Skillful Climate Forecasts of the Tropical Indo-Pacific Ocean Using Model-Analogs⊘

Hui Ding and Matthew Newman

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Earth System Research Laboratory, Boulder, Colorado

Michael A. Alexander

NOAA/Earth System Research Laboratory, Boulder, Colorado

Andrew T. Wittenberg

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

ABSTRACT

Seasonal forecasts made by coupled atmosphere–ocean general circulation models (CGCMs) undergo strong climate drift and initialization shock, driving the model state away from its long-term attractor. Here we explore initializing directly on a model's own attractor, using an analog approach in which model states close to the observed initial state are drawn from a ''library'' obtained from prior uninitialized CGCM simulations. The subsequent evolution of those ''model-analogs'' yields a forecast ensemble, without additional model integration. This technique is applied to four of the eight CGCMs comprising the North American Multimodel Ensemble (NMME) by selecting from prior long control runs those model states whose monthly tropical Indo-Pacific SST and SSH anomalies best resemble the observations at initialization time. Hindcasts are then made for leads of 1–12 months during 1982–2015. Deterministic and probabilistic skill measures of these model-analog hindcast ensembles are comparable to those of the initialized NMME hindcast ensembles, for both the individual models and the multimodel ensemble. In the eastern equatorial Pacific, model-analog hindcast skill exceeds that of the NMME. Despite initializing with a relatively large ensemble spread, model-analogs also reproduce each CGCM's perfect-model skill, consistent with a coarse-grained view of tropical Indo-Pacific predictability. This study suggests that with little additional effort, sufficiently realistic and long CGCM simulations provide the basis for skillful seasonal forecasts of tropical Indo-Pacific SST anomalies, even without sophisticated data assimilation or additional ensemble forecast integrations. The model-analog method could provide a baseline for forecast skill when developing future models and forecast systems.

1. Introduction

Seasonal forecast skill has significantly improved over the past three decades (Barnston et al. 2012; Barnston and Tippett 2017). Operational prediction centers worldwide use coupled atmosphere–ocean general circulation models (CGCMs) to conduct routine initialized

seasonal forecasts with actionable skill of 6–12 months in advance (Doblas-Reyes et al. 2013), much of which is related to predictions of tropical sea surface temperature (SST) and especially of El Niño–Southern Oscillation (ENSO) (Jin et al. 2008; Barnston et al. 2012; Kirtman et al. 2014). Both model and initialization improvements have driven advances in forecast skill, with considerable effort devoted to developing data assimilation schemes (Ji and Leetmaa 1997; Stockdale et al. 1998, 2011; Keenlyside et al. 2005; Saha et al. 2014).

While coupled models have improved, they remain imperfect. Models continue to suffer from both unconditional biases in their mean states, such as the

double ITCZ and cold tongue bias (e.g., Li and Xie 2014), and conditional biases in their variability, such as a westward shift of ENSO (e.g., Joseph and Nigam 2006). CGCM seasonal forecast skill is significantly impacted by these model errors (e.g., Barnston et al. 2015; Newman and Sardeshmukh 2017). Initialization shock, which can arise when forecast error rapidly develops from an initial imbalance between the analyzed initial state and all possible model states, degrades forecast skill in the presence of nonlinear air–sea interactions (Mulholland et al. 2015, 2016). At longer forecast leads, this problem may be worsened by the difficulty of spinning up the deeper oceanic component of the initialization.

When model-state space trajectories do not coincide with trajectories in nature, we could try to find and initialize those model trajectories that are closest (in some sense) to nature. This "shadowing" approach has been applied to relatively simple models (e.g., Judd et al. 2004), although the real-world problem may be considerably more complex (Smith 2001). Anomaly initialization (e.g., Hazeleger et al. 2013; Volpi et al. 2017) and nudging (e.g., DelSole et al. 2008; Meehl et al. 2014; Carrassi et al. 2016) both aim to remove effects of the initial model drift, but are only partly successful and do not directly confront conditional model error. Alternatively, model errors may be addressed through postprocessing hindcast ensembles, although their length and size are limited by available computational resources.

Here, we take the view that for forecast lead times where the model has drifted toward its own climatology and forecast states follow trajectories that lie within the model attractor, model initialization should be directly in the model phase space and not in the phase space of nature. To this end, we investigate initializing climate forecasts with states taken directly from long control simulations of operational CGCMs. These states are chosen as analogs to the observed initial state; then, their evolution within the control simulation immediately provides the forecast ensemble, and no additional model integration is necessary. In this paper we further simplify this approach by choosing these "model-analogs" solely from the current position of the state in phase space, without considering the phase-space trajectory, and determining them not from the entire climate state, but rather from monthly anomalies of SST and sea surface height (SSH). Nevertheless, applying this technique to tropical Indo-Pacific Ocean forecasts yields skill, such as shown in Fig. 1, that is not only surprisingly competitive with the traditional approach of executing an initialized forecast ensemble from the corresponding CGCMs, but actually appears to exceed it in the key ENSO region within the equatorial eastern Pacific.

The use of analogs in weather and climate forecasting goes back many decades (Lorenz 1969; Barnett and Preisendorfer 1978; Gutzler and Shukla 1984; Livezey and Barnston 1988; Van den Dool 1989). The underlying assumption is that a pair of states that are initially sufficiently similar to each other will follow similar trajectories for some subsequent finite period. Lorenz (1969) first noted that the increasing difference between two such analog trajectories represents error growth and therefore a measure of potential predictability. He suggested this difference might be empirically determined, but found that for daily weather, no pair of analogs taken from the observational record was close enough to be useful. In fact, Van den Dool (1994) suggested that global daily weather anomalies may have too many degrees of freedom for useful analogs to be found.

However, analogs may be more useful for problems with fewer effective degrees of freedom (Van den Dool 1994). Observational analogs have been applied with some success to short-term climate forecasting (Barnett and Preisendorfer 1978; Nicholls 1980) including U.S. surface temperature seasonal anomalies (Livezey and Barnston 1988; Barnston and Livezey 1989), climate downscaling (Kruizinga and Murphy 1983; Zorita and von Storch 1999; Wetterhall et al. 2005), and calibration of ensemble weather forecasts (Van den Dool et al. 2003; Fraedrich et al. 2003). However, these applications are still limited by the relatively short observational record. The recent easy availability of long climate model runs, providing access to larger data libraries than the observational record, has made analog techniques particularly attractive for applications such as paleoclimate reconstruction (Overpeck et al. 1992; Graham et al. 2007; Goodwin et al. 2014), estimating perfect-model climate predictability (Branstator et al. 2012), and separation of internal and forced trend components within climate runs (Deser et al. 2016; Lehner et al. 2017).

This study will investigate the use of a model-based analog technique for seasonal forecasts of tropical Indo-Pacific SST variability, which many studies have suggested can be represented by relatively few degrees of freedom. For example, most central and eastern Pacific ENSO events can be described through a linear combination of two dominant spatial patterns that together explain more than half of the variance in the tropical Indo-Pacific region (e.g., Takahashi et al. 2011). Also, the evolution of ENSO is well approximated by a low-dimensional, Markovian dynamical system with relatively few spatial degrees of freedom (Penland and Sardeshmukh 1995). Additionally, Wittenberg (2009) found that for the GFDL CM2.1 control simulation, 500 years of data was sufficient to sample the range of ENSO
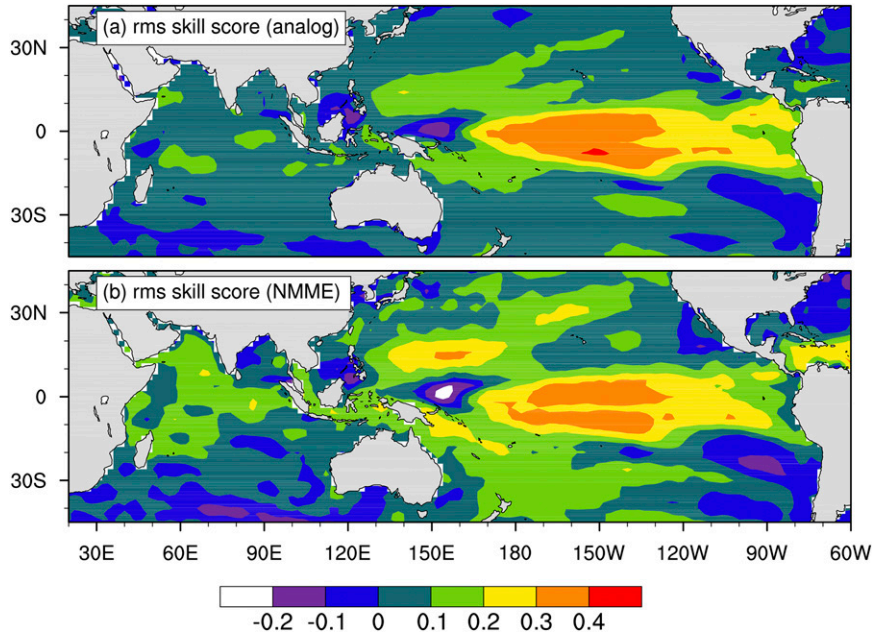
FIG. 1. (a) Model-analog and (b) NMME hindcast skills of observed SST variations at 6-month lead. Only RMS skill score is shown. The grand mean of 60 analogs is shown in (a), which consist of 15 best initial analogs from each of the four control runs (Table 1). The grand mean of the four NMME experiments is shown in (b) (Table 1).

events. We might thus expect that our model-analog approach to ENSO could be more tractable than other problems in terms of the amount of data needed.

The paper is organized as follows. Section 2 describes the details of the model-analog technique employed in this study, along with the CGCM simulations used. In section 3, model-analogs are first applied to forecast each model itself. The sensitivity of the technique to the size of historical data as well as ensemble size is explored, and it is shown that, in agreement with Lorenz (1969), model-analogs can be used to estimate a CGCM's perfect model seasonal tropical predictability. In section 4, the model-analog technique is used to make hindcasts for the 1982–2009 period, their skill is shown comparable to hindcast skill from the operational CGCM ensemble, and their sources of skill are evaluated. A summary and discussion of some implications of the model-analog technique are presented in section 5.

2. Models and method

a. Model-analog technique

We choose analogs at each time t by minimizing a metric of the distance between the target state $\mathbf{x}(t)$ and each library state $\mathbf{y}(t')$. Here, the target state is defined as the state at the initialization time, and the library consists of all preexisting states obtained from a CGCM

control simulation. In this study, we define a simple distance metric $d(t, t')$ from the root-mean-square (RMS) difference between a subset of variables within the full state vectors, or

$$d^2(t,t') = \sum_{i=1}^{I} \sum_{j=1}^{J} \left[\frac{x_j^i(t)}{\sigma_X^i} - \frac{y_j^i(t')}{\sigma_Y^i} \right]^2, \qquad (1)$$

where i represents a variable, with I as the total variables, and j represents a spatial degree of freedom (e.g., gridpoint index or principal component) within the training region, with J as the total spatial degrees of freedom. In (1), each variable is normalized by its respective domain-averaged standard deviation, σ_X^i and σ_Y^i, to have equal weight within this metric. Monthly mean data are used in (1), and to include seasonality, the library states are restricted to have the same calendar month as the target state. Consequently, only one analog is available from each year of the model simulation, so that the library length N is equivalent to the number of years in the training period.

Distances are 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 $\{\mathbf{y}(t_1'), \mathbf{y}(t_2'), \ldots, \mathbf{y}(t_k'), \ldots, \mathbf{y}(t_K')\}$, with k as the analog index and t_k' as the time of this analog in the library. The subsequent model evolution of this ensemble $\{\mathbf{y}(t_1' + \tau),$

$\mathbf{y}(t'_2 + \tau), \ldots, \mathbf{y}(t'_k + \tau), \ldots, \mathbf{y}(t'_K + \tau)\}$ is the model-analog forecast ensemble for $\mathbf{x}(t + \tau)$ at lead time τ months. The ensemble mean $\mathbf{z}(t) \equiv \sum_{k=1}^{K} \mathbf{y}(t'_k)/K$ represents the initial model-analog ensemble-mean reconstruction of the target $\mathbf{x}(t)$, and $\mathbf{z}(t+\tau) \equiv \sum_{k=1}^{K} \mathbf{y}(t'_k + \tau)/K$ is the model-analog ensemble-mean forecast at lead time τ months. In this paper, we do not weight the model-analog ensemble members. Note that a linear weighting such as used in the constructed analog method (Van den Dool 1994) can be shown to reduce to multivariate linear regression (Tippett and DelSole 2013). We also tested a quadratic weighting [i.e., inversely weighting each member by $d^2(t, t')$] but found no notable improvement.

