Diversity, Nonlinearity, Seasonality, and Memory Effect in ENSO Simulation and Prediction Using Empirical Model Reduction

CHEN CHEN, MARK A. CANE, NAOMI HENDERSON, DONG EUN LEE, AND DAVID CHAPMAN

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

DMITRI KONDRASHOV AND MICKAËL D. CHEKROUN

Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California

(Manuscript received 26 May 2015, in final form 16 December 2015)

ABSTRACT

A suite of empirical model experiments under the empirical model reduction framework are conducted to advance the understanding of ENSO diversity, nonlinearity, seasonality, and the memory effect in the simulation and prediction of tropical Pacific sea surface temperature (SST) anomalies. The model training and evaluation are carried out using 4000-yr preindustrial control simulation data from the coupled model GFDL CM2.1. The results show that multivariate models with tropical Pacific subsurface information and multilevel models with SST history information both improve the prediction skill dramatically. These two types of models represent the ENSO memory effect based on either the recharge oscillator or the time-delayed oscillator viewpoint. Multilevel SST models are a bit more efficient, requiring fewer model coefficients. Nonlinearity is found necessary to reproduce the ENSO diversity feature for extreme events. The nonlinear models reconstruct the skewed probability density function of SST anomalies and improve the prediction of the skewed amplitude, though the role of nonlinearity may be slightly overestimated given the strong nonlinear ENSO in GFDL CM2.1. The models with periodic terms reproduce the SST seasonal phase locking but do not improve the prediction appreciably. The models with multiple ingredients capture several ENSO characteristics simultaneously and exhibit overall better prediction skill for more diverse target patterns. In particular, they alleviate the spring/autumn prediction barrier and reduce the tendency for predicted values to lag the target month value.

1. Introduction

In the modeling hierarchy ranging from simplified conceptual models, empirical/statistical models, intermediate coupled models, and hybrid coupled models to the fully coupled general circulation models, empirical models not only assist in investigating ENSO behavior but are also skillful as prediction models. As data-based methods, they were popular before the arrival of fully coupled numerical model predictions and are still useful in the new era of "big data."

Advancing ENSO modeling depends on the understanding of ENSO behavior as well as the behavior of the models themselves. In this study, we carry out a

E-mail: cchen@ldeo.columbia.edu

DOI: 10.1175/JCLI-D-15-0372.1

series of empirical model experiments to study the important ENSO characteristics of diversity, nonlinearity, seasonality, and memory effect. We also address prediction problems, including the well-known "spring barrier" and the less known "slippage": that is, the tendency for predictions to incorporate too much persistence and thus lag behind the actual target month observations (Barnston et al. 2012).

We conduct the analysis using the empirical model reduction (EMR) framework (Kravtsov et al. 2005; Kondrashov et al. 2005; Kondrashov et al. 2015), which allows model settings to include additional quadratic terms, periodic terms, and multilevel predictors. The operational prediction version of EMR [University of California, Los Angeles, Theoretical Climate Dynamics (UCLA-TCD) model] participates in the real-time "ENSO prediction plume" of the International Research Institute for Climate and Society (IRI). Among eight empirical models in the plume, the UCLA-TCD

Corresponding author address: Chen Chen, Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, NY 10964.

model is very competitive, as measured by a 9-yr (2002– 11) real-time prediction and a 30-yr (1981–2010) hindcast, exceeded by only a few dynamical models. It also better copes with the spring barrier and target month slippage (Barnston et al. 2012).

The first task of this study is to decompose EMR in order to identify which ingredients contribute most to its good simulation and prediction skill. To do so, we construct a series of models with various settings, including adding additional quadratic terms, periodic terms, and multilevel and multivariate predictors. We then evaluate them in control groups (e.g., with or without nonlinearity, with or without seasonality, with or without memory, and multivariate or multilevel approach). Through panel comparisons of each model's ability to reproduce ENSO characteristics and its predictive ability, we show that deficiencies of the baseline settings are largely remedied in more complex models. Not surprisingly, we find that even our most "advanced" model has limitations.

Along with the general evaluation, we hope to contribute to the following aspects of ENSO or empirical modeling research. The first is to better understand the differences between two representations of the memory effect. The intrinsic predictability of ENSO largely comes from memory stored in the subsurface ocean, which may be described as a recharge oscillator (Cane et al. 1986; Zebiak and Cane 1987; Jin 1997a,b) or as a time-delayed oscillator (Suarez and Schopf 1988; Battisti and Hirst 1989). The approach of adding subsurface information [e.g., the 20°C isotherm depth (Z20)] as an additional variable is a natural extension in the linear inverse model (LIM) framework (e.g., Blumenthal 1991; Xue et al. 1994, 1997, 2000; Johnson et al. 2000a; Thompson and Battisti 2001; Newman et al. 2011a,b), while the approach of EMR fitting multiple levels in an SST time-delayed model is less common. We construct models of each type to illustrate their relation and make a comparison. Considering these two approaches together offers more choices to embed the memory in the model setting.

A second aspect of our work is to evaluate model performance on ENSO diversity (Capotondi et al. 2015). El Niños occur not only in the eastern Pacific but also in the central Pacific (Larkin and Harrison 2005; Ashok et al. 2007; Kao and Yu 2009; Kug et al. 2009; Di Lorenzo et al. 2010; Lee and McPhaden 2010; Karnauskas 2013; Vimont et al. 2014; Chen et al. 2015; Takahashi and Dewitte 2015; Fedorov et al. 2015). Most model evaluations use one-dimensional (1D) measures [e.g., Niño-3.4 or the leading principal component (PC1) of tropical Pacific SST anomalies (SSTAs)], but a two-dimensional (2D) framework is necessary to describe ENSO's spatial variation. Takahashi et al. (2011) showed various measures of ENSO diversity in the literature could be viewed as linear combinations of tropical Pacific SSTA PC1 and PC2. Following Takahashi et al. (2011) and Vimont et al. (2014), we introduce a measure using the PC1–PC2 pair for overall ENSO diversity performance. Takahashi et al. (2011) suggested that the curved shape in PC1-PC2 space implies that the two flavors of El Niño may not describe different phenomena, but rather the nonlinear evolution of ENSO. The EMR framework does not explicitly specify coefficients to characterize ENSO flavors, so model performance on ENSO diversity is implicitly determined by model dynamics. We conduct stochastically forced simulations of various model combinations and find that nonlinearity is necessary to reproduce the curved shape in PC1-PC2 space, supporting Takahashi et al.'s (2011) suggestion that ENSO diversity may emerge from nonlinear evolution.

ENSO nonlinearity is reflected in the El Niño–La Niña (EN–LN) asymmetry (e.g., Hoerling et al. 1997; Kang and Kug 2002; Larkin and Harrison 2002; An and Jin 2004; Okumura and Deser 2010; Choi et al. 2013). Given these nonlinear features of ENSO, one would expect that including nonlinearity in the empirical model will improve simulation and prediction. Here we find that the nonlinear model does not apparently improve simulation and prediction of Niño-3.4, but it does improve the skewed amplitude and the transition patterns from El Niño to La Niña.

We also investigate the spring barrier in ENSO prediction. Prediction skill in Niño-3.4 (similar to SST PC1) drops if initialized from the boreal spring but recovers after spring in both dynamic and statistical ENSO forecasts (Barnston et al. 2012; Barnston and Tippett 2013; Xue et al. 2013). There is also an "autumn barrier" in the prediction of warm water volume (similar to either Z20 PC2 or SST PC2) (McPhaden 2003). One would expect a model with seasonal-varying dynamics to improve ENSO prediction across the common spring barrier. We find that the seasonal setting does not contribute much to reducing the seasonal barrier until combined with the memory effect.

Another goal of our work is to gain more knowledge of the slippage problem in prediction. Recent studies (e.g., Tippett et al. 2012; Barnston et al. 2012; Barnston and Tippett 2013) found that both dynamic and statistical ENSO prediction models characteristically produce a delay to the target month. It is as if the model has too much persistence. The UCLA-TCD model is among the models with little slippage, and here we investigate which model ingredients contribute most to CHEN ET AL.

Limited observational data often bring in large uncertainties in model training and performance evaluation. To construct robust models and reduce evaluation uncertainty, we use a long record (4000 yr) of simulated data and conduct ensemble experiments to estimate uncertainty in training and verifying periods.

2. Data

This study uses monthly data from a 4000 year preindustrial simulation of the GFDL CM2.1, a coupled GCM at $1^{\circ} \times 0.33^{\circ}$ - 1° resolution (Delworth et al. 2006). This simulation has been shown to have a reasonably realistic ENSO (e.g., Wittenberg et al. 2006; Kug et al. 2010; Choi et al. 2013; Wittenberg et al. 2014), although its amplitude is too large (Wittenberg 2009), and it may be less predictable than the real ENSO (Karamperidou et al. 2014).

