change Fifth Assessment Report (Taylor et al., 2012); likewise, the model-analog technique can be used to assess upcoming CMIP6 models and develop seasonal forecasts from them. It may also be worth searching for model-analogs using large ensembles of externally forced model simulations (e.g., Kay et al., 2015) rather than using control simulations as was done here, both so that the externally forced component is more consistently included without the need to determine it separately and to determine how external forcing may impact ENSO predictability.

The power of the model-analog technique is evident in the size of our new tropical Indo-Pacific SST and precipitation hindcast database, which contains 10-member ensembles of 1- to 12-month lead forecasts, from each of 4 NMME and 28 CMIP5 models, initialized each month during 1961–2015, using only SSTA and SSHA as initial conditions. By doubling the length of current hindcast databases (e.g., the NMME) as well as substantially increasing the number and variety of the forecast models, we have obtained better estimates of overall ENSO forecast skill and have found that the lull in ENSO forecast skill in the early 2000s was not a unique event, with an even more notable reduction occurring in the late 1970s when ENSO activity was also at a minimum. More notably, the model-analog technique allows users to make seasonal ENSO forecasts without the development of a sophisticated forecast system, by taking advantage of freely available large model simulation databases. That is, anyone can be a climate forecaster.

Acknowledgments
This work has been funded by the NOAA Climate Program Office through the Earth System Modeling Program. All model-analog hindcasts discussed in this paper are available at ftp://ftp2.esrl.noaa.gov/Projects/ESM/hding/GRL2019/Hindcasts/. We acknowledge the CMIP5 climate modeling groups for producing and making available their model output. The CMIP5 data output is available at https://esgf-node.llnl.gov/search/cmip5/. Data from the four NMME control runs are available from https://data1.gfdl.noaa.gov and ftp://ftp2.esrl.noaa.gov/Projects/ESM/hding/GRL2019/NMME_ControlRuns. We also acknowledge the agencies that support the NMME-Phase II system, and we thank the participating climate modeling groups for producing and making their model output available. The NMME hindcasts are available at https://iridl.ldeo.columbia.edu.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Alexander, M., 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

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

Barnston, A. G., Tippett, M. K., van den Dool, H. M., & Unger, D. A. (2015). Toward an improved multimodel ENSO prediction. *Journal of Applied Meteorology and Climatology*, *54*(7), 1579–1595. https://doi.org/10.1175/JAMC-D-14-0188.1

Barsugli, J. J., & Sardeshmukh, P. D. (2002). Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *Journal of Climate*, *15*(23), 3427–3442. https://doi.org/10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., & Vialard, J. (2014). ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics*, *42*(7–8), 1999–2018. https://doi.org/10.1007/s00382-013-1783-z

Chen, C., Cane, M. A., Wittenberg, A. T., & Chen, D. (2017). ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, *30*(2), 775–801. https://doi.org/10.1175/JCLI-D-15-0901.1

Chen, L.-C. G., & Van Den Dool, H. (2017). Combination of multimodel probabilistic forecasts using an optimal weighting system. *Weather and Forecasting*, *32*(5), 1967–1987. https://doi.org/10.1175/WAF-D-17-0074.1

Dai, A., Fyfe, J. C., Xie, S.-P., & Dai, X. (2015). Decadal modulation of global surface temperature by internal climate variability. *Nature Climate Change*, *5*(6), 555–559. https://doi.org/10.1038/nclimate2605

DelSole, T., Nattala, J., & Tippett, M. K. (2014). Skill improvement from increased ensemble size and model diversity. *Geophysical Research Letters*, *41*, 7331–7342. https://doi.org/10.1002/2014GL060133

Deser, C., Phillips, A. S., & Alexander, M. A. (2010). Twentieth century tropical sea surface temperature trends revisited. *Geophysical Research Letters*, *37*, L10701. https://doi.org/10.1029/2010GL043321

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

Flügel, M., Chang, P., & Penland, C. (2004). The role of stochastic forcing in modulating ENSO predictability. *Journal of Climate*, *17*(16), 3125–3140. https://doi.org/10.1175/1520 -0442(2004)017<3125:TROSFI>2.0.CO;2

Huang, B., Shin, C.-S., Shukla, J., Marx, L., Balmaseda, M. A., Halder, S., et al. (2017). Reforecasting the ENSO events in the past 57 years (1958–2014). *Journal of Climate*, *30*(19), 7669–7693. https://doi.org/10.1175/JCLI-D-16-0642.1

Karamperidou, C., Cane, M. A., Lall, U., & Wittenberg, A. T. (2014). Intrinsic modulation of ENSO predictability viewed through a local Lyapunov lens. *Climate Dynamics*, *42*(1–2), 253–270. https://doi.org/10.1007/s00382-013-1759-z