To define analogs, we set $I = 2$ and constructed target and library states from SSH anomalies (SSHAs; $i = 1$) and SST anomalies (SSTAs; $i = 2$). SST is one of the most important quantities for seasonal forecasts (e.g., Latif et al. 1998; Barnston et al. 2012), and fluctuations in central and eastern tropical Pacific SST can exert a significant impact on atmospheric circulation and remote ocean regions via the atmospheric bridge (Alexander et al. 2002). Ocean heat content variations are also important for ENSO dynamics (Zebiak and Cane 1987; Neelin et al. 1998) and have been shown to be important for ENSO prediction (Kleeman et al. 1995; Ji and Leetmaa 1997; Rosati et al. 1997; Newman et al. 2011). Here, SSH variations are used to represent ocean heat content variations, since in the tropical oceans, SSH is intimately linked to the depth of the thermocline (Cane 1984). Furthermore, SSH is also an indicator of equatorial wave dynamics in the oceans (Gill and Clarke 1974), which affects the development and termination of ENSO events (Suarez and Schopf 1988). Using SSTAs alone substantially degraded the forecast skill of the model-analogs (not shown).

b. Model and observational datasets

The training (library) dataset consists of monthly mean data taken from CGCM control simulations of varying lengths, conducted using four climate models: CM2.1, CM2.5 FLOR, CCSM4, and CESM1, listed in Table 1. In each control simulation, only the portion of integration near equilibrium (i.e., spunup) is retained. Atmospheric greenhouse gas concentrations are fixed throughout each simulation, with the specified year also indicated in Table 1. The longest run, the 4000-yr preindustrial simulation of the GFDL CM2.1, has previously been analyzed to investigate record length needed to adequately sample ENSO events (Wittenberg 2009; Atwood et al. 2017), the predictability of decadal modulations of ENSO events (Karamperidou et al. 2014; Wittenberg et al. 2014) and their impact on

decadal variability (Ogata et al. 2013), and ENSO diversity (Kug et al. 2010; Capotondi et al. 2015; Chen et al. 2017).

Versions of all four models are also currently used by NCEP to make operational seasonal forecasts as part of the second phase of the North American Multimodel Ensemble (NMME) project, and were also used to generate ensembles of retrospective forecasts ("hindcasts") initialized from the beginning of each month from 1982 to 2009 and integrated forward for 12 months (Kirtman et al. 2014; see Table 1). To calculate anomalies, all these hindcasts were bias corrected: the mean hindcast drift as a function of lead and calendar month was removed separately for each model, as is common practice with CGCM seasonal forecasts (Stockdale 1997; Saha et al. 2006; Kirtman and Min 2009). For convenience, these hindcasts are denoted as NMME hindcasts, which we will compare with our model-analog hindcasts.

To determine initial observed states, we used SSTs from the monthly mean NOAA Optimum Interpolation SST, version 2, dataset (OISST; Reynolds et al. 2002) and SSHs from the ECMWF Ocean Reanalysis System 4 dataset (Balmaseda et al. 2013), for the years 1982–2015. SSTA and SSHA were determined by removing the monthly mean 1982–2009 climatology.

All model and observed data were interpolated onto a common 2° longitude by 2° latitude grid prior to our analysis. Analogs were determined within the domain bounded by 30°S–30°N, 30°E–80°W, covering the entire tropical Indo-Pacific Ocean.

c. Skill metrics

To measure how well the initial analog ensemble mean reproduces the target state, we use $d_K(t)$, the distance between the target state $\mathbf{x}(t)$ and the model-analog ensemble mean $\mathbf{z}(t)$, which from (1) is

$$d_K^2(t) = \sum_{i=1}^{2} \sum_{j=1}^{J} \left[\frac{x_j^i(t)}{\sigma_X^i} - \frac{z_j^i(t)}{\sigma_Y^i} \right]^2, \qquad (2)$$

where $z_j^i(t)$ indicates the ith component of $\mathbf{z}(t)$ at the jth grid point. Note that $d_K(t)$ will generally be different than the mean of the distances of the analog members themselves.

We use two deterministic skill measures: anomaly correlation (AC) and RMS error skill score (RMSSS; Barnston et al. 2015) defined as $1 - \varepsilon$, where $\varepsilon = \epsilon/\sigma$ is the standardized RMS error, ϵ is the RMS error of the ensemble-mean forecast, and σ is the observed climatological standard deviation. The overall skill within the tropical Indo-Pacific domain was assessed using the

TABLE 1. Model control run data used in this paper. The number of model vertical levels are in parentheses and prefixed with the letter L with the model resolution.

Model	Expanded model name	Year of fixed radiative gas concentrations	Length of control run (yr)	Resolution (lon × lat)		Corresponding NMME model	No. of NMME hindcast ensemble members
				Atmosphere	Ocean		
CM2.1	Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model, version 2.1 (CM2.1)	1860	4000	2° × 2.5° (L24)	1° × 1°, with a meridionally refined resolution of 0.33° near the equator (L50)	GFDL-CM2.1-aer04	10
CM2.5 FLOR	Forecast-Oriented Low Ocean Resolution (FLOR) version developed from the GFDL Climate Model, version 2.5 (CM2.5)	1990	700	50 km (L32)	1° × 1°, with a meridionally refined resolution of 0.33° near the equator (L50)	GFDL-CM2.5 FLOR	24
CCSM4	Community Climate System Model, version 4	1850	1100	1.25° × 0.9° (L26)	1.13° × 0.27° near the equator (L60)	COLA-Rosenstiel School of Marine and Atmospheric Science (RSMAS)-CCSM4	10
CESM1	Community Earth System Model, version 1	2000	700	1.25° × 0.9° (L30)	1.13° × 0.27° near the equator (L60)	NCAR-CESM1	10

standardized RMS error and the pattern correlation between the forecast and verification (OISST or perfect model) SSTA fields in the region 30°S–30°N, 30°E–80°W. Probabilistic skill is assessed using ranked probability skill score (RPSS; Weigel et al. 2007), reliability, and sharpness (Murphy 1973), applied to the observed and forecast tercile distributions.

3. Assessing perfect-model predictability with model-analogs

In this section, we evaluate the model-analog technique in a perfect model context. That is, for each CGCM, we construct model-analogs from a portion of its control run (''data library'') and use them to make forecasts within the withheld (''verification'') dataset. For each control simulation, data from the first 200 years are used for verification, with the remaining data used as the data library. Thus, the numbers of years N in the training periods for the CM2.1, CM2.5 FLOR, CCSM4, and CESM1 control simulations are 3800, 500, 900, and 500, respectively. For all four simulations, we set the analog ensemble size at $K = 10$. Parametric sensitivity is discussed in section 3b for choices of N and K, and in section 3c for the length of the verification period.

a. Evaluation of model-analogs from four control simulations

Figure 2 shows how well the model-analog ensemble means represent the target anomalies for each control simulation. In general, the AC between the target-state and the model-analog ensemble mean (shading) is above 0.7 throughout the training region (red box in Fig. 2a). For each simulation, the model-analog ensemble mean matches the target anomaly best along or near the equator, both for SSTA and for SSHA (not shown; values are generally similar with some differences in details). However, even in the eastern equatorial Pacific, the RMSSS (shown by white contours) does not exceed about 0.75; that is, on average there is at best about an initial $\pm 0.25\sigma$ difference between the model-analog ensemble mean and the target anomalies.

The model-analogs better capture larger-scale, more slowly decorrelating variations in the target anomalies than smaller-scale, more rapidly decorrelating variations. This is demonstrated in Fig. 3, which shows the monthly initial model-analog reconstruction error projected on the first 40 SSTA EOFs determined from the verification datasets. The model-analog ensemble means have a better representation of the target SSTA within the space of the leading principal components (PCs), which correspond to larger spatial scales and have the longest decorrelation time scales (not shown).

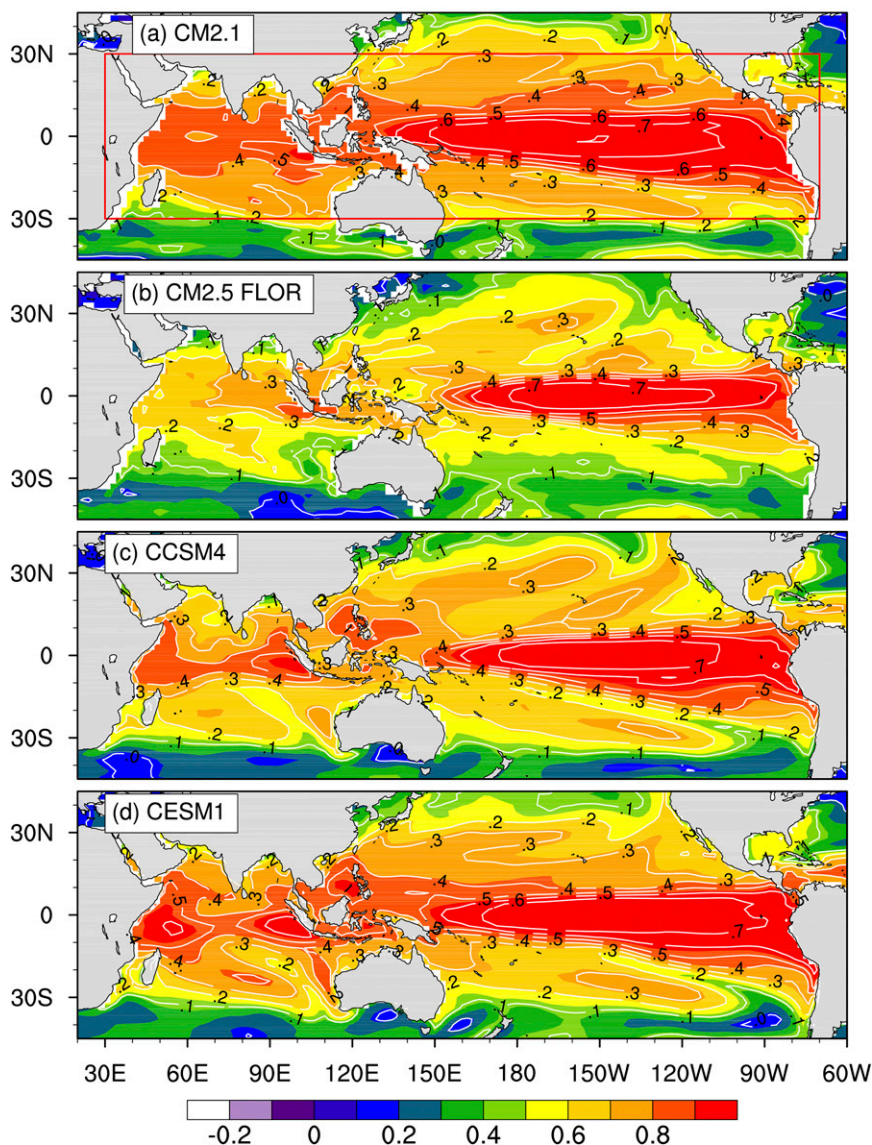## Skill of analogs at reconstructing out-of-sample model states

FIG. 2. Perfect-model skill of ensemble-mean model-analog reconstruction of SSTA at zero lag for the verification period in (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, and (d) CESM1. Shading denotes local anomaly correlation and contours RMSSS. For each model, the first 200 years of its control run was used as the verification period, with the remaining data used for the data library. Contour and shading interval is 0.1. Red box in (a) indicates the region used for evaluating the analog distances from the initial target state.

Most of the initial RMS error outside of the equatorial Pacific in Fig. 2 is due to the relatively larger error in the higher-order PCs, which dominate the variability there.

Given the nontrivial error in the analog representation of the target anomalies, it is notable that the 6-month-lead model-analog ensemble-mean forecasts have considerable skill over large parts of the tropical Indo-Pacific Ocean (Fig. 4); for some of the models, much of this skill is retained up to leads of 24 months

(Fig. 5; see also Fig. S1 in the supplemental material). For all four models, AC skill at 6-month lead exceeds 0.8 in the equatorial Pacific and 0.7 in large parts of the tropical Indian Ocean, and in the CESM1 even reaches 0.8 in portions of the Indian Ocean. Interannual variability in the tropical Indian Ocean is strongly related to ENSO (e.g., Nicholson 1997; Klein et al. 1999; Venzke et al. 2000; Alexander et al. 2002), although other sources exist (e.g., Saji et al. 1999; Yamagata et al. 2004).
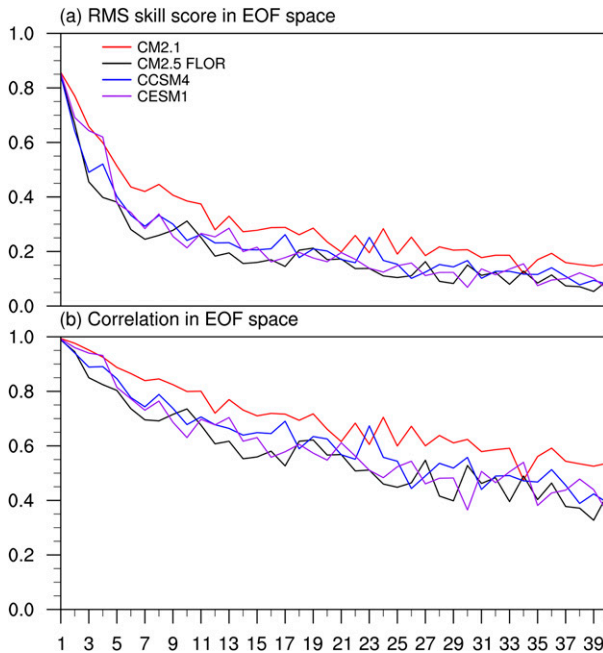
FIG. 3. Skill of ensemble-mean model-analog reconstruction of SSTA for the verification period, measured by (a) RMSSS and (b) correlation, as a function of EOF index. EOFs are calculated using monthly SSTAs from the verification period separately for each control run.