To justify using a model simulation as a substitute for observations, we present comparisons of several main ENSO characteristics between the GFDL CM2.1 simulation and the observational dataset of HadISST, version 1.1 (HadISST1.1), monthly $(1^{\circ} \times 1^{\circ})$ SSTs (Rayner et al. 2003) from 1870 to the present. We also compared with three other observational SST datasets: Centennial In Situ Observation-Based Estimates version 2 (COBEv2) (Hirahara et al. 2014), Kaplan (Kaplan et al. 1998), and ERSST.v3b (Smith et al. 2008). Results with these three are nearly identical to those with HadISST1.1. For simplicity, the observations are referred to as OBS, and the simulation is referred to as GCM here.

All data are preprocessed as follows. SST data are restricted to the tropical Pacific domain (30°S–30°N, 108°E–72°W). SSTA in OBS are calculated by removing a monthly climatology based on the 1950–2010 period, while SSTAs in GCM are calculated by removing a monthly climatology based on the entire 4000-yr record. To remove the influence of the global warming trend in OBS and model drift in GCM, linear detrending is applied to the SSTA at each grid point. Then a 3-month running average is applied to smooth the temporal noise. In addition to SST, we also use Z20 in the tropical Pacific (20°S–20°N, 108°E–72°W), as a proxy for thermocline depth to incorporate subsurface information. The processing of Z20 is similar to SST.

SST and Z20 variability are then decomposed using empirical orthogonal function (EOF) analysis. Figures 1a-d show the leading two EOF patterns of SSTA. In both OBS and GCM, the leading EOF (EOF1) is the familiar El Niño pattern. The second EOF (EOF2), which adds flavor to the main El Niño pattern, has a zonal dipole pattern with positive loading in the western Pacific when the loading in the eastern Pacific is negative. The first and second modes explain 50% and 8% of the total variance in OBS and 52% and 11% in GCM. EOF patterns from GCM are reasonably realistic and are generally consistent with observations, although slightly shifted west and narrower in the meridional direction, as noted by Wittenberg et al. (2006). Besides the EOF patterns (Fig. 1), the results show good consistency between observation and simulation in the sign of the skewed distribution, seasonal standard deviation, and lagged autocorrelation. The joint probability distribution of (PC1, PC2) is also reasonably consistent between OBS and GCM (see Figs. 3a and 3b). Additional discussion is given in section 5.

These comparisons suggest that ENSO characteristics in GFDL CM2.1 are reasonably realistic, so the understanding from the modeling study based on the GFDL CM2.1 may be applicable to nature. But the nonlinearity in the GFDL CM2.1 is much stronger than in nature, so the nonlinear setting may not be as helpful in modeling nature.

3. Modeling methods

a. Baseline model

Linear inverse models are widely used empirical modeling techniques for ENSO, in which the full dynamics of the tropical Pacific SST variability from seasonal-to-interannual time scales is approximated as a linear system of ordinary differential equations (e.g., Penland 1989; Blumenthal 1991; Penland and Magorian 1993; Penland and Sardeshmukh 1995; Penland 1996; Xue et al. 1994, 1997, 2000; Johnson et al. 2000a; Thompson and Battisti 2001; Compo and Sardeshmukh 2010; Newman et al. 2011a,b). LIMs have been shown successful in predicting ENSO (e.g., Penland and Sardeshmukh 1995) and tropical Atlantic variability (Penland and Matrosova 1998) as well as extratropical variability (Alexander et al. 2008). In the LIM family, features like seasonality and memory effect may be implemented in many possible ways.

In the EMR framework, we construct linear and nonlinear models, nonseasonal and seasonal models, and models with and without memory effects. In this study, a nonseasonal SST linear model without memory effect is considered as the baseline model. By construction, the baseline model lacks nonlinearity and seasonality and assumes the next step SSTA only depends on the current state of SSTA, ignoring memory



FIG. 1. Comparison of ENSO statistics in SST observations (1871–2013) and GFDL CM2.1. (a)–(d) EOF1 and EOF2 of tropical Pacific SST anomalies. Shading is regression coefficient of tropical Pacific SST anomalies on PC1 and PC2. Contour is the percentage of explained variance of PC1 and PC2. (e)–(h) ENSO asymmetry and nonlinearity represented in the PDFs of PC1 and PC2. The skewness averaged among four observation datasets and GFDL CM2.1 segments (SKm) is shown. (i)–(l) ENSO seasonality is represented in STD varying as to each calendar month. (m)–(p) ENSO memory and seasonal break of persistence are represented in lagged Corrcoeff. Only HadISST1.1 results are shown in (a)–(d),(m)–(p). The four observation datasets in (e),(f),(i),(j) are color-coded: HadISST1.1 (black), COBEv2 (green), ERSST.v3b (blue), and Kaplan (red). The 4000-yr preindustrial run of GFDL CM2.1 is divided into 20 mutually exclusive 200-yr segments for calculations in (g),(h),(k),(l). Each segment is in gray, and averaged result is in black. The x axis is the initial calendar month, and y axis is the lag month.

effects from subsurface processes or from the past history of SSTA.

b. Model with nonlinearity

Nonlinear features of ENSO include the asymmetry between El Niño and La Niña in amplitude (El Niño is often stronger than La Niña), duration (La Niña is often more persistent), transition (an extreme El Niño is more often followed by a La Niña than vice versa), and teleconnections (e.g., Hoerling et al. 1997; Kang and Kug 2002; Larkin and Harrison 2002; An and Jin 2004; Okumura and Deser 2010; Choi et al. 2013).

There are many possible ways to construct a nonlinear model [e.g., multiple nonparametric regression (Timmermann et al. 2001) or neural networks (Tangang et al. 1998)]. In the polynomial modeling framework of EMR, one straightforward way is to add quadratic terms, which has been shown to successfully reproduce the skewed probability density function (PDF) characteristics of the observed ENSO (Kravtsov et al. 2005; Kondrashov et al. 2005).

Taking this approach, the empirical model is constructed in a reduced phase space of K EOFs. For the main level,

$$dx_i = (\mathbf{x}^{\mathrm{T}} \mathbf{A}_i \mathbf{x} + \mathbf{b}_i \mathbf{x} + c_i)dt + dr_i; \quad i = 1, \dots, K, \quad (1)$$

where $\mathbf{x} = \{x_i\}$ is the *K* dimensional state vector (i.e., PCs). The quadratic terms on the right-hand side are

dropped for linear models. The model coefficients in the matrices \mathbf{A}_i , the vectors \mathbf{b}_i of matrix \mathbf{B} , the components c_i of vector **c** and the components r_i of the residual **r** are determined by multiple polynomial regression (MPR) (McCullagh and Nelder 1989), which is the generalized version of multiple linear regression (MLR) (Wetherill 1986). The noise in the climate system usually has spatial dependence, as addressed in Penland and Sardeshmukh (1995). Therefore the noise covariance matrix, implicitly embedding the spatial coherent information, is estimated along with other deterministic coefficients and further used to generate spatially coherent white noise for the stochastically forced simulations. For the nonlinear EMR models, nonzero c accounts for the mean nonlinear drift (Kondrashov et al. 2011). For consistency the c coefficients are retained in the linear model as well, but their estimated values are negligible, as expected for a dataset of anomalies around a zero time mean and without a trend.

c. Model with seasonality

Another notable feature of ENSO is seasonal phase locking, the tendency of El Niño and La Niña to peak toward the end of the calendar year, which is the result of competing coupling feedbacks (Tziperman et al. 1995, 1997, 1998; Neelin et al. 2000; An and Wang 2001). There are several possible ways to introduce seasonality in a model. One way is to fit a varying linear propagator for each calendar month (e.g., Blumenthal 1991; Xue et al. 2000). Another is to treat seasonality as periodic terms, as in EMR (Kravtsov et al. 2005; Kondrashov et al. 2005), with consideration of the interaction between the annual cycle and variability as in Tziperman et al. (1995).

Following the EMR framework, we include seasonality by adding additional coefficients into the main level of the model:

$$\mathbf{B} = \mathbf{B}_n + \mathbf{B}_s \sin(2\pi t/T) + \mathbf{B}_c \cos(2\pi t/T) \text{ and } (2)$$

$$\mathbf{c} = \mathbf{c}_n + \mathbf{c}_s \sin(2\pi t/T) + \mathbf{c}_c \cos(2\pi t/T), \qquad (3)$$

where the matrix \mathbf{B}_n and vector \mathbf{c}_n are the original nonseasonal terms as in Eq. (1), matrices \mathbf{B}_s and \mathbf{B}_c and vectors \mathbf{c}_s and \mathbf{c}_c are seasonal terms and for the period T = 12 months. All these coefficients are determined simultaneously with the other coefficients in the main level. Note that the seasonal dependence is allowed for the linear part of the model on the main level, but not for the residual.

d. Model with memory effect

ENSO is often viewed as a recharge oscillator where the system memory is stored in the subsurface ocean that reemerges through the thermocline feedback in the eastern Pacific SST variability (Cane et al. 1986; Zebiak and Cane 1987; Jin 1997a,b). It could be also viewed as a time-delayed oscillator where the subsurface ocean memory is embedded in the SST history (Suarez and Schopf 1988; Battisti and Hirst 1989).