Kay, J., 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

Kirtman, B. P., & Min, D. (2009). Multimodel ensemble ENSO prediction with CCSM and CFS. *Monthly Weather Review*, *137*(9), 2908–2930. https://doi.org/10.1175/2009MWR2672.1

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L. III, 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

Knutson, T. R., Zeng, F., & Wittenberg, A. T. (2014). Multimodel assessment of extreme annual-mean warm anomalies during 2013 over regions of Australia and the western tropical Pacific. *Bulletin of the American Meteorological Society*, *95*(9), S1–S104. https://doi.org/10.1175/1520-0477-95.9.S1.1

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), 14,375–14,393. https://doi.org/10.1029/97JC03413

Lee, J.-W., Yeh, S.-W., & Jo, H.-S. (2016). Weather noise leading to El Niño diversity in an ocean general circulation model. *Climate Dynamics*, 1–13. https://doi.org/10.1007/s00382-016-3438-3

Lee, T., & McPhaden, M. J. (2010). Increasing intensity of El Niño in the central-equatorial Pacific. *Geophysical Research Letters*, *37*, L14603. https://doi.org/10.1029/2010GL044007

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

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

Newman, M., Shin, S.-I., & Alexander, M. A. (2011). Natural variation in ENSO flavors. *Geophysical Research Letters*, *38*, L14705. https://doi.org/10.1029/2011GL047658

Newman, M., Wittenberg, A. T., Cheng, L., Compo, G. P., & Smith, C. A. (2018). The extreme 2015/16 El Niño, in the context of historical climate variability and change. Section 4 of: "Explaining extreme events of 2016 from a climate perspective". *Bulletin of the American Meteorological Society*, *99*(1), S16–S20. https://doi.org/10.1175/BAMS-D-17-0116.1

Ogata, T., Xie, S.-P., Wittenberg, A., & Sun, D.-Z. (2013). Interdecadal amplitude modulation of El Niño–Southern Oscillation and its impact on tropical Pacific decadal variability. *Journal of Climate*, *26*(18), 7280–7297. https://doi.org/10.1175/JCLI-D-12-00415.1

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

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, *15*(13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., et al. (2006). The NCEP climate forecast system. *Journal of Climate*, *19*(15), 3483–3517. https://doi.org/10.1175/JCLI3812.1

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

Solomon, A., & Newman, M. (2012). Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nature Climate Change*, *2*(9), 691–699. https://doi.org/10.1038/nclimate1591

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

Stockdale, T. N. (1997). Coupled ocean-atmosphere forecasts in the presence of climate drift. *Monthly Weather Review*, *125*(5), 809–818. https://doi.org/10.1175/1520-0493(1997)125<0809:COAFIT>2.0.CO;2

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Vecchi, G., Wittenberg, A., & Rosati, A. (2006). Reassessing the role of stochastic forcing in the 1997–1998 El Niño. *Geophysical Research Letters*, *33*, L01706. https://doi.org/10.1029/2005GL024738

Watanabe, M., Kug, J.-S., Jin, F.-F., Collins, M., Ohba, M., & Wittenberg, A. T. (2012). Uncertainty in the ENSO amplitude change from the past to the future. *Geophysical Research Letters*, *39*, L20703. https://doi.org/10.1029/2012GL053305

Watanabe, M., & Wittenberg, A. T. (2012). A method for disentangling El Niño-mean state interaction. *Geophysical Research Letters*, *39*, L14702. https://doi.org/10.1029/2012GL052013

Webster, P. J., Moore, A. M., Loschnigg, J. P., & Leben, R. R. (1999). Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997–98. *Nature*, *401*(6751), 356–360. https://doi.org/10.1038/43848

Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly Weather Review*, *135*(1), 118–124. https://doi.org/10.1175/MWR3280.1

Weisheimer, A., Doblas-Reyes, F., Palmer, T., Alessandri, A., Arribas, A., Deqúe, M., et al. (2009). ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters*, *36*, L21711. https://doi.org/10.1029/2009GL040896

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*, L12702. https://doi.org/10.1029/2009GL038710

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. https://doi.org/10.1175/JCLI-D-13-00577.1

Yeh, S.-W., Kug, J.-S., Dewitte, B., Kwon, M.-H., Kirtman, B. P., & Jin, F. F. (2009). El Niño in a changing climate. *Nature*, *461*(7263), 511–514. https://doi.org/10.1038/nature08316

Zhao, M., Hendon, H. H., Alves, O., Liu, G., & Wang, G. (2016). Weakened eastern Pacific El Niño predictability in the early twenty-first century. *Journal of Climate*, *29*(18), 6805–6822. https://doi.org/10.1175/JCLI-D-15-0876.1