The model-analog technique also captures some notable skill differences between the four models. For example, in CM2.1 (Fig. 4a), skill is greatest in the western–central equatorial Pacific, while in CESM1 it peaks in the central to eastern equatorial Pacific (Fig. 4d), consistent with the greater westward displacement of ENSO variability in the CM2.1 compared to the CESM1 (not shown).

Note that, for all four models, the patterns of RMSSS (contours) and AC (shading) are very similar. This does not appear to be coincidental. In fact, it can be shown that for the mean of an infinite ensemble of any perfect model (i.e., a model whose forecasts have no conditional or unconditional bias), AC skill ρ and standardized RMS forecast error ε, standardized by the standard deviation of the verification data, are directly related by

$$\rho^2 + \varepsilon^2 = 1 \qquad (3)$$

(von Storch and Zwiers 2002; Newman and Sardeshmukh 2017). This relationship is largely satisfied for the model-analog skill shown in Fig. 4, apart from some small (~1%–3%) differences in some regions (Fig. S2 in the supplemental material). Additionally, $\rho^2 + \varepsilon^2$ computed from standardized RMS error and pattern correlation values determined over the entire tropical Indo-Pacific training domain largely satisfies (3) for all forecast leads

(Fig. S3). That is, despite the relatively large spread of its initial ensemble, the model-analog technique applied in a perfect model sense provides an excellent estimate of the potential predictability of each model within the tropical Indo-Pacific region, consistent with Lorenz (1969).

Potential predictability of equatorial SSTA AC skill (Fig. 5) is greater than persistence (not shown) for all forecast lead times ranging from 0 to 24 months. Following previous studies (e.g., Collins et al. 2002), we defined a maximum predictable lead time (MPLT) as the forecast lead time at which the AC reaches 0.6. Figure 5 shows that MPLTs for equatorial SSTAs have a strong dependence upon both model and longitude. For instance, in CM2.1, MPLT ranges from 20 months in the western equatorial Pacific to only about 8 months in the eastern part of the basin (Fig. 5a). The other three models have MPLT in the central equatorial Pacific varying between 16, 12, and 24 months for CM2.5 FLOR, CCSM4, and CESM1, respectively.

Past studies of potential (i.e., perfect model) ENSO predictability (e.g., Boer 2000; Collins et al. 2002; Power et al. 2006; Wittenberg et al. 2014) employed the more commonly used approach, known as reforecast experiments, in which a target trajectory in a model is forecasted by an ensemble of perturbed runs. The model-analog technique appears to yield similar results at less computational expense. For example, we obtain an MPLT of about 20 months in CM2.1 through the model-analog method, which agrees with the estimate based on CM2.1 reforecasts in Wittenberg et al. (2014).

b. Sensitivity to library length and ensemble size

One potential area of parametric dependence for analog forecasts is the library length necessary to achieve useful skill. For our model-analogs, this question has a practical aspect, since it might give guidance on how long to run the control simulation. The library needs to sample the attractor for the phenomenon of interest (e.g., ENSO) well enough to provide sufficient analogs near a given target state. The required library length will thus depend on the required number of analogs (i.e., the required resolution of the forecast PDF, and the required robustness of the forecast ensemble mean), the required precision of the analogs (i.e., the acceptable neighborhood width in phase space), and the attractor's recurrence time for the target state(s) of interest (which is related to the attractor's dimension and the ergodic time scale of the phenomenon). The library can be expanded by either lengthening trajectories (e.g., running a longer control simulation) or diversifying trajectories (e.g., increasing the number of independent ensemble members). Van den Dool (1994) estimates that the required library length depends on the number

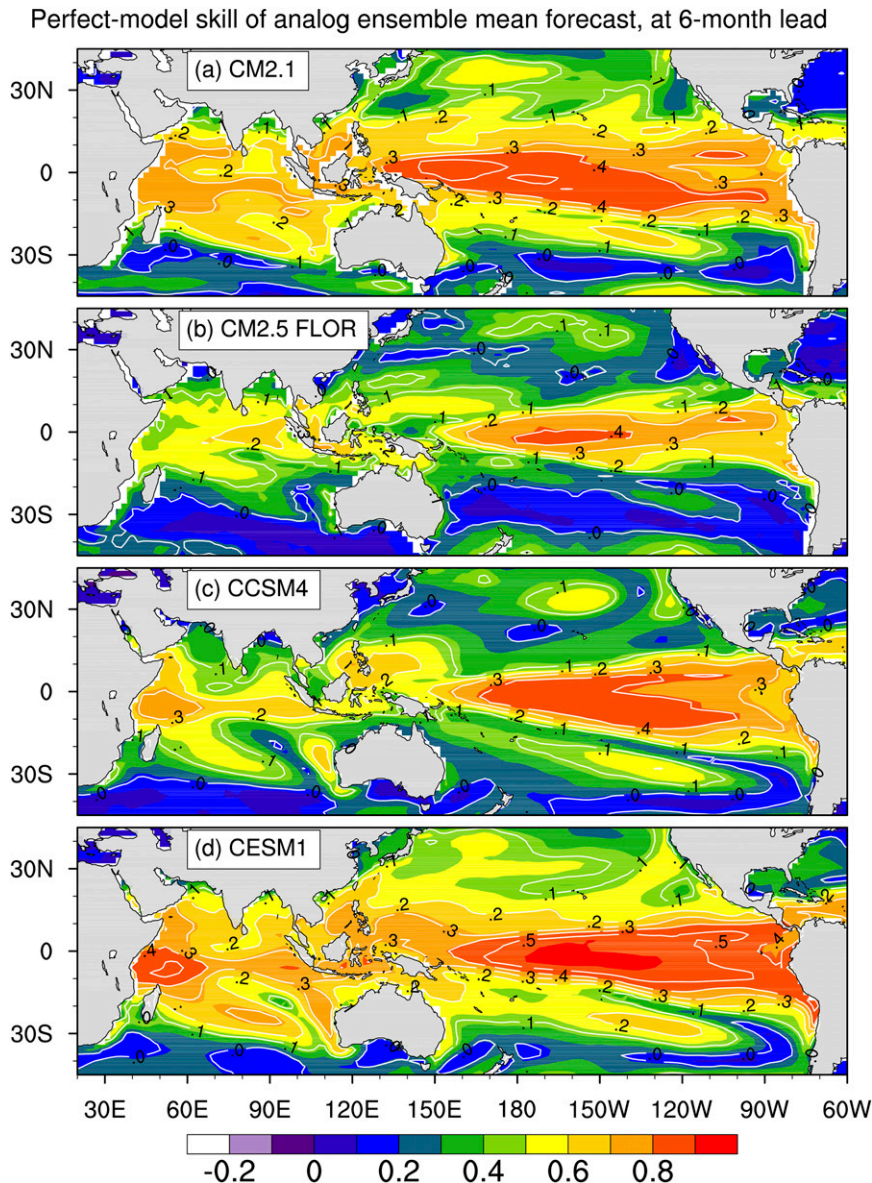Perfect-model skill of analog ensemble mean forecast, at 6-month lead

FIG. 4. As in Fig. 2, but for perfect-model skill of ensemble-mean forecasts at 6-month lead.

of spatial degrees of freedom. Here, we consider this question using the GFDL CM2.1 control simulation, the longest simulation available in this study, although results are qualitatively similar for all four models (not shown).

Using the first 200 years of the GFDL CM2.1 simulation as the verification period, we tested varying N from 50 to 3800 years. Extending the training period in this way corresponds to extending the control simulation. Also, since more analogs are potentially available with increasing library length, we tested ensemble size by letting K range from 1 to 70. Setting $K = 1$ corresponds to using the nearest analog with distance d_1.

The impacts of library length and ensemble size are first evaluated in terms of the representation of the initial target states, using the distance d_K between the target states $\mathbf{x}(t)$ and corresponding ensemble-mean model-analog state vectors $\mathbf{z}(t)$ (section 2). Figure 6a shows the time mean of $d_K(t)$ determined over the verification period $\langle d_K \rangle$ scaled by the climatological root-mean-square distance of states from the verification period, the square of which is given by $\sigma^2 = \sum_{t=1}^{T}\sum_{i=1}^{2}\sum_{j=1}^{J}[x_j^i(t)/\sigma_X^i]^2/T$, where $T = 12M$, and M is the number of verification years. In all cases, the initial ensemble-mean distance is minimized by using an analog ensemble mean rather than the single best analog
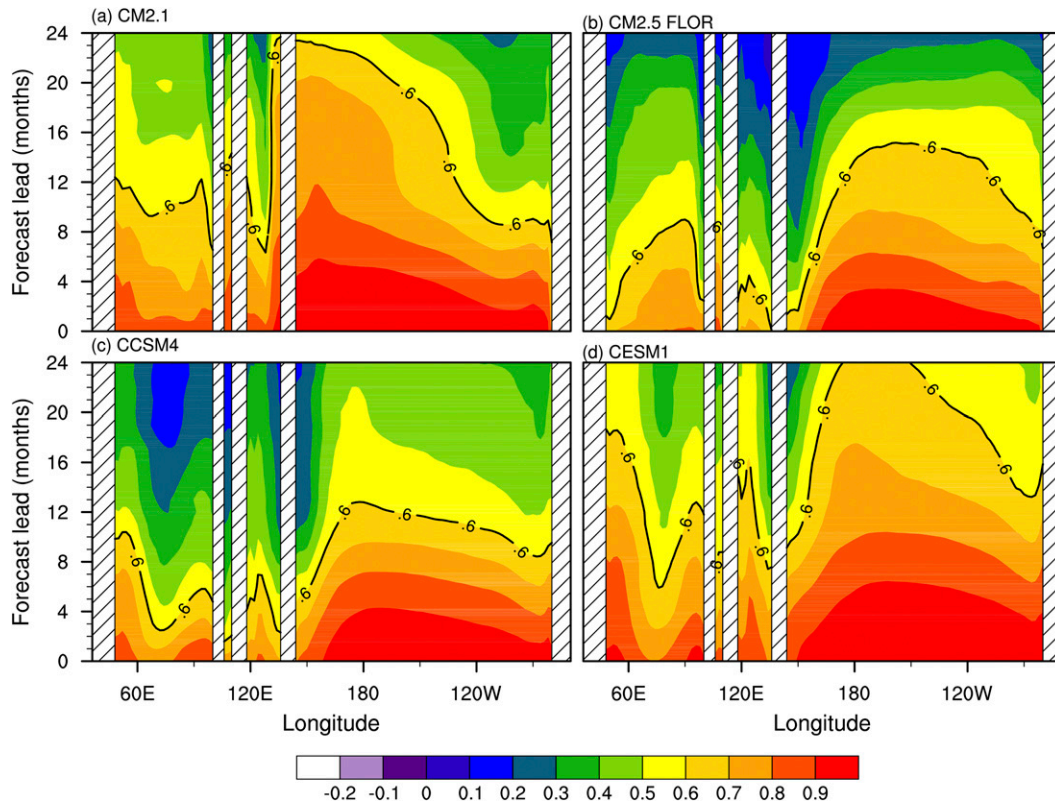
FIG. 5. Perfect-model equatorial SSTA forecast skill from AC as a function of forecast lead for each model's 200-yr verification period for (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, and (d) CESM1. SSTA is averaged between 5°S and 5°N; areas with hatching are continents. Results shown are a function of longitude (abscissa) and lead time (ordinate).

(Fig. 6a). Most of this improvement comes from simply increasing K from 1 to 5, and in fact results are optimal for $K > 10$–20 (with a weak dependence upon N), becoming worse for further increase of ensemble size. Clearly, the gain in additional ensemble members can be offset by adding members that are increasingly far from the target state. On the other hand, the minimum value of $\langle d_K \rangle$ decreases monotonically as N is increased, consistent with the availability of more library states, increasing the chance of better matches to the target state. However, much of the improvement is realized by $N \sim 500$ with relatively less improvement thereafter, even as the length of the training period is further extended.

Potential predictability estimates similarly depend upon N and K. This is shown in Fig. 6 for Niño-3.4 SSTA forecast skill at a lead time of 6 months, for both AC (Fig. 6b) and RMSSS (Fig. 6c). Forecast skill again maximizes for $K \sim 10$–20, with most of the improvement in forecast skill apparent for $K \sim 5$, and again most of the benefits of library size are reached for $N \sim 500$, with more modest improvement thereafter.