These theories suggest two alternative approaches that could be used to add memory to the models. A natural extension directly based on the recharge oscillator viewpoint is to extend the SST-only state vector to a multivariate state vector by adding the leading PCs of ocean heat content, 20°C isotherm depth, or sea level variation in the tropical Pacific (e.g., Blumenthal 1991; Xue et al. 1994, 1997, 2000; Johnson et al. 2000a; Thompson and Battisti 2001; Newman et al. 2011a,b). A direct representation from the time-delayed oscillator viewpoint is to use only SST PCs but fit a multiple timelevel model using previous time steps rather than just the current time step. The multilevel fit could be executed at once (Chapman et al. 2015; Lee et al. 2015) or conducted recursively (Kondrashov et al. 2005; Kravtsov et al. 2005).

In this study, we construct models in both multilevel and multivariate settings to compare these two approaches. Following Kondrashov et al. (2005), we add the memory effect via multiple levels but confine nonlinearity and seasonality to the main (first) level. Adding an additional level, the temporal increment of the residual at the main level $d\mathbf{r}^1$ is further modeled as a linear function of an extended state vector $[\mathbf{x}, \mathbf{r}^1] \equiv [\mathbf{x}^T, (\mathbf{r}^1)^T]^T$. More levels are added in the same way, until the *L*th level's residual \mathbf{r}^{L+1} becomes uncorrelated in time with a lag-0 correlation matrix that converges to a constant matrix (details are given in appendix A of Kondrashov et al. 2015):

$$dx_{i} = (\mathbf{x}^{\mathrm{T}} \mathbf{A}_{i} \mathbf{x} + \mathbf{b}_{i}^{1} \mathbf{x} + c_{i}^{1})dt + dr_{i}^{1},$$

$$dr_{i}^{1} = \mathbf{b}_{i}^{2} [\mathbf{x}, \mathbf{r}^{1}]dt + dr_{i}^{2},$$

$$dr_{i}^{2} = \mathbf{b}_{i}^{3} [\mathbf{x}, \mathbf{r}^{1}, \mathbf{r}^{2}]dt + dr_{i}^{3},$$

$$\vdots$$

$$dr_{i}^{L} = \mathbf{b}_{i}^{L+1} [\mathbf{x}, \mathbf{r}^{1}, \mathbf{r}^{2}, \dots, \mathbf{r}^{L}]dt + dr_{i}^{L+1},$$

(4)

where i = 1, ..., K and the \mathbf{b}_i^j and r_i^j for level j are determined recursively.

In addition to SST-only multilevel models, we also construct multivariate models, in which the state vector \mathbf{x} is in K = m + n dimensions with the leading m normalized SST PCs and the leading n normalized Z20 PCs given equal weighting. Sensitivity tests show that slight weighting changes do not change the conclusion.

TABLE 1. List of ENSO empirical models with different model settings: L(m) is a linear model in the form of Eq.(1) without quadratic terms, constructed from a state vector of the leading *m* PCs of tropical Pacific SSTA, m = 3 and 8 are presented; M(k) is the linear model constructed from a multivariate state vector of the leading 3 PCs of tropical Pacific SSTA and the leading *n* PCs of tropical Pacific Z20 anomalies, k = 3 + n with n = 5 and 10 presented here; 2L and 5L are a multilevel linear models with 2 and 5 time levels and 3 SST PCs; NL denotes the nonlinear models with quadratic terms; 2L+NL is a combined model with quadratic nonlinearity in the main level and linear terms in the one additional time level; S is the seasonal model with additional periodic terms; S+NL, 2L+S, and 2L+S+NL are combined models including seasonality; and 2L+S+NL is the most comprehensive model of the 12 models studied here. Coef No. denotes the total number of coefficients in the design matrix. Coefficients in the noise covariance matrix are not included. Condition No. denotes the condition number of design matrix for each level. See the appendix of Kravtsov et al. (2005) for more details.

No.	Label	SST	Z20	Level	Nonlinear	Seasonal	Coef No.	Condition No.
1	L(3)	3	_	1			12	1.11
2	L(8)	8	_	1			72	1.26
3	M(8)	3	5	1			72	5.88
4	M(13)	3	10	1			182	6.90
5	2L	3	_	2			30	1.11 and 6.81
6	5L	3	_	5			138	1.11, 6.81, 14.61, 16.13, and 16.31
7	NL	3	_	1			30	3.74
8	2L+NL	3	_	2			48	3.74 and 3.72
9	S	3	_	1		1	36	1.80
10	S+NL	3	_	1		1	54	4.61
11	2L+S	3	_	2		1	54	1.80 and 8.25
12	2L+S+NL	3	_	2	1		72	4.61 and 8.92

e. Modeling experiments

In this modeling framework, there are many possible model settings and coefficient choices: m SST PCs, n Z20 PCs, j levels, linear or nonlinear, and nonseasonal or seasonal. Since we cannot include all possible settings in this study, we offer 12 typical model settings to conceptually compare simulation and prediction performance. The 12 models are specified in Table 1.

There is evidence that ENSO is a low-dimensional system with only a few degrees of freedom (Tziperman et al. 1994). Rather than simply increasing the size of the state vector, here we assess how adding other ingredients may improve the skill. For the baseline model, we choose a linear model (L) with three SST PCs [L(3)]. Models with a large state vector often encounter overfitting problems given insufficient data; thus, fewer coefficients is a desired property. Nonlinear models are especially prone to the problem of an ill-conditioned matrix and numerical instability (Kondrashov et al. 2005). To avoid this, all the empirical models here have $K \leq 13$ PC predictors. Consequently, all the models included in this study behave well and are numerically stable without any constraint or regularization.

Besides the baseline model L(3), we include a linear model L(8) with eight SST PCs for comparison. Nonlinear models (NL), all with quadratic terms, are denoted as NL, S+NL, 2L+NL, and 2L+S+NL. Seasonal models (S), all with periodic terms, are denoted as S, 2L+S, S+NL, and 2L+S+NL. Both the nonlinear and seasonal models have three SST PCs, as for L(3). Memory effects are represented either in multivariate models (M) with 3 SST PCs and 5 or 10 Z20 PCs [M(8) and M(13)] or in SST models with multiple time levels (2L, 5L, 2L+NL, 2L+S, and 2L+S+NL). The sensitivity test shows the qualitative conclusions still hold when using a linear model with more than three SST PCs as the baseline model, although the quantitative improvement given by additional model complexity is less.

SST and Z20 PCs for all 4000 yr are obtained once, and then the data are divided into two parts for model training and evaluation as follows: 1) Model training and evaluation for goodness of fit are carried out in the first 2000 yr using 10 nonoverlapping segments of 200 yr. The model used for further simulation and prediction is obtained by averaging the models from all 10 segments. 2) Evaluation of simulation performance is made by conducting twenty 1000-yr stochastically forced simulations and then comparing simulated statistics with the GCM data statistics for the training period. 3) We carry out 1-12-month-lead out-of-sample predictions in the last 2000 yr using 10 segments of 200 yr. We carry out 20-member stochastic ensemble forecasts, and the predicted value for each initial condition is the ensemble averaged result. An ensemble size of 20 is generally enough to sample the model spread in this study.

4. Model goodness of fit

Although goodness of fit does not guarantee good outof-sample prediction skill (in section 6), it does indicate in-sample theoretical skill. We measure the model fit by applying the model on the training period and



FIG. 2. Goodness of fit measures are shown for all models in Table 1. (a),(b) The residual STD of the two leading PCs of tropical Pacific SSTA and (c),(d) the residual skewness. Models are displayed in three groups: models without memory effect, models with multivariate (SST and Z20) memory, and models with multilevel memory. The total number of model coefficients is given across the bottom. Each model carries its own color code. Error bars (one standard deviation) indicate the spread of the fit among ten 200-yr training segments. Small residual STD and residual skewness close to zero indicate a good fit.

examining the residual statistics, including residual standard deviation (STD) and residual skewness. A good model fit is characterized by small residual STD and near-zero residual skewness.