Note that forecast skill is less sensitive to the ensemble size K as the library size increases, especially for $N \gtrsim 1000$.

c. Sensitivity to verification period

A historical record of 200 years may not be long enough to completely sample ENSO events (Wittenberg 2009), so we also tested sensitivity to the choice of the verification period. First, we repeated the analog calculation using the GFDL CM2.1, but used the first 500 years instead of the first 200 years of the simulation as the verification period, with the remainder used for the library as before. However, we found no appreciable change (not shown); for example, Niño-3.4 SSTA 6-month-lead forecast skill increased from 0.81 to 0.83. Next, the CESM1 simulation was divided into two nonoverlapping portions of 350 years each, with each half of the simulation using the verification period with the other half serving as the library. The skill was the same in both cases and very close to the CESM1 results shown above. Our sensitivity experiments all suggest that potentially useful model-analog
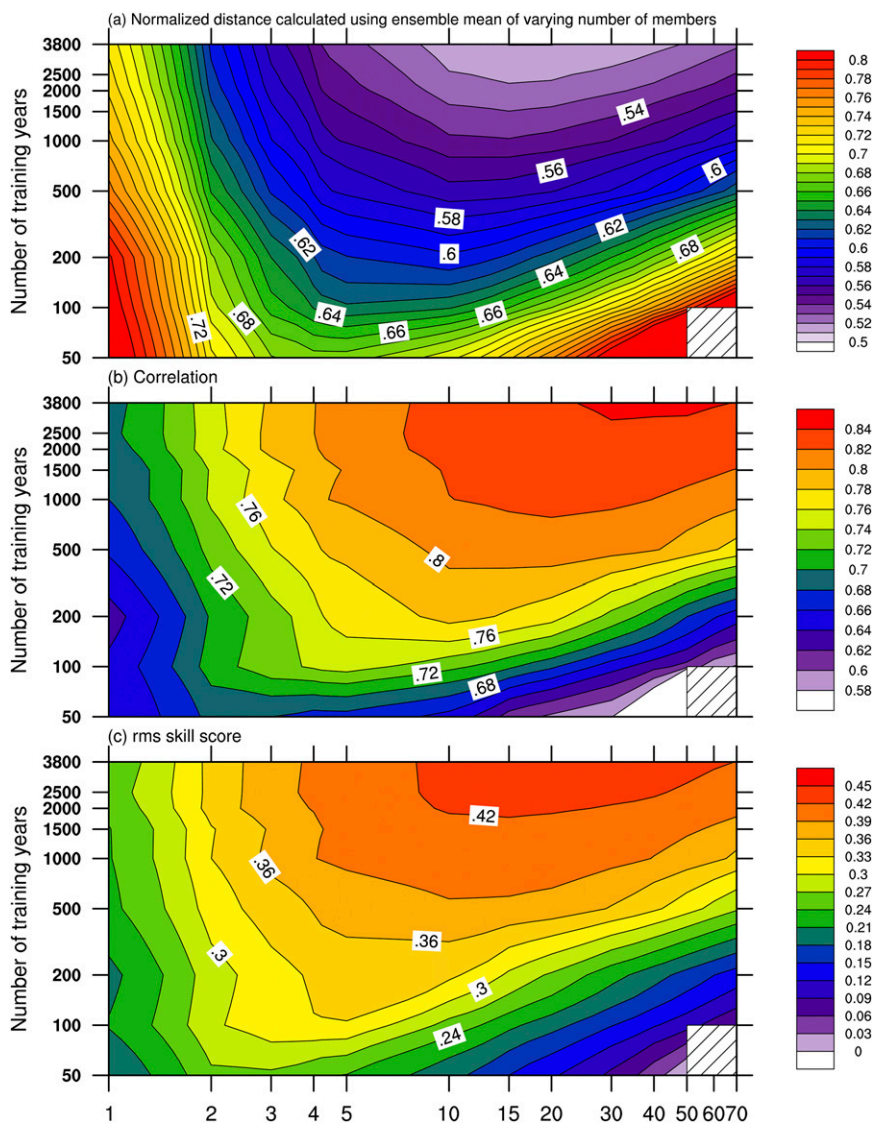
FIG. 6. (a) Time mean of initial d_K calculated using the ensemble mean of analogs and 6-month-lead forecast skills of Niño-3.4 SST anomalies measured by (b) correlation and (c) RMSSS as a function of ensemble size of analogs (abscissa) and the number of years in a training period (ordinate). The first 200 years of the CM2.1 control run is used as the verification period. In (a), distance is scaled by climatological-mean distance of states from the verification period. All values tested are indicated next to the tick marks of both axes.

ENSO forecast skill can be obtained with surprisingly small ensemble size, library length, and verification periods.

4. Retrospective forecasts using model-analogs

In this section we use the model-analogs to make retrospective forecasts (hindcasts) of observed SSTAs in the tropical Indo-Pacific region, and compare their skill to initialized NMME seasonal forecasts.

a. Description of model-analog forecast technique

To make real-world forecasts, we find model-analogs to target state vectors constructed from observed anomalies. Note that now we use the entire dataset from each model simulation as its data library, in contrast to the analysis in section 3. We constructed model-analog hindcasts for forecast leads of 0–12 months, where month 0 is the initial (target) state and the ensuing forecast leads are chosen to match the forecast range of
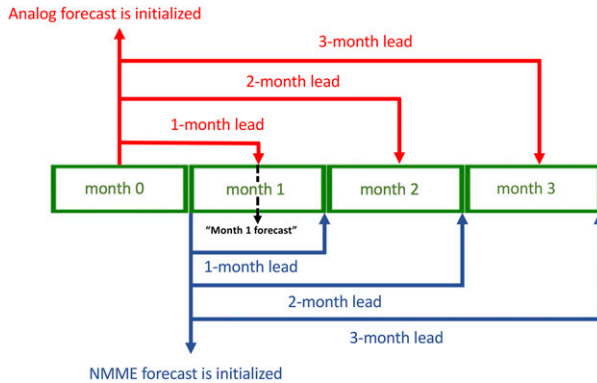
FIG. 7. A schematic of how forecast lead time is defined for the model-analog and NMME hindcasts.

the NMME hindcasts. (Clearly, the model-analog forecasts could be extended beyond 12 months.) The NMME forecasts were initialized on or near (for staggered starts) the first day of each month. The month-0.5 forecast was then the mean of the first month of the forecast run, that is, centered in the middle of the calendar month. The equivalent model-analog forecast was initialized with the monthly mean observations centered on the previous month (i.e., month 0) so that the 1-month-lead model-analog forecast and the NMME month-0.5 forecast verified at the same time. We renamed both these forecasts the month-1 forecast, and so on for increasing forecast lead times. Thus, to match NMME hindcasts initialized between January 1982 and December 2009, the model-analog hindcasts should be initialized from December 1981 to November 2009. Figure 7 shows a schematic of how each forecast system is initialized. In the following, both model-analog and NMME hindcasts are verified against OISST data.

As in section 3b, we investigated the sensitivity of the model-analog reconstruction of observations to ensemble size K and library length N. Figure 8 shows $\langle d_K \rangle$ for each of the four models. In general, the effects of library length and ensemble sizes resemble those in Fig. 6a, with more improvement from using even a small ensemble rather than from increasing library length. Note that for CM2.1, $\langle d_K \rangle$ is much greater for observations than for the perfect model (cf. Figs. 6a and 8a), which is not surprising given that the models all suffer from biases in simulating ENSO variability (Delworth et al. 2006; Wittenberg et al. 2006). It also appears that $\langle d_K \rangle$ has a similar dependence on N for all four models. Consistent with our findings in section 3b, it appears to be sufficient to construct model-analogs using a historic record of approximately 500 years with $K \sim 10$–15. In the following, we use $K = 15$ for each model case.

b. Comparing model-analog and NMME hindcast skill

How well each set of model-analog ensemble means represent observed target anomalies is assessed using AC and RMSSS (Fig. 9). Some observed states, notably the mature stages of the 1982/83 and 1997/98 El Niño events, are not as well reconstructed in model space as others (not shown, but see $d_K(t)$ in Fig. S4 of the supplemental material, which is essentially a measure of how well each model reproduces observed anomalies). As in the perfect-model analysis (Fig. 2), the model-analogs best reproduce SSTA variability in the equatorial Pacific, where the AC and RMSSS are as high as 0.9 and 0.6, respectively. Model-analogs also reproduce some observed SSTA variability in the subtropical Pacific, particularly as seen in relatively high AC (0.6) and RMSSS (0.3) values in the western part of the basin. In the northern subtropical Pacific, the model-analogs all represent aspects of the North Pacific meridional mode (Chiang and Vimont 2004), which has been linked to ENSO variability (Chang et al. 2007) via the seasonal footprinting mechanism (Vimont et al. 2003; Alexander et al. 2010).

However, outside of these regions the model-analogs provide a poorer representation of observed target anomalies than they did for the model states (cf. Figs. 9 and 2). This is most notable in the equatorial Indian Ocean, where the perfect-model-analogs mostly have initial correlation values that were nearly as high as in the Pacific, while the model-analogs of observations do not. So, compared to the perfect model, variability in the Indian Ocean plays a much smaller role in defining the model-analogs. This is also the case in the western equatorial Pacific, most dramatically for the CM2.1, where the model space for the GFDL CM2.1 control run with preindustrial forcings is apparently so different from nature in the late twentieth century that it is not possible for any model-analog to simultaneously match observed SSTAs in both the eastern and western parts of the basin. The bias in the model degrades analog forecast skill as well as corresponding NMME forecast skill in the western equatorial Pacific (see below).

Month-6 model-analog and NMME hindcast skill are shown in Figs. 10a,c,e,g,i and 10b,d,f,h,j, respectively. The bottom row shows the four-model grand ensemble mean of the model-analog and NMME hindcasts. The model-analogs clearly reproduce the essential details of skill from the NMME hindcasts. In particular, both sets of hindcasts are skillful in the central and eastern equatorial Pacific, where correlation is as high as 0.7. The model-analogs have higher skill along the equator
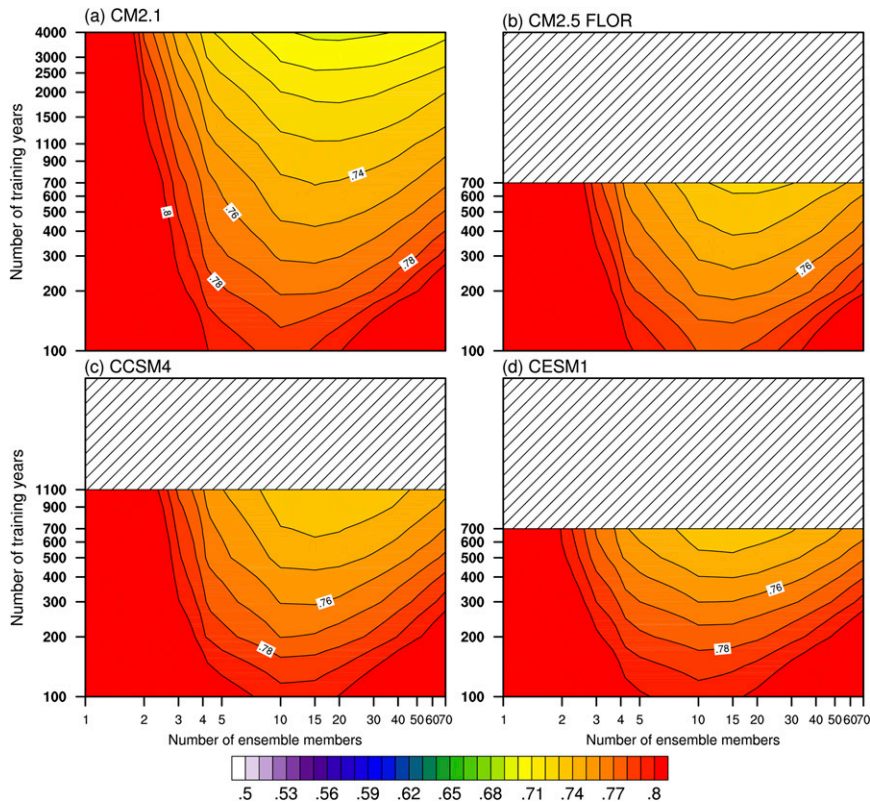
FIG. 8. Time mean of d_K calculated using the ensemble mean of analogs calculated from (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, and (d) CESM1. Analogs are searched for observed states from December 1981 to November 2009.

and South American coast than the NMME hindcasts. Month-6 model-analog hindcasts also display skill in the northwestern tropical Pacific and in the South Pacific near 30°N, 150°W, with a correlation of 0.5. In these regions, SSTA variability is largely driven by ENSO via the atmospheric bridge (Alexander et al. 2002). In general, NMME hindcasts outperform model-analog hindcasts in the northwestern tropical Pacific except for CESM1. In the South Pacific, skill is comparable. However, the model-analogs appear to have poorer skill than the corresponding NMME hindcasts in much of the western tropical Pacific and Indian Ocean, except for CM2.1.

The evolution of equatorial skill as a function of forecast lead time is shown in Fig. 11 for the model-analogs and their corresponding hindcasts, with the bottom row again showing the grand ensemble means of each. Model-analog skill is comparable to the NMME hindcast skill throughout the Pacific for all forecast leads. The model-analogs are more skillful than the NMME hindcasts near the date line and along the South American coast. The poorer Indian Ocean skill for the model-analogs remains evident.

Figure 12 shows how well the model-analogs capture individual events by comparing the month-6 model-analog hindcasts of the Niño-3.4 time series to observations and the month-6 NMME hindcasts. The model-analog hindcasts reproduce most of the strong ENSO events (blue lines in Fig. 12), particularly the 1982/83 and 1997/98 El Niño events and the 1988/89 and 1998/99 La Niña events, and again appear to have skill comparable to that of the NMME. We carried out a more detailed skill comparison by applying the random walk test of DelSole and Tippett (2016) to RMSSS values of month-6 Niño-3.4 hindcasts (Fig. S5 in the supplemental material). We found that the model-analogs for the CM2.1 and CCSM4 were both significantly more skillful (at the 95% significance level) than the corresponding NMME hindcasts. The model-analogs for the other two models, as well as the multimodel ensembles, were equally as skillful as the corresponding NMME hindcasts.