Results for all the models in Table 1 are shown in Fig. 2. The total number of coefficients is shown for each model. This count does not include elements of the noise covariance matrix. Although the constant **c** is kept in the count, it is negligible for linear models. Error bars indicate the spread of residual STD in 10 mutually exclusive training segments. The L(3) model is the baseline model with its residual STD as the benchmark. For PC1, adding seasonality (S) or quadratic terms (NL) or the combination of both (S+NL) does not reduce residual

STD much. Extending the SST state vector to eight PCs [L(8)] reduces residual STD by 20%, while extending the state vector to include Z20 PCs [M(8) and M(13)] reduces residual STD by 30%. The other way to add memory effect, by fitting more time levels of the residual (2L) and its variations (2L+NL, 2L+S, and 2L+S+NL) all show reductions of approximately 50%. More levels (5L) reduce residual STD by 60%. For PC2, adding memory through Z20 or adding levels gives a better fit than adding seasonality and nonlinearity. Adding 2 levels (2L) reduces residual STD by 35% for PC2, less than the 50% reduction for PC1. Note that including memory effect also reduces residual skewness to near zero (Figs. 2c,d).

Compared to the baseline model L(3), models adding seasonality, nonlinearity, or extending the state vector [i.e., L(8), M(8), and M(13)] require more coefficients and do show a better fit. However, compared to the multivariate model M(13), the multilevel model (2L) fits better with fewer coefficients. The comparison shows that the multilevel approach successfully captures the effects of ocean memory in terms of SST itself, a betterobserved quantity than subsurface variables.

These results show that the memory effect is the most important addition to the baseline model. From the physical perspective, a large fraction of the total variance of the tropical Pacific SST anomalies is associated with the recharge oscillation, so representing this memory effect in a data-driven model is essential. Nonlinearity and seasonality do not help much unless combined with the memory effect.

5. Results of ENSO simulation

We will first present ENSO characteristics in OBS and GCM and then compare them with the modeled statistics.

a. Nonlinearity

In both OBS and GCM, asymmetries about zero and the curved shape of the two-dimensional PDF in PC1– PC2 space indicate ENSO nonlinearity (Figs. 3a,b). The other plots in Fig. 3 show that various linear models [L(3), 2L, M(8), and S] are not able to reproduce this curved shape but have an ellipsoid shape centered at (0,0). On the contrary, the various nonlinear models (NL, 2L+NL, S+NL, and 2L+S+NL) show a curved shape in PC1–PC2 space in approximate agreement with the data. The first three models (NL, 2L+NL, and S+NL) show a small deficiency with a spike toward negative PC1 and negative PC2, but agreement is improved in the comprehensive model (2L+S+NL).

In both OBS and GCM (Fig. 1), the PC1 skewness (0.39 and 0.35, respectively) is positive, meaning warm states occur less often than cold states but with stronger amplitude in El Niño than La Niña. PC2 skewness is negative, and GCM has stronger nonlinearity than OBS with a more strongly skewed PDF in PC2 (-1.23 versus -0.40). This is associated with the fact that GCM has more extreme El Niño events. For all the models trained using GCM data, the linear models show skewness close to zero, while the four nonlinear models approximately match the skewness in GCM and resemble its 1D PDF for PC1 and PC2 (not shown).

b. Diversity

Both OBS and GCM show ENSO diversity in PC1– PC2 space, represented by the curved shape of the two-dimensional PDF (Figs. 3a,b). Takahashi et al. (2011) inspired us to distinguish various SSTA patterns using the PC1-PC2 space. The central region in PC1-PC2 space (Fig. 3) is associated with neutral patterns with small variability. The upper-right quadrant is the warming anomaly patterns occurring at the central Pacific. The lower-right quadrant is the warming anomaly patterns peaking at the eastern Pacific, and far end points are where extreme El Niño events locate. The upper-left quadrant is the cooling anomaly patterns occurring at the eastern Pacific. The lower-left quadrant is the cooling anomaly patterns peaking at the central Pacific, and far end points are where extreme La Niña events locate. The negative PC2 region is usually the transition patterns from the extreme El Niño events to the following La Niña events.

To reproduce the ENSO diversity, the model needs to not only generate anomaly patterns with realistic amplitude but also the correct location of the variability. Among the eight models in Fig. 3, the linear models do generate SSTA patterns with large variabilities, but their extreme El Niño events are not in the far eastern Pacific (lower-right region), and their extreme La Niña events are not toward the central Pacific (lower-left region). The ENSO patterns produced by the nonlinear models more closely resemble the diversity characteristics of the data.

c. Seasonality

Both OBS and GCM exhibit seasonal phase locking in monthly STDs, with PC1 showing larger variability in boreal winter and PC2 showing larger variability in boreal summer (Fig. 1). The correlation coefficient (Corrcoeff) between the monthly STD of GCM and that from a stochastically forced simulation is used as a skill measure for seasonality. Nonseasonal models produce nearly constant monthly STDs and thus a Corrcoeff near zero. Seasonal models have seasonal variations that are similar to the training data, which is consistent with Kondrashov et al. (2005). For the simplest seasonalityonly model, Corrcoeff is close to one, but adding nonlinearity or levels reduces this skill measure slightly.

d. Seasonal memory

In both OBS and GCM (Fig. 1), PC1 is not autocorrelated across boreal spring but retains a substantial correlation after spring almost until the next spring. PC2 autocorrelation (Fig. 4h) shows a similar break, but in late autumn instead of spring.

The modeling comparison (Figs. 4b–g,i–n) shows that both seasonality and memory effects are needed to mimic the lagged autocorrelation. For models without the memory effect [L(3), NL, and S], autocorrelation



FIG. 3. (a) The common logarithm of the 2D PDF for the two leading SSTA PCs in OBS. PCs are normalized to have STD equal to one. The gridcell size is 0.267×0.267 . (b) As in (a), but for GCM, the training data. (c)–(j) As in (a) but for eight models in Table 1, based on 1000-yr stochastically forced simulations in each case. Shown are the linear models in (c),(e),(g),(i) and the nonlinear models in (d),(f),(h),(j). A good simulation of ENSO nonlinearity is indicated by resemblance to the curved pattern and the PDF value distribution of the GCM data in (b).

slowly decays as a function of lag. Models with the memory effect [M(8) and 2L] better resemble the data. Nonseasonal models [L(3), M(8), 2L, and NL] all show a uniform autocorrelation across all calendar months.

Adding seasonality alone produces seasonally varying lagged autocorrelations, but without the memory effect it does not resemble the GCM autocorrelation decay as a function of lag. Thus, the model with seasonality and



FIG. 4. Lagged autocorrelation coefficients for each calendar month. (a),(h) The average of twenty 200-yr segments from the GCM data. (b)–(g),(i)–(n) Model results calculated by averaging over twenty 200-yr stochastically forced simulations. The 6 models shown are described in Table 1. Every second calendar month is labeled on the x axis.

memory effect (2L+S) shows the best overall resemblance to the data.

6. Results of ENSO prediction

In this section, the evaluation aims to identify each model ingredient's contribution to prediction skill.

a. PC1 and PC2

Our principal measures of skill are the Corrcoeff and root-mean-square error (RMSE) between target PC time series and predicted PC time series at given lead time τ . PC1 represents main variability and thus is more important than PC2, so a model with poor skill for PC1 but good skill for PC2 is not viewed as a good model.

Figure 5 shows the improvement of the 12-month forecasts as measured using the Corrcoeff for different model settings compared to persistence and the baseline model L(3). For PC1 (Fig. 5c), the seasonal model does not help much. The models with more SST PCs [L(8)] and with nonlinearity show slight improvement. Models with memory effect (both Z20 and levels) improve the most. For PC2 (Fig. 5d), the 12-month forecast Corrcoeff for persistence and L(3) are relatively high; thus, improvement using other settings is not as remarkable as for PC1. The models adding seasonality, more SST PCs, or the memory effect do not help. Models with nonlinearity show a better skill. Note that even though seasonal models do not show much improvement compared to the baseline model L(3), the model with both memory effect and seasonality (2L+S) does show slightly better skill than the memory model (2L) itself.

The RMSE results (not shown) are consistent with Corrcoeff results, showing that the memory effect efficiently reduces PC1 RMSE, and nonlinearity works best to reduce the PC2 RMSE. Models with both memory and nonlinearity have overall good performance on the PC1–PC2 pair.