The pattern correlation within the Indo-Pacific training region of observed and month-6 hindcast SSTA is shown for both the model-analog and NMME in Fig. 13. Both sets of hindcasts show similar variations of skill,
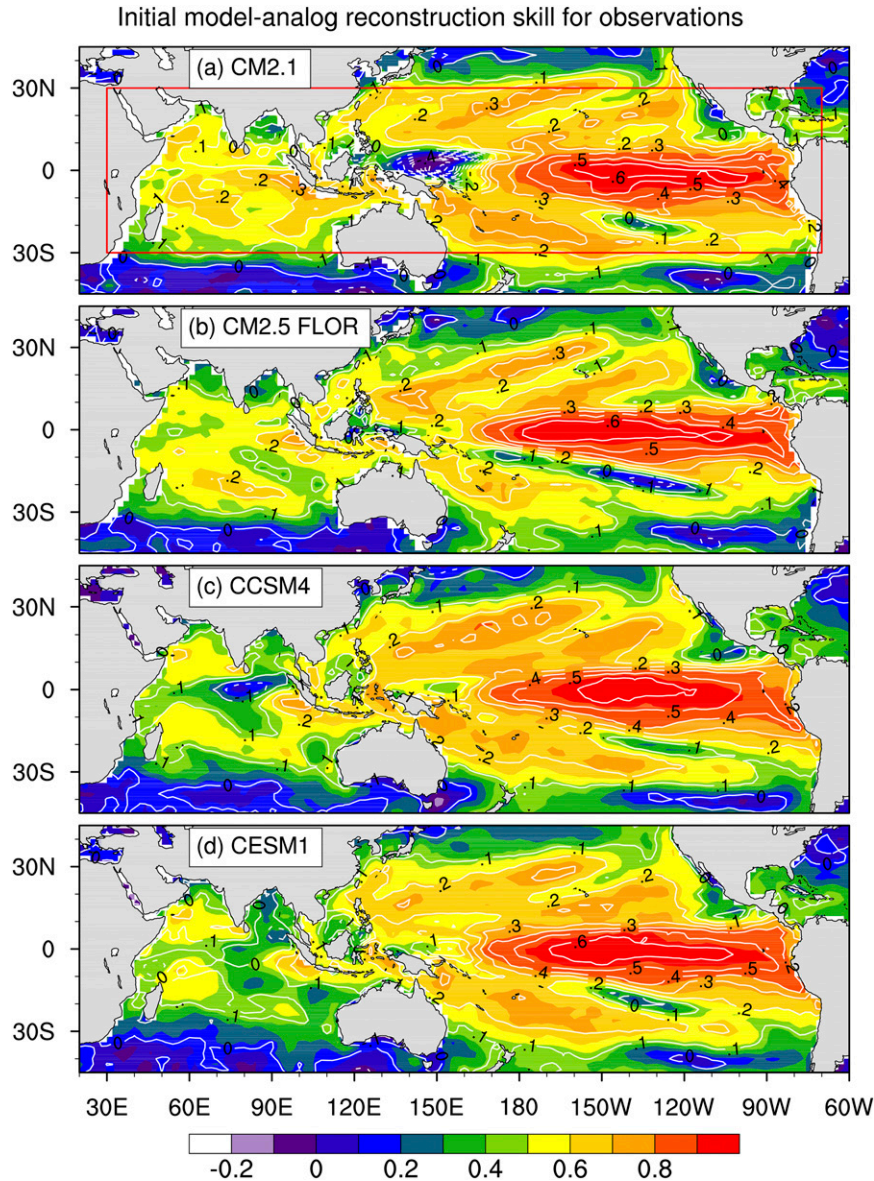
FIG. 9. Initial model-analog ensemble-mean reconstruction skill for observed (1982–2009) SSTA in (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, and (d) CESM1. Shading and contours denote correlation and RMSSS, respectively. In each model, all data are used to define analogs.

although there is a tendency for the model-analogs to perform better in ENSO events than in neutral conditions. Again, the model-analog hindcasts perform as well as the NMME hindcasts although there are discrepancies between the two forecast methods. The time-mean values of the pattern correlations are 0.34 (0.29, 0.32, and 0.34) and 0.34 (0.39, 0.37, and 0.37) for the model-analog and NMME hindcasts using CM2.1 (CM2.5 FLOR, CCSM4, and CESM1, respectively).

Finally, Fig. 14 shows probabilistic skill based on the upper (El Niño), middle (ENSO neutral), and lower

(La Niña) terciles of the SSTA distributions. RPSS calculated from the multimodel ensembles of model-analog and NMME month-6 hindcasts are shown in Figs. 14a and 14b, respectively. Again, the model-analog technique clearly reproduces the essential details of skill from the NMME hindcasts in the tropical Indo-Pacific region (see also Tippett et al. 2018). In particular, both sets of hindcasts are more skillful than climatology in the central and eastern equatorial Pacific, with RPSS above 0.3. Plots of the corresponding reliability (Figs. 14c–e, top panels) and frequency of occurrence (Figs. 14c–e,
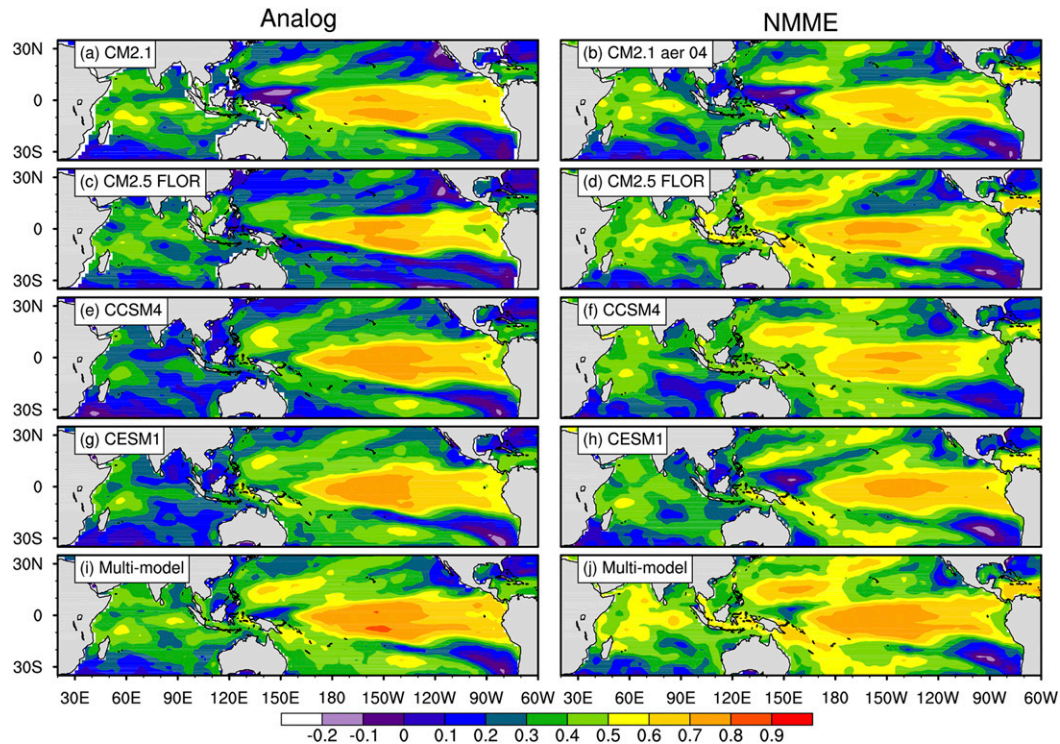
FIG. 10. (left) Model-analog and (right) NMME hindcast skills of observed SST variations at 6-month lead. Shading denotes correlation. Analog and NMME hindcasts are based on (a),(b) CM2.1, (c),(d) CM2.5 FLOR, (e),(f) CCSM4, (g),(h) CESM1, and (i),(j) multimodels. The grand mean of 60 analogs is shown in (i), which consist of 15 best initial analogs from each of the four control runs. The grand mean of the four NMME experiments is shown in (j).

bottom panels) for the same three categories, averaged over the central and eastern equatorial Pacific (6°S–6°N, 160°E–80°W), likewise show that the model-analog and NMME hindcast ensemble distributions are comparable within the ENSO region. For the upper and lower terciles, the model-analogs appear somewhat less sharp but more reliable than the NMME hindcasts, which corresponds to the slightly higher overall values of RPSS; both hindcast ensembles are overconfident when forecasting neutral conditions.

c. Evaluating sources of model-analog skill

Recently, Newman and Sardeshmukh (2017) showed that a low-dimensional empirical dynamical model of monthly SSTA and SSHA variability, a linear inverse model (LIM) with approximately 30 degrees of freedom, had similar tropical Indo-Pacific SSTA skill to the grand ensemble mean of all eight NMME operational models, and that both LIM and CGCM spatial and temporal skill variations were largely predicted by the LIM itself. Therefore, they suggested, the seasonal predictability of tropical SSTAs is effectively linear, and both forecast systems may already be near the intrinsic

limit of tropical Indo-Pacific SSTA predictability, with two exceptions: potential nonlinear predictability in the far eastern tropical Pacific, where the NMME-mean hindcast skill was greater than the LIM, and significant NMME forecast error in the western tropical Pacific, where the LIM alone had positive skill.

We might then ask whether the skill of the model-analogs is also a consequence of low-dimensional linear predictability. Certainly, the model-analogs appear effectively low-dimensional. While the operational models have detailed full-field initializations, the model-analogs are initialized only with tropical Indo-Pacific SSTA and SSHA, and have relatively large initial errors apart from the leading PCs (Fig. 3). In fact, training the model-analogs using only the leading 10–20 combined SSTA–SSHA PCs (i.e., with 100% initial error in all higher-order PCs) led to initial error and forecast skill nearly identical to that presented in this paper, a result that is consistent with previous reduced-space empirical dynamical models of ENSO (e.g., Penland and Sardeshmukh 1995; Berliner et al. 2000; Chen et al. 2017). Capturing both perfect-model and actual NMME hindcast skill with analogs determined from such
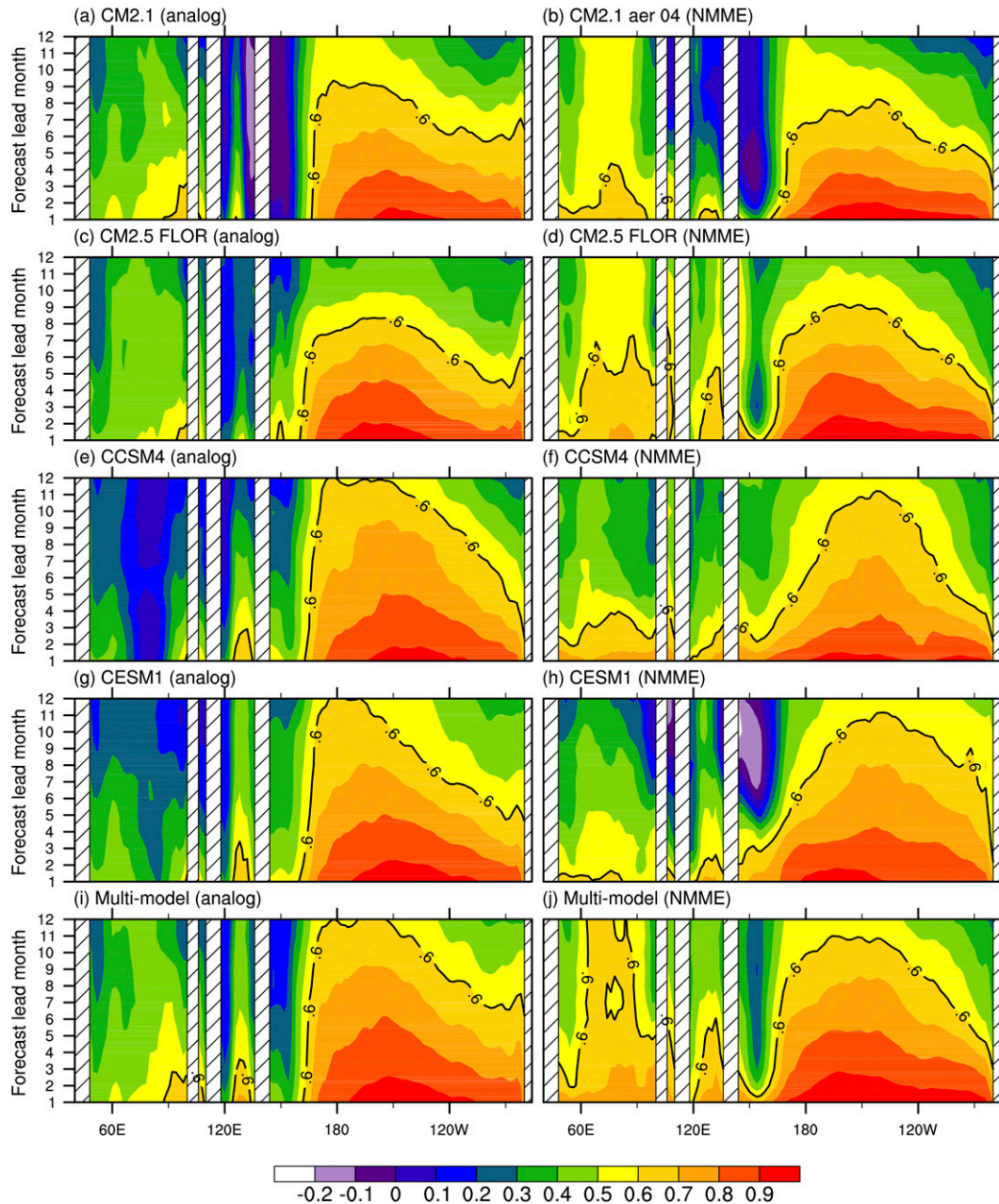
FIG. 11. (left) Model-analog and (right) NMME hindcast skill (relative to observations) for SST averaged between 5°S and 5°N, as a function of longitude (abscissa) and lead time (ordinate). Both shading and contours denote correlation. Analog and NMME hindcasts are based on (a),(b) CM2.1, (c),(d) CM2.5 FLOR, (e),(f) CCSM4, and (g),(h) CESM1. The grand mean of 60 analogs is shown in (i), which consist of the 15 best initial analogs from each of the four control runs. The grand mean of the four NMME experiments is shown in (j). Hatching indicates continents.

relatively short data libraries also suggests that the ef-fective number of degrees of freedom must be $O(10)$ as opposed to $O(100)$ or higher (Van den Dool 1994). Likewise, the ensemble members of the model-analogs are close enough [in the Lorenz (1969) sense] that their evolution can be used to estimate predictability (i.e., trajectory divergence) within the CGCMs.