Among all the models, the skill difference for PC1 is mainly between models with and without memory effect, rather than between linear and nonlinear models. The seasonal linear model with memory (2L+S) and the seasonal nonlinear model with memory (2L+S+NL) give overall the same performance for PC1. For

Prediction performance: Corrcoeff



FIG. 5. Prediction performance: correlation coefficients for PC1 and PC2 of SSTA. For each model in Table 1, 20-member ensemble forecasts are carried out for 2000 yr that are out of sample (i.e., distinct from the data used to construct the models). Correlations are between the data and the ensemble average. (a),(b) Forecast correlations at leads of 0–12 months. Persistence (black line) is given as a reference. (c),(d) Correlations of 12-month-lead forecasts. Error bars are created by dividing the forecasts into 10 nonoverlapping segments of 200 yr each and showing the standard deviation. For PC1, models with or without memory effect are grouped separately, and for PC2, linear and nonlinear models are grouped separately. The dashed line is the correlation coefficient for the benchmark model L(3). Colors for models in (a),(b) correspond to the colors in (c),(d).

different evaluation periods with fewer or more extreme ENSO events, 2L+S may have higher or lower skill than 2L+S+NL, although within the range of the error bar. Linear and nonlinear models have generally the same skill for PC1 once they include the memory effect, which is consistent with Kondrashov et al. (2005). To see this, note that in their Fig. 2c, 1-L is the same as the S+NL model in this study, and 2-L is our 2L+S+NL model. In their Fig. 6a, Linear is the 2L+S model in this study, and Nonlinear is our 2L+S+NL model. The Corrcoeff difference between Linear and Nonlinear is approximately 0.05, much smaller than the difference between 1-L and 2-L (\sim 0.2).

b. Spring barrier and autumn barrier

The spring barrier in PC1 and autumn barrier in PC2 we saw in the SSTA seasonal lagged autocorrelations (Figs. 4a,h) are reflected in model predictions (Fig. 6). All model settings inherit this barrier problem to differing degrees. The prediction skill of persistence is the benchmark for comparison. Its skill for PC1 quickly drops if initiated from spring, and its skill for PC2 drops





FIG. 6. Prediction performance: seasonal variations and the boreal spring and autumn barriers. (a),(b) For all models in Table 1, the correlation coefficient between the GCM target time series and 6-month predictions initialized at each calendar month. Persistence is shown in black. Note that seasonal barrier exists for PC1 when initialized from boreal spring (e.g., March) and for PC2 when initialized from boreal autumn (e.g., September). (c) PC1 6-month-lead prediction initialized in March. Models with or without a memory effect are grouped separately. (d) PC2 6-month-lead prediction initialized in September. Linear and nonlinear models are grouped separately. See Fig. 5 for additional descriptions.

quickly if initiated from autumn. The 6-month-lead forecast from each calendar month (Fig. 6) shows that models with memory effect [M(8), M(13), 2L, 5L, 2L+NL, 2L+S, and 2L+S+NL] significantly reduce the problem for PC1. For the PC2 autumn barrier, nonlinear models (NL, 2L+NL, S+NL, and 2L+S+NL) have better skill than other models. The seasonal model itself does not show much ability until combined with memory effect and nonlinearity.

In Barnston et al. (2012), dynamical and statistical models with various data assimilation and advanced features all show some drop in skill around spring. Following the same layout as the Fig. 5 in Barnston et al. (2012), the influence of the spring barrier is shown for each target calendar month (Fig. 7). The results show adding more features to the base model reduces the seasonal barrier, but even the most complex models in this study still show signs of the seasonal barrier, which is similar to results shown in Kondrashov et al. (2005). Also note that while the spring barrier in PC1 is the main problem in SST prediction, reducing the autumn barrier in PC2 improves the transition between El Niño and La Niña, thus increasing the total prediction skill for SST.



FIG. 7. Prediction performance for PC1 and PC2 of each target calendar month, for six models in Table 1. Corrcoeff between 1–12-month-lead predicted time series and the target time series according to each calendar month. The result shown is averaged among ten 200-yr prediction segments. Every second calendar month is labeled on the x axis. Good prediction skill is indicated by retaining large Corrcoeff toward 12-month-lead prediction as well as having a small seasonal skill drop.

c. Target month slippage

Slippage denotes the tendency of forecasts to retain initial conditions for too long (i.e., to overdo persistence). Thus, for example, a forecast intended to verify six months ahead may actually verify best at four months ahead: the prediction slipped by two months.

Following Barnston et al. (2012), we present the slippage performance by showing the lagged correlation coefficient between the target time series and the predicted time series (Fig. 8). Persistence is shown as a reference, and large slippage is indicated by the region with high Corrcoeff (>0.8) pointing to the right rather than being vertical. The comparison among all models indicates that the memory effect [M(8), 2L, and 2L+S+NL] substantially corrects the tilt for PC1, while memory and nonlinearity (NL, 2L, and 2L+S+NL) slightly correct the tilt for PC2.

We then identify the degree of slippage using the lag that gives the maximum Corrcoeff and plot this slippage at each lead time (Fig. 9). Persistence, which is used as a benchmark, necessarily slips τ months for a τ -month prediction. For PC1, the slippage of each model slowly increases as the lead time increases. For a 6-month prediction, L(3) slips 5 months for PC1 and 6 months for PC2. Models with the memory effect reduce the PC1 slippage to approximately 2 months. Nonlinear models reduce the PC2 slippage to approximately 3 months.

d. ENSO diversity

Model ability to predict ENSO diversity is evaluated in PC1–PC2 space (Fig. 10). We introduce a measure D_2 defined as the Euclidean distance between the target PC1–PC2 pair (x_1 , x_2) and the predicted PC1–PC2 pair [$y_1(\tau)$, $y_2(\tau)$] at a τ -month lead:

$$D_2(\tau) = \{ [x_1 - y_1(\tau)]^2 + [x_2 - y_2(\tau)]^2 \}^{1/2}.$$

We show the D_2 values at leads of 3, 6, and 12 months in Fig. 10. The value in a given grid cell represents the model skill for the target SSTA patterns that projected to this given grid cell. The central points of PC1–PC2 are associated with neutral patterns with small variability. The lower-right corner is the patterns of strong El Niño events, and the lower-left corner is the patterns of strong





FIG. 8. Prediction performance: slippage as measured by the correlation coefficients between 0–12-month-lead predictions and from -12- to 12-month-lagged target GCM data. (c)–(n) Results for six of the models in Table 1. (a),(b) Persistence, shown as a reference, necessarily has a τ -month lag to the target month for a τ -month-lead prediction. Good prediction performance is indicated by large correlation coefficients for the target month (i.e., along the vertical line at zero lag). Less slippage is indicated by reduced tilt with time of the maximum correlation coefficient. See Fig. 5 for additional descriptions of prediction methodology.

La Niña events. Given D_2 measuring the absolute errors rather than the errors normalized by the amplitude of the variability, large errors are associated with large variability.

Comparison of different models shows that persistence loses skill quickly. Baseline model L(3) gives a slight improvement over persistence. Models with memory effect show better skill than L(3), with the primary improvement occurring for strong La Niña patterns (negative PC1). Nonlinear models account for ENSO skewness and mainly improve the skill in the negative PC2 region. The comprehensive model (2L+S+NL) has the best skill overall and shows improvement for diverse target patterns. Note that at long lead (12 month) prediction, the comprehensive model settings do not show much improvement in predicting the extreme El Niño events but somewhat improve the skill for the transition patterns from extreme El Niño to La Niña events (negative PC2 region) as well as the extreme La Niña events (negative PC1 region). Many



FIG. 9. Prediction performance: slippage defined as the lag which has the maximum correlation coefficient. Good prediction performance is indicated by fewer months of slippage. (a),(b) The slippage for 0–12-month-lead predictions for each model in Table 1. Persistence, shown as a reference in black, necessarily has a slippage of τ months for a τ -month-lead prediction. (c),(d) Slippage for 6-month-lead predictions. See Fig. 5 for additional descriptions of prediction methodology.

studies (e.g., Vecchi et al. 2006; Chen et al. 2015) have shown that the amplitude of an extreme El Niño is largely a result of anomaly growth driven by the statedependent noise, such as the westerly wind bursts, and is thus less predictable. Our current 12 models do not consider such state-dependent noise, which is proposed for future study.

e. Summary

From the preceding results, some general conclusions emerge. The memory effect contributes most to PC1 prediction, and nonlinearity contributes most to PC2 prediction. They both contribute to overcoming the seasonal barriers and reducing target month slippage. Adding seasonality does not help much for prediction until combined with memory effect and nonlinearity. Much room for improvement remains, especially for extreme El Niño patterns.

7. Discussion

a. Nonlinearity

Nonlinear models are theoretically superior to linear models in terms of mimicking ENSO behavior, in particular, reproducing the skewed 1D and 2D PDFs. Linear models do not produce the skewed amplitude of

Prediction performance: ENSO diversity-- D₂



FIG. 10. Prediction performance: ENSO diversity. Prediction skill is measured here by the Euclidean distance D_2 between the predicted and GCM target PC1–PC2 pairs. The space is divided into grid cells, and D_2 is averaged according to the target PC1–PC2 grid cell. The gridcell size is 0.33 for PC1 and 0.3 for PC2. (a)–(c) Persistence is shown as a reference. Prediction results are shown for (left) 3-, (center) 6-, and (right) 12-month leads. Good predictions of ENSO diversity are indicated by small D_2 in a greater number of grid cells and for longer leads. See Fig. 5 for additional descriptions of prediction methodology.

ENSO, while quadratic terms correct this deficiency. Nonlinearity does not contribute much to prediction of the slightly skewed PC1 but does matter for PC2. Thus, its advantage may be overlooked if using PC1 (or Niño-3.4) as the only metric, as seen in Kondrashov et al. (2005). On the other hand, we may not be able to obtain robust nonlinear coefficients if the training data lack sufficient extreme events, in which case a linear model may be the more practical choice. For the low-dimension state vectors and long training dataset used in this study, nonlinear models behave well without imposing energy conservation constraints. However, given a large state vector or insufficient training data, numerical instability may make constraints necessary (appendix C of Kondrashov et al. 2015). Therefore, we examined the change in model performance after applying the energy conserving constraint. The results (Fig. 11) show that the advantageous







FIG. 11. Influence of an energy conserving constraint on the four nonlinear models in Table 1. (a),(b) The quality of the model simulations measured by skewness; the GCM data skewness to be simulated is shown on the left. Skewness is chosen because it depends on nonlinearity. (c),(d) Prediction performance, as measured by the correlation coefficient for 12 month forecasts is shown. Models without and with a constraint are grouped separately and the constrained models carry the same color code as the unconstrained versions, but with a "c" subscript. See Fig. 5 for additional descriptions of prediction methodology.

feature of nonlinear models to reproduce data skewness almost disappears; thus, the energy conserving nonlinear model performance ends up close to that of the linear model. The predictive skill for PC2 also slightly decreases.

Nonlinear empirical models have many possible variations in format and algorithm beyond the quadratic terms. In addition to embedding nonlinearity explicitly in the formulation, the skewed nonlinear effect could also be represented in state-dependent noise. Studies suggest that atmospheric noise like westerly wind bursts (WWBs) may play a role in developing extreme El Niño events (Vecchi et al. 2006), but WWBs are in turn modulated by SST variation (Gebbie et al. 2007). Chen et al. (2015) added SST-modulated WWB-like perturbations to an intermediate oceanatmosphere coupled model (Zebiak–Cane model) and successfully reproduced ENSO diversity, asymmetry, and extremes. Chekroun et al. (2011) introduced a past noise forecast (PNF) methodology by embedding the state-dependent noise into empirical prediction models and showed it can improve the ENSO prediction skill up to 16-month lead. Penland and Sardeshmukh (2012) showed the multiplicative and additive noise forcing in a linear Markov model could also produce skewed distribution. Thus, a study of state-dependent noise may be a promising direction for further investigation.

b. Seasonality

The seasonal model with periodic noise is able to reproduce seasonal phase locking for both PC1 and PC2. In contrast, a nonseasonal model even with a large SST state vector cannot reproduce ENSO seasonality, although notable seasonal variations are present in these training PCs. It was expected that the seasonal model may improve the ENSO prediction and, in particular, help overcome the seasonal barrier, but it turned out the seasonal models did not improve prediction skill very much. A similar conclusion is also reached in Flügel and Chang (1998) and Johnson et al. (2000b). On the other hand, our results do show that adding seasonality to a model with memory (2L+S) has slightly better skill than the memory model itself (2L). This is similar to Xue et al. (2000), who found that considering seasonality improved a multivariate LIM. Therefore, it is fair to say that seasonality could improve the ENSO prediction if combined with a memory effect. Penland and Sardeshmukh (1995) and Newman et al. (2003) take an alternative approach of embedding seasonality into the noise forcing rather than in the dynamics. Further studies may explore other possible approaches to represent the seasonality in empirical modeling.

c. Memory effect

We incorporated the memory effect by adding Z20 in a multivariate model or adding more time levels. We confirmed that these approaches indeed improve the ENSO prediction skill, as shown in many previous studies using a multivariate approach (e.g., Xue et al. 1997; Newman et al. 2011b) or multilevel approach (e.g., Kondrashov et al. 2005). We also found that adding time levels of SST adds fewer coefficients than adding Z20 for similar skill improvement. If subsurface information is not available, one could construct an SST-only timedelayed model without loss of skill. We also showed that the memory effect is the most important factor for improving prediction skill, especially to reduce the seasonal barrier and target month slippage. Further study may explore more effective ways to implement and maximize the memory effect.

d. Seasonal barrier

In addition to the well-known spring barrier in SST PC1 (Niño-3.4), we investigated an autumn barrier in SST PC2. It is consistent with the autumn barrier seen in Z20 PC2 (warm water volume mode) in McPhaden (2003), given the strong linear relation between SST PC2 and Z20 PC2. The decorrelation across spring for

PC1 and across autumn for PC2 strongly influence the models that rely heavily on SSTA persistence, but models with memory and nonlinearity can largely retain skill across all four seasons. We also note that, compared to the spring barrier problem in the GFDL CM2.1, the problem in the real world (Figs. 5 and 7 in Barnston et al. 2012) seems more difficult to tackle given various sources of noises.

The seasonal barrier appears not only in statistical models but also in dynamic models (Barnston et al. 2012). Our results from empirical models may help us better understand the seasonal barrier in dynamic model prediction. In spring, the increase of atmospheric noise triggers large anomaly growth (e.g., Larson and Kirtman 2015), and the SSTA field becomes noisy with a less distinguishable pattern (Xue et al. 1994; Wu et al. 2009). Models that initialize with additional subsurface state information work better (e.g., Chen et al. 1995; 2004). It is consistent with our understanding that adding a memory effect using the subsurface variable or SST history compensates the seasonal memory loss in the SST state alone.

e. Target month slippage

Slippage, a measure of forecast mistiming, occurs when models retain too much persistence. Barnston et al. (2012) evaluated forecast models available at IRI and showed this slippage problem is common in both statistical and dynamical models. Tippett et al. (2012) introduced a statistical postprocessing of model output to correct the slippage. Barnston and Tippett (2013) also showed that the Climate Forecast System, version 2, (CFSv2) (Saha et al. 2014) greatly improves on slippage as compared to CFSv1 (Saha et al. 2006). Adding multivariate or multilevel approach substantially reduces slippage, suggesting that the slippage is a consequence of insufficient representation of the memory effect.

f. ENSO diversity

ENSO diversity studies reveal the rich behaviors of the tropical coupled climate system. They also raise new challenges for ENSO modeling and prediction. In this study, we identified summer phase locking and an autumn barrier in PC2, in addition to the better-known spring barrier seen in PC1. The results suggest that if the modeling goal is merely the main ENSO signal (i.e., PC1), then memory effect and seasonality may be sufficient. If the goal is to capture the skewed amplitude and have the ability to predict two flavors of El Niño, then also adding nonlinearity is useful. The new ENSO diversity measure D_2 reveals that the comprehensive model (2L+S+NL) is able to predict a wider range of SSTA patterns than the baseline model. Usually the extreme La Niña events tightly follow the extreme El Niño events (Choi et al. 2013; Cai et al. 2015). Among all the model ingredients, adding nonlinearity most improves the skill for the transition patterns connecting extreme El Niño to La Niña events, thus strengthening the model's ability to simulate and predict diverse target SSTA patterns. This provides evidence supporting that ENSO diversity emerges from the nonlinear evolution of the ENSO cycle, as in Takahashi et al. (2011).

g. Transferring from GFDL CM2.1 to nature

Although there are discrepancies between GFDL CM2.1 and observations in the upper-ocean stratification (Wittenberg et al. 2006), the general understanding, that the memory effect is the most important factor to improve the model skill, which is also supported by previous EMR and LIM studies conducted using observation, is applicable to nature. Here we also show that the memory effect largely reduces the seasonal barrier and slippage problems, although the improvement may be smaller in the real world with more stochasticity.

For a strongly nonlinear system like GFDL CM2.1, nonlinear settings should by construction fit data better than linear settings and should show better skill to simulate and predict the given data. But in the real climate, although nonlinear features do show in extreme El Niño events, the majority of ENSO events are moderate. The overall degree of nonlinearity is much weaker in the observation; the observed SSTA PC2 skewness (-0.4) is much less than that in GFDL CM2.1 (-1.23). So the advantage of the nonlinear model compared with the linear setting may not be that notable in the real world, as in Kondrashov et al. (2005).

8. Conclusions

We use a 4000-yr GFDL CM2.1 preindustrial simulation to help us gain better understanding of four important features in ENSO simulation and prediction: ENSO seasonality, diversity, nonlinearity, and the memory effect. First, we compare the ENSO statistics in simulations with observations, and we find that the GFDL CM2.1 produces reasonably realistic ENSO statistics. It resembles the ENSO diversity feature embedded in the curved shape in the two leading principal components of the tropical Pacific SSTA. It also agrees with observations as to the El Niño–La Niña asymmetry in the skewed probability density function of PC1 (similar to Niño-3.4) and in the winter phase locking. Thus, the model experiments based on GFDL CM2.1 may inform us about nature. But we also note that GFDL CM2.1 has much greater nonlinearity than observed.