To test whether the model-analog skill is effectively linear, the analysis of the previous section was repeated using "antianalogs" from the control simulations, de-termined by 1) reversing the target anomaly sign; 2) constructing an analog ensemble, its subsequent evo-lution, and ensemble mean as before; and 3) reversing the sign again to yield the forecast. Skill of these antianalogs
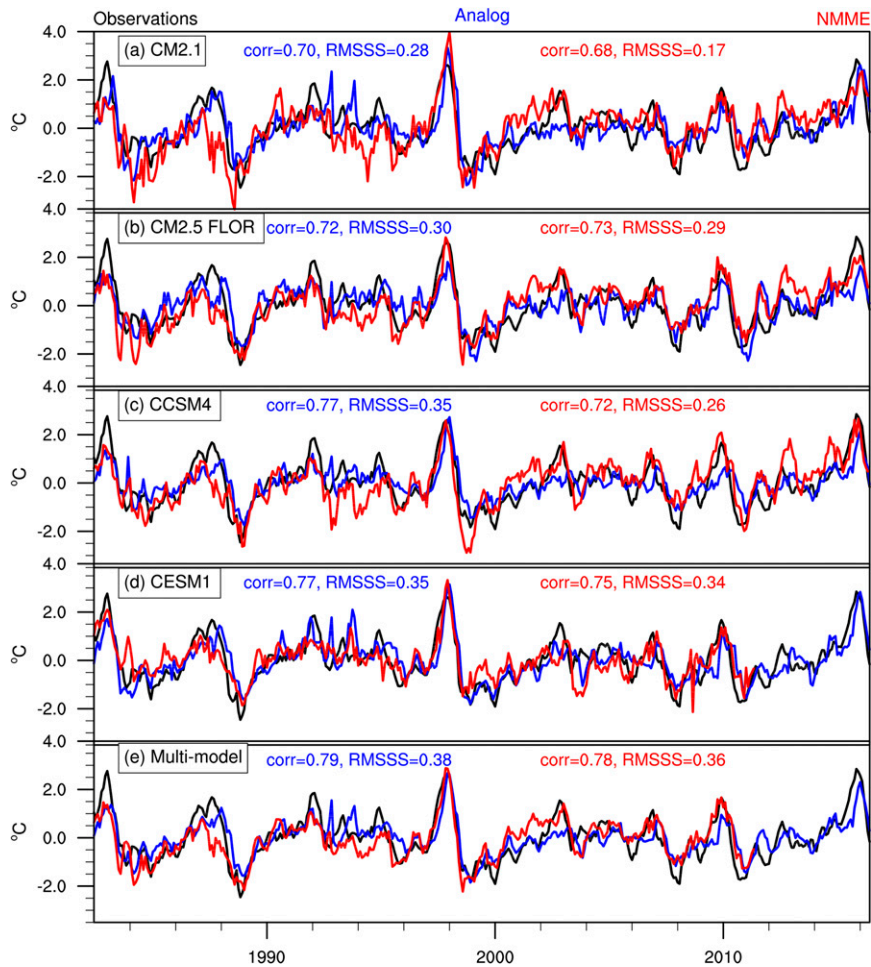
FIG. 12. Niño-3.4 SST anomalies from observations (black lines), compared to month-6 model-analog (blue lines) and NMME (red lines) hindcasts in (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, (d) CESM1, and (e) four-model grand mean.

for month 6, shown in Fig. 15, can then be compared to the corresponding model-analog skill (Figs. 10a,c,e,g). To first approximation, the patterns and values are largely similar, consistent with effectively linear skill, particularly in the central equatorial Pacific, north-western tropical Pacific, and Indian Ocean. However, all four models show substantial differences between analog and antianalog skill in the far eastern Pacific, notably the Niño-1.2 region, supporting the importance of predictable nonlinearity there that is captured by the model-analogs. Other areas where the skill of the analogs exceeded the antianalogs are generally regions of low skill and where these differences may also not be significant.

The LIM is a Markovian model; that is, its forecasts depend only on the current state, not on any past states. This is a justification for the approach applied to determine analogs in this paper, but of course it is not a necessary component of the model-analog technique.

For example, we can incorporate information about trajectory velocity by defining analogs using a different metric than in (1), such as

$$d^2(t,t') = \sum_{i=1}^{2} \sum_{j=1}^{J} \left\{ \left[\frac{x_j^i(t)}{\sigma_X^i} - \frac{y_j^i(t')}{\sigma_Y^i} \right]^2 + \left[\frac{dx_j^i(t)}{\sigma_X^i} - \frac{dy_j^i(t')}{\sigma_Y^i} \right]^2 \right\},$$

where $dx_j^i(t) = x_j^i(t) - x_j^i(t-1)$ and $dy_j^i(t') = y_j^i(t') - y_j^i(t'-1)$. However, using this metric to select model-analogs led to no forecast skill improvement (Fig. S6 in the supplemental material).

In the key ENSO regions, the model-analogs have higher skill than their NMME hindcast counterparts, and perfect-model skill is also comparable to actual skill. Elsewhere, model-analog skill is lower than NMME hindcast skill and much lower than perfect-model skill. The reasons for this remain to be determined. One
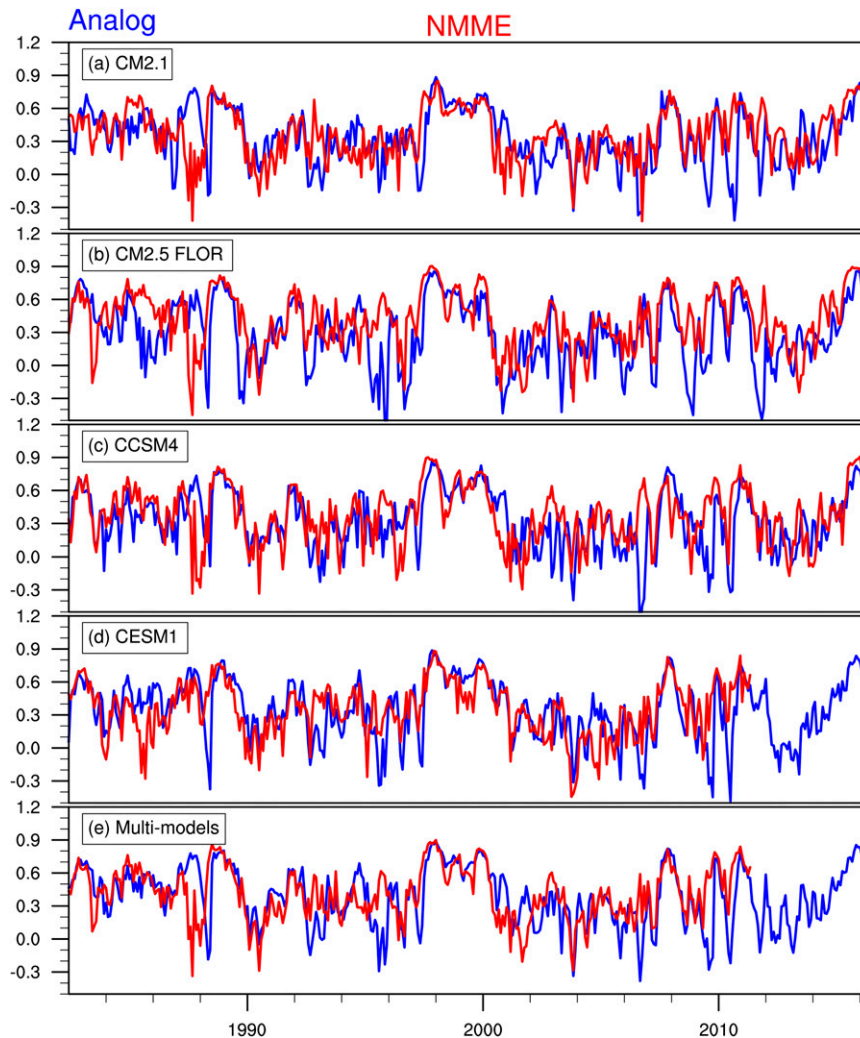
FIG. 13. Evaluation of retrospective forecast of SST at 6-month lead by pattern correlation in (a) CM2.1, (b) CM2.5 FLOR, (c) CCSM4, (d) CESM1, and (e) four-model grand mean. SST in the training region is involved in the calculation of pattern correlation. Blue lines denote analog hindcasts; red lines denote NMME hindcasts.

possibility is suggested by the notably poorer representation of observed initial conditions provided by the model-analogs outside the equatorial Pacific compared to the perfect-model analysis (cf. Figs. 9 and 2). During the hindcast period, the tropical SST warming trend was also much larger outside of the equatorial eastern Pacific, especially in the Indian Ocean and west Pacific warm pool (e.g., Solomon and Newman 2012). However, all the control simulations used for the model-analogs were run under fixed radiative forcing conditions. Interestingly, we found that the CM2.5 model-analogs had higher Indian Ocean hindcast skill when trained from the entire CM2.5 library, which included a (spurious) trend resulting from a spinup period of several hundred years rather than from the

equilibrated last 700 years that we ultimately employed here (not shown). Moreover, the LIM hindcasts, which used a fixed dynamical operator but whose initial conditions included the trend, had Indian Ocean skill that was similar to the NMME hindcasts. This suggests that our model-analog skill was negatively impacted by external radiative forcing. If so, training model-analogs with large ensembles of externally forced model simulations [such as in Kay et al. (2015)] could improve skill.

5. Summary and discussion

We have investigated the use of analogs based on long climate simulations, rather than observational datasets,
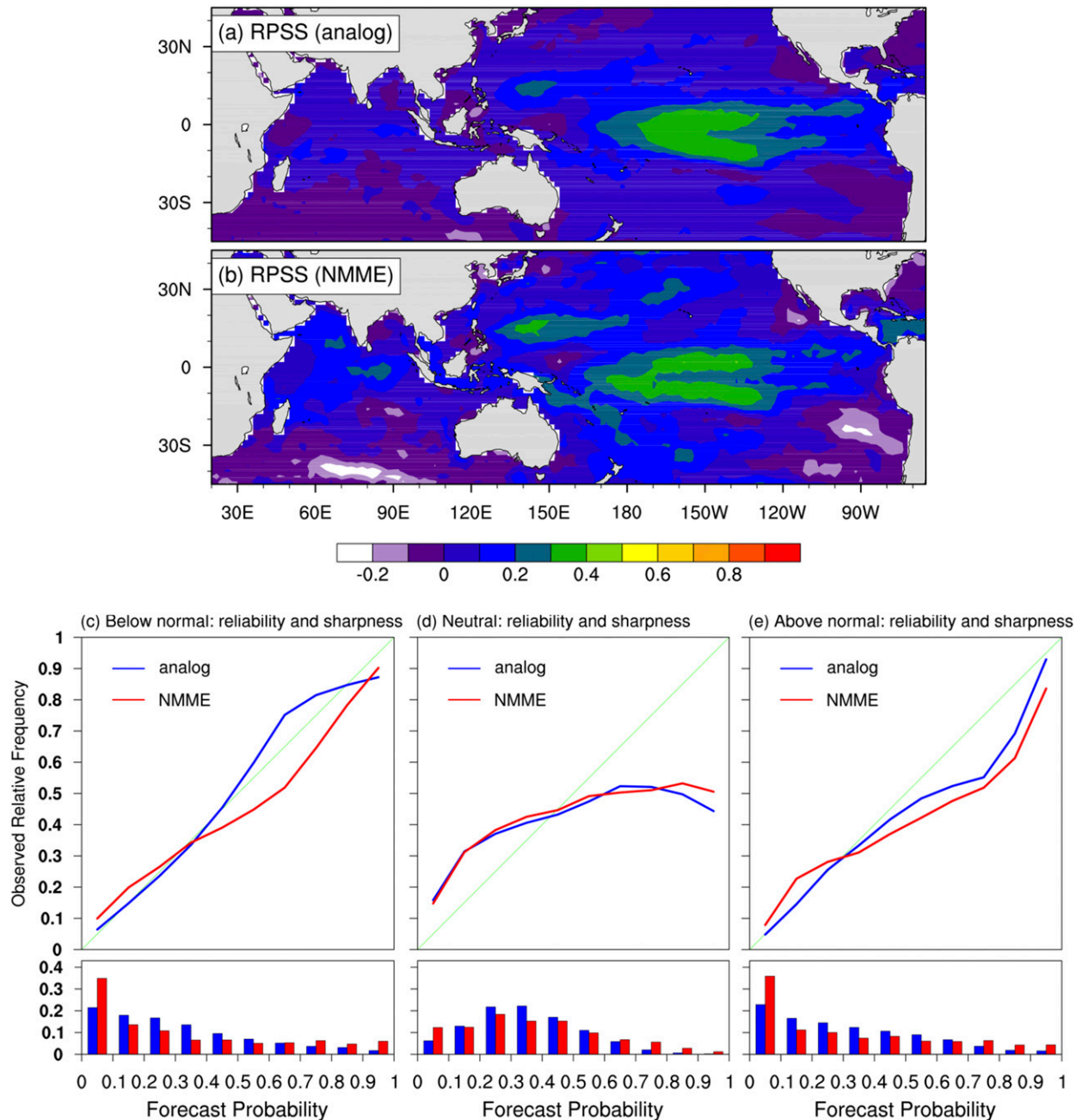
FIG. 14. Metrics of month-6 SSTA hindcast probabilistic skill. RPSS of multimodel hindcast ensembles for (a) model-analog and (b) NMME hindcasts. Positive (negative) values indicate probabilistic skill that is better (worse) than a forecast of climatology. (c) Reliability and sharpness (averaged over the central and eastern equatorial Pacific region; 6°S–6°N, 160°E–80°W) for multimodel model-analog (blue) and NMME (red) hindcast ensembles, shown for the (c) lower (below normal), (d) middle (neutral), and (e) upper (above normal) terciles. In (c)–(e), the bottom panels show the frequency of occurrence (i.e., sharpness) for each tercile category for model-analog and NMME hindcasts, corresponding to the reliability diagram in the top panels.