In this study, a series of modeling experiments are carried out using empirical models ranging from a simple SSTA linear model to more refined models with additional model coefficients and terms. The conclusions are as follows: The memory effect, either by adding additional tropical Pacific subsurface information (e.g., a multivariate model with SST and 20°C isotherm depth, as in the recharge oscillator viewpoint) or by adding additional SST history information (e.g., an SST-only model with multiple time levels, following the timedelayed oscillator viewpoint), improves the SSTA prediction significantly, although it is more efficient in a multilevel setting with fewer required model coefficients. The nonlinear models with quadratic terms reconstruct the skewed probability density function of SSTA and improve the prediction of the skewed amplitude. The memory effect and nonlinearity enhance the model's ability to retain prediction skill across the so-called spring-autumn prediction barriers and to substantially correct the prediction slippage (i.e., predicted value lags the target month value). The periodic terms enable the model to reproduce the seasonal phase locking of SSTA, even though they do not improve prediction by much. The comprehensive models with combined coefficients have the ability to capture several ENSO characteristics simultaneously and exhibit overall better prediction skill, in agreement with Kondrashov et al. (2005), although they still have difficulty with the prediction of ENSO diversity, especially in predicting extreme El Niño patterns. In summary, this study contributes to our understanding of ENSO diversity, nonlinearity, and seasonality, as well as the memory effect in ENSO simulation and prediction.

Acknowledgments. We thank Michael K. Tippett and Alexey Kaplan for inspiring discussions and valuable suggestions. We also thank three anonymous reviewers for their helpful comments. This research is supported by the Office of Naval Research under the research Grant MURI (N00014-12-1-0911). D. Kondrashov and M. Chekroun are also supported by NSF-OCE 1243175.

REFERENCES

- Alexander, M. A., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. J. Climate, 21, 385–402, doi:10.1175/2007JCLI1849.1.
- An, S.-I., and B. Wang, 2001: Mechanisms of locking of the El Niño and La Niña mature phases to boreal winter. *J. Climate*, 14, 2164–2176, doi:10.1175/1520-0442(2001)014<2164: MOLOTE>2.0.CO;2.

VOLUME 29

—, and F.-F. Jin, 2004: Nonlinearity and asymmetry of ENSO. J. Climate, 17, 2399–2412, doi:10.1175/1520-0442(2004)017<2399: NAAOE>2.0.CO:2.

- Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible teleconnection. J. Geophys. Res., 112, C11007, doi:10.1029/2006JC003798.
- Barnston, A. G., and M. K. Tippett, 2013: Predictions of Niño3.4 SST in CFSv1 and CFSv2: A diagnostic comparison. *Climate Dyn.*, 41, 1615–1633, doi:10.1007/s00382-013-1845-2.

—, —, 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, doi:10.1175/BAMS-D-11-00111.1.

- Battisti, D. S., and A. C. Hirst, 1989: Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry and nonlinearity. J. Atmos. Sci., 46, 1687–1712, doi:10.1175/1520-0469(1989)046<1687:IVIATA>2.0.CO;2.
- Blumenthal, M. B., 1991: Predictability of a coupled oceanatmosphere model. J. Climate, 4, 766–784, doi:10.1175/ 1520-0442(1991)004<0766:POACOM>2.0.CO;2.
- Cai, W., and Coauthors, 2015: ENSO and greenhouse warming. *Nat. Climate Change*, 5, 849–859, doi:10.1038/nclimate2743.
- Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832, doi:10.1038/ 321827a0.
- Capotondi, A., and Coauthors, 2015: Understanding ENSO diversity. Bull. Amer. Meteor. Soc., 96, 921–938, doi:10.1175/ BAMS-D-13-00117.1.
- Chapman, D., M. A. Cane, N. Henderson, D. E. Lee, and C. Chen, 2015: A vector autoregressive ENSO prediction model. *J. Climate*, 28, 8511–8520, doi:10.1175/JCLI-D-15-0306.1.
- Chekroun, M. D., D. Kondrashov, and M. Ghil, 2011: Predicting stochastic systems by noise sampling, and application to the El Niño–Southern Oscillation. *Proc. Natl. Acad. Sci. USA*, **108**, 11766–11771, doi:10.1073/pnas.1015753108.
- Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for EI Niño forecasting: Implications for predictability. *Science*, 269, 1699–1702, doi:10.1126/ science.269.5231.1699.
- —, M. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, 428, 733–736, doi:10.1038/nature02439.

—, and Coauthors, 2015: Strong influence of westerly wind bursts on El Niño diversity. *Nat. Geosci.*, 8, 339–345, doi:10.1038/ ngeo2399.

- Choi, K.-Y., G. A. Vecchi, and A. T. Wittenberg, 2013: ENSO transition, duration, and amplitude asymmetries: Role of the nonlinear wind stress coupling in a conceptual model. *J. Climate*, 26, 9462–9476, doi:10.1175/JCLI-D-13-00045.1.
- Compo, G. P., and P. D. Sardeshmukh, 2010: Removing ENSOrelated variations from the climate record. J. Climate, 23, 1957–1978, doi:10.1175/2009JCLI2735.1.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. J. Climate, 19, 643–674, doi:10.1175/ JCLI3629.1.
- Di Lorenzo, E., K. M. Cobb, J. C. Furtado, N. Schneider, B. T. Anderson, A. Bracco, M. A. Alexander, and D. J. Vimont, 2010: Central Pacific El Niño and decadal climate change in the North Pacific Ocean. *Nat. Geosci.*, 3, 762–765, doi:10.1038/ ngeo984.
- Fedorov, A. V., S. Hu, M. Lengaigne, and E. Guilyardi, 2015: The impact of westerly wind bursts and ocean initial state on the

development, and diversity of El Niño events. *Climate Dyn.*, **44**, 1381–1401, doi:10.1007/s00382-014-2126-4.

- Flügel, M., and P. Chang, 1998: Does the predictability of ENSO depend on the seasonal cycle? J. Atmos. Sci., 55, 3230–3243, doi:10.1175/1520-0469(1998)055<3230:DTPOED>2.0.CO;2.
- Gebbie, G., I. Eisenman, A. Wittenberg, and E. Tziperman, 2007: Modulation of westerly wind bursts by sea surface temperature: A semistochastic feedback for ENSO. J. Atmos. Sci., 64, 3281–3295, doi:10.1175/JAS4029.1.
- Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. J. Climate, 27, 57–75, doi:10.1175/JCLI-D-12-00837.1.
- Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. J. Climate, 10, 1769–1786, doi:10.1175/1520-0442(1997)010<1769: ENOLNA>2.0.CO;2.
- Jin, F.-F., 1997a: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. J. Climate, 54, 811–829, doi:10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.
- —, 1997b: An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model. J. Climate, 54, 830–847, doi:10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2.
- Johnson, S. D., D. S. Battisti, and E. S. Sarachik, 2000a: Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies. J. Climate, 13, 3–17, doi:10.1175/1520-0442(2000)013<0003:EDMMAP>2.0.CO;2.
- —, —, and —, 2000b: Seasonality in an empirically derived Markov model of tropical Pacific sea surface temperature anomalies. J. Climate, **13**, 3327–3335, doi:10.1175/ 1520-0442(2000)013<3327:SIAEDM>2.0.CO;2.
- Kang, I.-S., and J.-S. Kug, 2002: El Niño and La Niña sea surface temperature anomalies: Asymmetry characteristics associated with their wind stress anomalies. J. Geophys. Res., 107, 4372, doi:10.1029/2001JD000393.
- Kao, H.-Y., and J.-Y. Yu, 2009: Contrasting eastern-Pacific and central-Pacific types of ENSO. J. Climate, 22, 615–632, doi:10.1175/2008JCLI2309.1.
- Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. J. Geophys. Res., 103, 18567– 18589, doi:10.1029/97JC01736.
- 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, doi:10.1007/s00382-013-1759-z.
- Karnauskas, K. B., 2013: Can we distinguish canonical El Niño from Modoki? *Geophys. Res. Lett.*, 40, 5246–5251, doi:10.1002/ grl.51007.
- Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A hierarchy of data-based ENSO models. J. Climate, 18, 4425– 4444, doi:10.1175/JCL13567.1.

—, —, and M. Ghil, 2011: Signatures of nonlinear dynamics in an idealized atmospheric model. J. Atmos. Sci., 68, 3–12, doi:10.1175/2010JAS3524.1.

- —, M. D. Chekroun, and M. Ghil, 2015: Data-driven non-Markovian closure models. *Physica D*, **297**, 33–55, doi:10.1016/ j.physd.2014.12.005.
- Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multilevel regression modeling of nonlinear processes. J. Climate, 18, 4404– 4424, doi:10.1175/JCL13544.1.
- Kug, J.-S., F.-F. Jin, and S.-I. An, 2009: Two types of El Niño events: Cold tongue El Niño and warm pool El Niño. *J. Climate*, **22**, 1499–1515, doi:10.1175/2008JCLI2624.1.

- —, 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, doi:10.1175/ 2009JCLI3293.1.
- Larkin, N. K., and D. E. Harrison, 2002: ENSO warm (El Niño) and cold (La Niña) event life cycles: Ocean surface anomaly patterns, their symmetries, asymmetries, and implications. *J. Climate*, **15**, 1118–1140, doi:10.1175/1520-0442(2002)015<1118: EWENOA>2.0.CO;2.
 - —, and —, 2005: On the definition of El Niño and associated seasonal average U.S. weather anomalies. *Geophys. Res. Lett.*, 32, L13705, doi:10.1029/2005GL022738.
- Larson, S. M., and B. P. Kirtman, 2015: Revisiting ENSO coupled instability theory and SST error growth in a fully coupled model. *J. Climate*, 28, 4724–4742, doi:10.1175/JCLI-D-14-00731.1.
- Lee, D. E., D. Chapman, N. Henderson, C. Chen, and M. A. Cane, 2015: Multilevel vector autoregressive prediction of sea surface temperature in the north tropical Atlantic Ocean and the Caribbean Sea. *Climate Dyn.*, 1–12, doi:10.1007/ s00382-015-2825-5.
- Lee, T., and M. J. McPhaden, 2010: Increasing intensity of El Niño in the central-equatorial Pacific. *Geophys. Res. Lett.*, 37, L14603, doi:10.1029/2010GL044007.
- McCullagh, P., and J. A. Nelder, 1989: Generalized Linear Models. 2nd ed. CRC Monographs on Statistics and Applied Probability, Vol. 37, Chapman and Hall, 511 pp.
- McPhaden, M. J., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.*, 30, 1995–1998, doi:10.1029/2003GL016872.
- Neelin, J. D., F.-F. Jin, and H.-H. Syu, 2000: Variations in ENSO phase locking. J. Climate, 13, 2570–2590, doi:10.1175/ 1520-0442(2000)013<2570:VIEPL>2.0.CO;2.
- Newman, M., P. D. Sardeshmukh, C. R. Winkler, and J. S. Whitaker, 2003: A study of subseasonal predictability. *Mon. Wea. Rev.*, **131**, 1715–1732, doi:10.1175//2558.1.
- —, M. A. Alexander, and J. D. Scott, 2011a: An empirical model of tropical ocean dynamics. *Climate Dyn.*, **37**, 1823–1841, doi:10.1007/s00382-011-1034-0.
- —, S.-I. Shin, and M. A. Alexander, 2011b: Natural variation in ENSO flavors. *Geophys. Res. Lett.*, **38**, L14705, doi:10.1029/ 2011GL047658.
- Okumura, Y. M., and C. Deser, 2010: Asymmetry in the duration of El Niño and La Niña. J. Climate, 23, 5826–5843, doi:10.1175/ 2010JCLI3592.1.
- Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165–2185, doi:10.1175/1520-0493(1989)117<2165:RFAFUP>2.0.CO;2.
- —, 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98**, 534–558, doi:10.1016/ 0167-2789(96)00124-8.
- —, and T. Magorian, 1993: Prediction of Niño 3 sea surface temperatures using linear inverse modeling. J. Climate, 6, 1067–1076, doi:10.1175/1520-0442(1993)006<1067:PONSST>2.0.CO;2.
- —, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. J. Climate, 8, 1999–2024, doi:10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.
- —, and L. Matrosova, 1998: Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. J. Climate, 11, 483–496, doi:10.1175/1520-0442(1998)011<0483: POTASS>2.0.CO:2.
- —, and P. D. Sardeshmukh, 2012: Alternative interpretations of power-law distributions found in nature. *Chaos*, **22**, 023119, doi:10.1063/1.4706504.

- Rayner, N. A., D. E. Parker, E. B. Horton, C. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. J. Climate, 19, 3483–3517, doi:10.1175/JCLI3812.1.
- —, and Coauthors, 2014: The NCEP Climate Forecast System version 2. J. Climate, 27, 2185–2208, doi:10.1175/ JCLI-D-12-00823.1.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land– ocean surface temperature analysis (1880–2006). J. Climate, 21, 2283–2296, doi:10.1175/2007JCLI2100.1.
- Suarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. J. Atmos. Sci., 45, 3283–3287, doi:10.1175/ 1520-0469(1988)045<3283:ADAOFE>2.0.CO:2.
- Takahashi, K., and B. Dewitte, 2015: Strong and moderate nonlinear El Niño regimes. *Climate Dyn.*, 1–19, doi:10.1007/ s00382-015-2665-3.
- —, A. Montecinos, K. Goubanova, and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophys. Res. Lett.*, 38, L10704, doi:10.1029/2011GL047364.
- Tangang, F. T., B. Tang, A. H. Monahan, and W. W. Hsieh, 1998: Forecasting ENSO events: A neural network– extended EOF approach. J. Climate, 11, 29–41, doi:10.1175/ 1520-0442(1998)011<0029:FEEANN>2.0.CO;2.
- Thompson, C. J., and D. S. Battisti, 2001: A linear stochastic dynamical model of ENSO. Part II: Analysis. J. Climate, 14, 445–466, doi:10.1175/1520-0442(2001)014<0445:ALSDMO>2.0.CO;2.
- Timmermann, A., H. U. Voss, and R. Pasmanter, 2001: Empirical dynamical system modeling of ENSO using nonlinear inverse techniques. J. Phys. Oceanogr., 31, 1579–1598, doi:10.1175/ 1520-0485(2001)031<1579:EDSMOE>2.0.CO;2.
- Tippett, M. K., A. G. Barnston, and S. Li, 2012: Performance of recent multimodel ENSO forecasts. J. Appl. Meteor. Climatol., 51, 637–654, doi:10.1175/JAMC-D-11-093.1.
- Tziperman, E., L. Stone, M. A. Cane, and H. Jarosh, 1994: El Nino chaos: Overlapping of resonances between the seasonal cycle and the Pacific Ocean–atmosphere oscillator. *Science*, 264, 72– 74, doi:10.1126/science.264.5155.72.
- —, M. A. Cane, and S. E. Zebiak, 1995: Irregularity and locking to the seasonal cycle in an ENSO prediction model as explained by the quasi-periodicity route to chaos. *J. Atmos. Sci.*, **52**, 293–306, doi:10.1175/1520-0469(1995)052<0293: IALTTS>2.0.CO:2.
- —, S. E. Zebiak, and M. A. Cane, 1997: Mechanisms of seasonal– ENSO interaction. J. Atmos. Sci., 54, 61–71, doi:10.1175/ 1520-0469(1997)054<0061:MOSEI>2.0.CO;2.
- M. A. Cane, S. E. Zebiak, Y. Xue, and B. Blumenthal, 1998: Locking of El Niño's peak time to the end of the calendar year in the delayed oscillator picture of ENSO. *J. Climate*, 11, 2191–2199, doi:10.1175/1520-0442(1998)011<2191: LOENOS>2.0.CO;2.
- Vecchi, G. A., A. T. Wittenberg, and A. Rosati, 2006: Reassessing the role of stochastic forcing in the 1997–1998 El Niño. *Geophys. Res. Lett.*, **33**, L01706, doi:10.1029/2005GL024738.
- Vimont, D. J., M. A. Alexander, and M. Newman, 2014: Optimal growth of central and east Pacific ENSO events. *Geophys. Res. Lett.*, 41, 4027–4034, doi:10.1002/2014GL059997.
- Wetherill, G. B., 1986: Regression Analysis with Applications. Monographs on Statistics and Applied Probability, Vol. 27, Chapman and Hall, 311 pp.

- Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, 36, L12702, doi: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, doi: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, doi:10.1175/JCLI-D-13-00577.1.
- Wu, R., B. P. Kirtman, and H. van den Dool, 2009: An analysis of ENSO prediction skill in the CFS retrospective forecasts. *J. Climate*, 22, 1801–1818, doi:10.1175/2008JCLI2565.1.
- Xue, Y., M. A. Cane, S. E. Zebiak, and M. B. Blumenthal, 1994: On the prediction of ENSO: A study with a low-order Markov model. *Tellus*, 46A, 512–528, doi:10.1034/j.1600-0870.1994.00013.x.
- , —, —, and T. N. Palmer, 1997: Predictability of a coupled model of ENSO using singular vector analysis. Part II: Optimal growth and forecast skill. *Mon. Wea. Rev.*, 125, 2057–2073, doi:10.1175/1520-0493(1997)125<2057: POACMO>2.0.CO;2.
- —, A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: The impact of sea level. J. Climate, 13, 849–871, doi:10.1175/1520-0442(2000)013<0849:EPWMMT>2.0.CO;2.
- —, M. Chen, A. Kumar, Z.-Z. Hu, and W. Wang, 2013: Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2. *J. Climate*, 26, 5358–5378, doi:10.1175/JCLI-D-12-00600.1.
- Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278, doi:10.1175/ 1520-0493(1987)115<2262:AMENO>2.0.CO;2.