to make monthly forecasts of observed tropical Indo-Pacific Ocean SST and SSH anomalies. Analogs are chosen from model states that are most similar to the observed target state in a given calendar month, where both are defined only from tropical Indo-Pacific Ocean SSTAs and SSHAs. An analog ensemble is defined from those model states that minimize the distance to the target state. This model-analog ensemble and its subsequent time evolution then yields the time-evolving forecast ensemble from the target state.
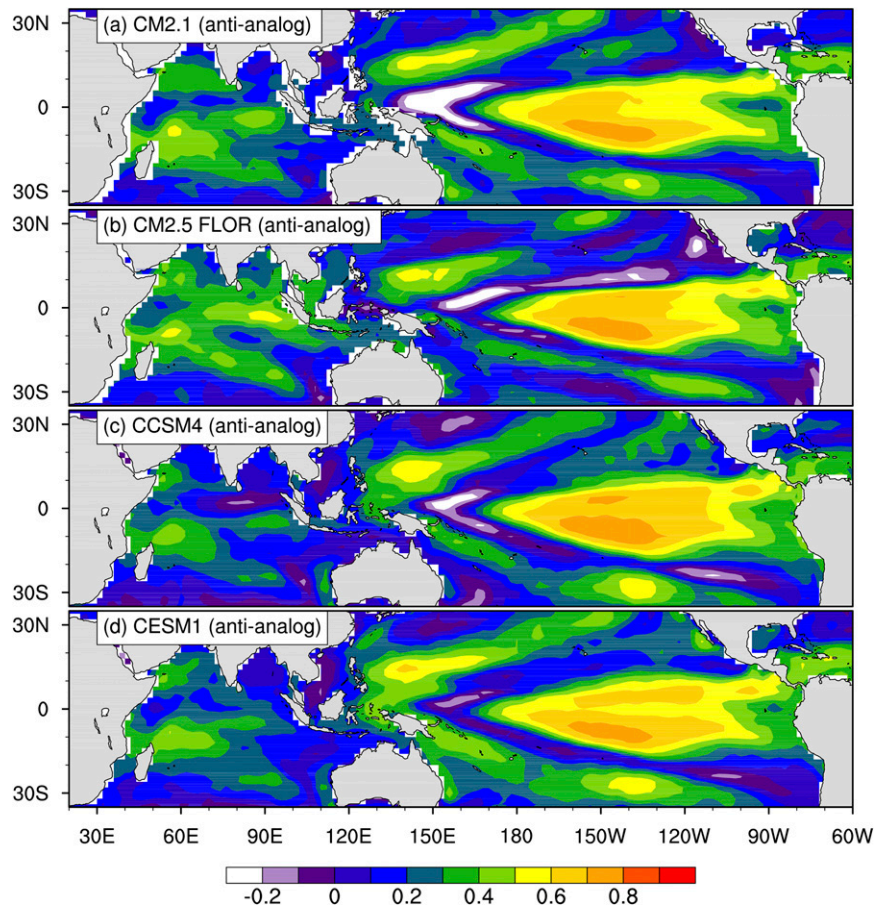
FIG. 15. Antianalog hindcast skill of observed SST variations at 6-month lead. Compare with Figs. 10a,c,e,g.

For ensemble sizes of about 10–15 members, 500 years of model output appears sufficient to generate estimates of ensemble-mean perfect-model skill, and to make hindcasts of the 1982–2009 period that have generally similar deterministic and probabilistic skill as the initialized hindcast ensembles from four climate models that are also used to make operational seasonal forecasts as part of the NMME project. This appears consistent with Wittenberg (2009), who showed that a record of 500 years is required in order to obtain a complete sampling of ENSO events in the CM2.1. Extending the model simulation record to 4000 years yielded only slight skill improvement. More important was the initial increase of ensemble members from 1 to 10–15, and in fact, skill was degraded for ensemble sizes greater than about 20–30. Our results thus raise the possibility that the tropical Indo-Pacific forecast skill of any CGCM, both in a historical and a perfect model sense, could be assessed using the model-analog technique as soon as a sufficiently long control run is produced, an assessment that could even be made part of the model development cycle itself.

The skill of the model-analog hindcasts is all the more striking given the relatively small amount of data that "initializes" them; we used only monthly anomalies of tropical Indo-Pacific SST and SSH, neither field drawn directly from a data assimilation system employing the CGCM in question. The relatively large initial ensemble spread also raises the possibility that for the tropical Indo-Pacific region, seasonal forecasts may only require accurate initialization within a low-dimensional subspace involving relatively large spatial scales. Moreover, this spread may be sufficient to estimate model potential predictability based on the divergence of an ensemble of analog trajectories (e.g., Lorenz 1969), consistent with earlier suggestions that on monthly time scales, the tropical Indo-Pacific region could be considered a coarse-grained, Markovian dynamical system (e.g., Newman and Sardeshmukh 2017).

Our model-analog technique is, in essence, an attempt to reconstruct the phase space of a CGCM from a library of its states. The theoretical basis for phase-space reconstruction is that the underlying attractor may be

reconstructed using a succession of states, represented within so-called embedding vectors, to reproduce past trajectories (Takens 1981; Sugihara and May 1990; see also Ye et al. 2015). In practice, this means that analogs determined using multiple time lags could capture information about the phase-space trajectory near the target state (e.g., McDermott and Wikle 2016; Lguensat et al. 2017; McDermott et al. 2018). Weighting analog ensemble members by their distance from the target state, such as with kernel-weighting strategies (e.g., Zhao and Giannakis 2016), may also enhance this approach. Note, however, that for Markovian dynamics, previous states provide no predictive information, which might justify the simpler method of our study. Moreover, the larger data libraries available from climate model simulations, compared to the more limited observational datasets used in previous analog studies, likely aided our analog determination. Still, these issues will need to be reevaluated as the model-analog technique is applied to other forecast problems.

The model-analog method potentially has some notable advantages compared with the conventional seasonal forecasting approach in which an ensemble of initialized climate model forecasts is run (e.g., Latif et al. 1998; Keenlyside et al. 2005; Kirtman et al. 2014). First, initialization shock may be avoided in the model-analog forecasts, since both initial conditions and their subsequent trajectories are taken from a free-running model in which each component of the model is well balanced and the consistency of model physics is retained. Recall that model-analog skill is higher than the corresponding model hindcast skill throughout the ENSO region. The improvement is greatest in the Niño-1.2 region, with differences large enough to be statistically significant for each CGCM's model-analog and for their ensemble mean, suggesting that the NMME hindcast error in this region is at least partly due to initialization shock. Moreover, since the model-analog anomalies are always in the model space, the mean bias in their forecasts is removed by construction. This may also improve model-analog skill relative to the NMME hindcasts, whose bias correction is determined retrospectively from a relatively short 28-yr training period.

Second, the use of the model-analog technique could mean the reduction of technological barriers for new models to participate in seasonal forecasting, which would further enhance the diversity of multimodel ensembles and potentially allow for improved ensemble forecast calibration. Certainly, it is encouraging that a skillful ensemble of model-analogs can be constructed from a control simulation of just a few hundred years in length, since such runs are often routinely executed, such as for the IPCC assessments (Meehl et al. 2007; Taylor et al. 2012). The model-analog technique could therefore allow for the construction of forecast ensembles based on a much larger number of CGCMs than are currently employed operationally, since it is far less computationally expensive than running a model in both initialization and forecast modes and does not require additional data assimilation or new suites of retrospective ensemble forecast integrations. For example, a 15-member NMME hindcast ensemble covering 1982–2009, with monthly initializations and forecast leads of up to one year, would be computationally equivalent to a control simulation of over 5000 years, or about 10 times what the model-analogs would require. Also, since only SSTAs and SSHAs are needed to construct the model-analogs, and monthly data are available since 1958, hindcast skill could be evaluated over a much longer period than 28 years at no additional computational cost (although the quality of SSH prior to the satellite era may be a concern).

In this paper, we have only displayed SSTA skill, although we also found similar results for SSHA. Of course, all other quantities associated with a target SSTA–SSHA state (e.g., rainfall and air temperature) can be forecasted as well, using corresponding quantities associated with the preidentified analogs. We could also test the inclusion of such variables in the analog selection metric itself. Given the results shown in this paper, further development of the model-analog technique along the lines outlined above seems promising.

REFERENCES

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231, https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

——, D. J. Vimont, P. Chang, and J. D. Scott, 2010: The impact of extratropical atmospheric variability on ENSO: Testing the seasonal footprinting mechanism using coupled model experiments. *J. Climate*, **23**, 2885–2901, https://doi.org/10.1175/2010JCLI3205.1.

Atwood, A. R., D. S. Battisti, A. T. Wittenberg, W. H. G. Roberts, and D. J. Vimont, 2017: Characterizing unforced multidecadal variability of ENSO: A case study with the GFDL CM2.1 coupled GCM. *Climate Dyn.*, **49**, 2845–2862, https://doi.org/10.1007/s00382-016-3477-9.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quart. J. Roy. Meteor. Soc.*, **139**, 1132–1161, https://doi.org/10.1002/qj.2063.

Barnett, T. P., and R. W. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci.*, **35**, 1771–1787, https://doi.org/10.1175/1520-0469(1978)035<1771:MAPOST>2.0.CO;2.

Barnston, A. G., and R. E. Livezey, 1989: An operational multifield analog/anti-analog prediction system for United States seasonal temperatures. Part II: Spring, summer, fall and intermediate 3-month period experiments. *J. Climate*, **2**, 513–541, https://doi.org/10.1175/1520-0442(1989)002<0513:AOMAAP>2.0.CO;2.

——, and M. K. Tippett, 2017: Do statistical pattern corrections improve seasonal climate predictions in the North American Multimodel Ensemble models? *J. Climate*, **30**, 8335–8355, https://doi.org/10.1175/JCLI-D-17-0054.1.

——, ——, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, https://doi.org/10.1175/BAMS-D-11-00111.1.

——, ——, H. M. Van den Dool, and D. A. Unger, 2015: Toward an improved multimodel ENSO prediction. *J. Appl. Meteor. Climatol.*, **54**, 1579–1595, https://doi.org/10.1175/JAMC-D-14-0188.1.

Berliner, L. M., C. K. Wikle, and N. Cressie, 2000: Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate*, **13**, 3953–3968, https://doi.org/10.1175/1520-0442(2001)013<3953:LLPOPS>2.0.CO;2.

Boer, G. J., 2000: A study of atmosphere-ocean predictability on long time scales. *Climate Dyn.*, **16**, 469–477, https://doi.org/10.1007/s003820050340.

Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827–1846, https://doi.org/10.1175/JCLI-D-11-00227.1.

Cane, M. A., 1984: Modeling sea level during El Niño. *J. Phys. Oceanogr.*, **14**, 1864–1874, https://doi.org/10.1175/1520-0485(1984)014<1864:MSLDEN>2.0.CO;2.

Capotondi, A., and Coauthors, 2015: Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, **96**, 921–938, https://doi.org/10.1175/BAMS-D-13-00117.1.

Carrassi, A., V. Guemas, F. J. Doblas-Reyes, D. Volpi, and M. Asif, 2016: Sources of skill in near-term climate prediction: Generating initial conditions. *Climate Dyn.*, **47**, 3693–3712, https://doi.org/10.1007/s00382-016-3036-4.

Chang, P., L. Zhang, R. Saravanan, D. J. Vimont, J. C. H. Chiang, L. Ji, H. Seidel, and M. K. Tippett, 2007: Pacific meridional mode and El Niño–Southern Oscillation. *Geophys. Res. Lett.*, **34**, L16608, https://doi.org/10.1029/2007GL030302.

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

Chiang, J. C. H., and D. J. Vimont, 2004: Analogous Pacific and Atlantic meridional modes of tropical atmosphere–ocean variability. *J. Climate*, **17**, 4143–4158, https://doi.org/10.1175/JCLI4953.1.

Collins, M., D. Frame, B. Sinha, and C. Wilson, 2002: How far ahead could we predict El Niño? *Geophys. Res. Lett.*, **29**, 1492, https://doi.org/10.1029/2001GL013919.

DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks. *Mon. Wea. Rev.*, **144**, 615–626, https://doi.org/10.1175/MWR-D-15-0218.1.

——, M. Zhao, P. A. Dirmeyer, and B. P. Kirtman, 2008: Empirical correction of a coupled land–atmosphere model. *Mon. Wea. Rev.*, **136**, 4063–4076, https://doi.org/10.1175/2008MWR2344.1.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, https://doi.org/10.1175/JCLI3629.1.

Deser, C., L. Terray, and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237–2258, https://doi.org/10.1175/JCLI-D-15-0304.1.

Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev.: Climate Change*, **4**, 245–268, https://doi.org/10.1002/wcc.217.

Fraedrich, K., C. C. Raible, and F. Sielmann, 2003: Analog ensemble forecasts of tropical cyclone tracks in the Australian region. *Wea. Forecasting*, **18**, 3–11, https://doi.org/10.1175/1520-0434(2003)018<0003:AEFOTC>2.0.CO;2.

Gill, A. E., and A. J. Clarke, 1974: Wind-induced upwelling, coastal currents and sea-level changes. *Deep-Sea Res. Oceanogr. Abstr.*, **21**, 325–345, https://doi.org/10.1016/0011-7471(74)90038-2.

Goodwin, I. D., and Coauthors, 2014: A reconstruction of extra-tropical Indo-Pacific sea-level pressure patterns during the Medieval Climate Anomaly. *Climate Dyn.*, **43**, 1197–1219, https://doi.org/10.1007/s00382-013-1899-1.

Graham, N. E., and Coauthors, 2007: Tropical Pacific–mid-latitude teleconnections in medieval times. *Climatic Change*, **83**, 241–285, https://doi.org/10.1007/s10584-007-9239-2.

Gutzler, D. S., and J. Shukla, 1984: Analogs in the wintertime 500 mb height field. *J. Atmos. Sci.*, **41**, 177–189, https://doi.org/10.1175/1520-0469(1984)041<0177:AITWMH>2.0.CO;2.

Hazeleger, W., V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian, 2013: Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett.*, **40**, 1794–1798, https://doi.org/10.1002/grl.50355.

Ji, M., and A. Leetmaa, 1997: Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev.*, **125**, 742–753, https://doi.org/10.1175/1520-0493(1997)125<0742:IODAOO>2.0.CO;2.

Jin, E. K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dyn.*, **31**, 647–664, https://doi.org/10.1007/s00382-008-0397-3.

Joseph, R., and S. Nigam, 2006: ENSO evolution and teleconnections in IPCC's twentieth-century climate simulations: Realistic representation? *J. Climate*, **19**, 4360–4377, https://doi.org/10.1175/JCLI3846.1.

Judd, K., L. Smith, and A. Weisheimer, 2004: Gradient free descent: Shadowing, and state estimation using limited derivative information. *Physica D*, **190**, 153–166, https://doi.org/10.1016/j.physd.2003.10.011.

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

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Keenlyside, N., M. Latif, M. Botzet, J. Jungclaus, and U. Schulzweida, 2005: A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature. *Tellus,* **57A**, 340–356, https://doi.org/10.1111/j.1600-0870.2005.00107.x.

Kirtman, B. P., and D. Min, 2009: Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon. Wea. Rev.*, **137**, 2908–2930, https://doi.org/10.1175/2009MWR2672.1.

——, and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Kleeman, R., A. M. Moore, and N. R. Smith, 1995: Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean–atmosphere forecast model. *Mon. Wea. Rev.*, **123**, 3103–3114, https://doi.org/10.1175/1520-0493(1995)123<3103:AOSTDI>2.0.CO;2.

Klein, S. A., B. J. Soden, and N.-C. Lau, 1999: Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge. *J. Climate*, **12**, 917–932, https://doi.org/10.1175/1520-0442(1999)012<0917:RSSTVD>2.0.CO;2.

Kruizinga, S., and A. H. Murphy, 1983: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. *Mon. Wea. Rev.*, **111**, 2244–2254, https://doi.org/10.1175/1520-0493(1983)111<2244:UOAAPT>2.0.CO;2.

Kug, J.-S., J. Choi, S.-I. An, F.-F. Jin, and A. T. Wittenberg, 2010: Warm pool and cold tongue El Niño events as simulated by the GFDL 2.1 coupled GCM. *J. Climate*, **23**, 1226–1239, https://doi.org/10.1175/2009JCLI3293.1.

Latif, M., and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103**, 14 375–14 393, https://doi.org/10.1029/97JC03413.

Lehner, F., C. Deser, and L. Terray, 2017: Toward a new estimate of "time of emergence" of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *J. Climate*, **30**, 7739–7756, https://doi.org/10.1175/JCLI-D-16-0792.1.

Lguensat, R., P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, 2017: The analog data assimilation. *Mon. Wea. Rev.*, **145**, 4093–4107, https://doi.org/10.1175/MWR-D-16-0441.1.

Li, G., and S.-P. Xie, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765–1780, https://doi.org/10.1175/JCLI-D-13-00337.1.

Livezey, R. E., and A. G. Barnston, 1988: An operational multifield analog/antianalog prediction system for United States seasonal temperatures: 1. System design and winter experiments. *J. Geophys. Res.*, **93**, 10 953–10 974, https://doi.org/10.1029/JD093iD09p10953.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

McDermott, P. L., and C. K. Wikle, 2016: A model-based approach for analog spatio-temporal dynamic forecasting. *Environmetrics*, **27**, 70–82, https://doi.org/10.1002/env.2374.

——, ——, and J. Millspaugh, 2018: A hierarchical spatiotemporal analog forecasting model for count data. *Ecol. Evol.*, **8**, 790–800, https://doi.org/10.1002/ece3.3621.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, https://doi.org/10.1175/BAMS-88-9-1383.

——, and Coauthors, 2014: Decadal climate prediction: An update from the trenches. *Bull. Amer. Meteor. Soc.*, **95**, 243–267, https://doi.org/10.1175/BAMS-D-12-00241.1.

Mulholland, D. P., P. Laloyaux, K. Haines, and M. A. Balmaseda, 2015: Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Mon. Wea. Rev.*, **143**, 4631–4644, https://doi.org/10.1175/MWR-D-15-0076.1.

——, K. Haines, and M. A. Balmaseda, 2016: Improving seasonal forecasting through tropical ocean bias corrections. *Quart. J. Roy. Meteor. Soc.*, **142**, 2797–2807, https://doi.org/10.1002/qj.2869.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600, https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, 1998: ENSO theory. *J. Geophys. Res.*, **103**, 14 261–14 290, https://doi.org/10.1029/97JC03424.

Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.*, **44**, 8520–8529, https://doi.org/10.1002/2017GL074088.

——, M. A. Alexander, and J. D. Scott, 2011: An empirical model of tropical ocean dynamics. *Climate Dyn.*, **37**, 1823–1841, https://doi.org/10.1007/s00382-011-1034-0.

Nicholls, N., 1980: Long-range weather forecasting: Value, status, and prospects. *Rev. Geophys.*, **18**, 771–788, https://doi.org/10.1029/RG018i004p00771.

Nicholson, S. E., 1997: An analysis of the ENSO signal in the tropical Atlantic and western Indian Oceans. *Int. J. Climatol.*, **17**, 345–375, https://doi.org/10.1002/(SICI)1097-0088(19970330)17:4<345::AID-JOC127>3.0.CO;2-3.

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

Overpeck, J. T., R. S. Webb, and T. Webb III, 1992: Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology*, **20**, 1071–1074, https://doi.org/10.1130/0091-7613(1992)020<1071:MENAVC>2.3.CO;2.

Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024, https://doi.org/10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.

Power, S., M. Haylock, R. Colman, and X. Wang, 2006: The predictability of interdecadal changes in ENSO activity and ENSO teleconnections. *J. Climate*, **19**, 4755–4771, https://doi.org/10.1175/JCLI3868.1.

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

Rosati, A., K. Miyakoda, and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.*, **125**, 754–772, https://doi.org/10.1175/1520-0493(1997)125<0754:TIOOIC>2.0.CO;2.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, https://doi.org/10.1175/JCLI3812.1.

——, and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

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

Smith, L. A., 2001: Disentangling uncertainty and error: On the predictability of nonlinear systems. *Nonlinear Dynamics and Statistics*, A. I. Mees, Ed., Springer, 31–64.

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

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

——, D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373, https://doi.org/10.1038/32861.

——, and Coauthors, 2011: ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dyn.*, **37**, 455–471, https://doi.org/10.1007/s00382-010-0947-3.

Suarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. *J. Atmos. Sci.*, **45**, 3283–3287, https://doi.org/10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2.

Sugihara, G., and R. M. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, 734–741, https://doi.org/10.1038/344734a0.

Takahashi, K., A. Montecinos, K. Goubanova, and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophys. Res. Lett.*, **38**, L10704, https://doi.org/10.1029/2011GL047364.

Takens, F., 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds., Vol. 898, Lecture Notes in Mathematics, Springer, 366–381.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tippett, M. K., and T. DelSole, 2013: Constructed analogs and linear regression. *Mon. Wea. Rev.*, **141**, 2519–2525, https://doi.org/10.1175/MWR-D-12-00223.1.

——, M. Ranganathan, M. L. L'Heureux, A. G. Barnston, and T. DelSole, 2018: Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dyn.*, https://doi.org/10.1007/s00382-017-3721-y, in press.

Van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247, https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

——, 1994: Searching for analogues, how long must we wait? *Tellus*, **46A**, 314–324, https://doi.org/10.3402/tellusa.v46i3.15481.

——, J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *J. Geophys. Res.*, **108**, 8617, https://doi.org/10.1029/2002JD003114.

Venzke, S., M. Latif, and A. Villwock, 2000: The coupled GCM ECHO-2. Part II: Indian Ocean response to ENSO. *J. Climate*, **13**, 1371–1383, https://doi.org/10.1175/1520-0442(2000)013<1371:TCGE>2.0.CO;2.

Vimont, D. J., J. M. Wallace, and D. S. Battisti, 2003: The seasonal footprinting mechanism in the Pacific: Implications for ENSO. *J. Climate*, **16**, 2668–2675, https://doi.org/10.1175/1520-0442(2003)016<2668:TSFMIT>2.0.CO;2.

Volpi, D., V. Guemas, and F. J. Doblas-Reyes, 2017: Comparison of full field and anomaly initialisation for decadal climate prediction: Towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. *Climate Dyn.*, **49**, 1181–1195, https://doi.org/10.1007/s00382-016-3373-3.

Von Storch, H., and F. W. Zwiers, 2002: *Statistical Analysis in Climate Research.* Cambridge University Press, 496 pp.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, https://doi.org/10.1175/MWR3280.1.

Wetterhall, F., S. Halldin, and C.-Y. Xu, 2005: Statistical precipitation downscaling in central Sweden with the analogue method. *J. Hydrol.*, **306**, 174–190, https://doi.org/10.1016/j.jhydrol.2004.09.008.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, L12702, https://doi.org/10.1029/2009GL038710.

——, A. Rosati, N.-C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate*, **19**, 698–722, https://doi.org/10.1175/JCLI3631.1.

——, ——, T. L. Delworth, G. A. Vecchi, and F. Zeng, 2014: ENSO modulation: Is it decadally predictable? *J. Climate*, **27**, 2667–2681, https://doi.org/10.1175/JCLI-D-13-00577.1.

Yamagata, T., S. K. Behera, J.-J. Luo, S. Masson, M. R. Jury, and S. A. Rao, 2004: Coupled ocean-atmosphere variability in the tropical Indian Ocean. *Earth's Climate*, *Geophys. Monogr.*, Vol. 147, Amer. Geophys. Union, 189–211, https://doi.org/10.1029/147GM12.

Ye, H., R. J. Beamish, S. M. Glaser, S. C. H. Grant, C.-H. Hsieh, L. J. Richards, J. T. Schnute, and G. Sugihara, 2015: Equation-free ecosystem forecasting. *Proc. Natl. Acad. Sci. USA*, **112**, E1569–E1576, https://doi.org/10.1073/pnas.1417063112.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278, https://doi.org/10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2.

Zhao, Z., and D. Giannakis, 2016: Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, **29**, 2888–2939, https://doi.org/10.1088/0951-7715/29/9/2888.

Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.