

# Seasonal predictions using a simple ocean initialization scheme

Jieshun Zhu<sup>1,2</sup> · Arun Kumar<sup>1</sup> · Hyun-Chul Lee<sup>1,3</sup> · Hui Wang<sup>1</sup>

Received: 12 July 2016 / Accepted: 25 January 2017  
© Springer-Verlag Berlin Heidelberg (outside the USA) 2017

**Abstract** In this study, a simple ocean data assimilation scheme was applied to initialize a set of seasonal hindcasts, which started from each January, April, July and October during 1982–2010 with six ensemble members. In the scheme, sea surface temperature (SST) was the only observed information used to estimate the ocean initial states. Predictions of SST, 2-m temperature (T2m) and precipitation (Prate) over land were assessed and compared with hindcasts from the North American Multimodel Ensemble (NMME) project which were all based on sophisticated ocean initialization schemes with subsurface observations assimilated. The skill comparison indicated that, for all variables evaluated, the prediction skill by the simple ocean initialization procedure was well within the range of skills from individual NMME models. The result suggests that even though sophisticated initialization schemes have the potential to best capture the seasonal climate predictability, most present-day capabilities of seasonal predictions can also be accomplished by utilizing SST only. Further, significant seasonal dependence of prediction skill was also identified in hindcasts by the simple initialization scheme. Specifically, the ENSO SST predictions were featured by the significant “spring barrier” problem. It is argued that the ENSO prediction skill seasonality could be due to “double dip” contribution from the seasonality of climatic feedbacks in the tropical Pacific. In particular, these

feedbacks not only play a role during the forecast leading to the largest errors in spring, but also cause the largest errors in the SST-derived ocean initial conditions in spring. Complementary aspects of seasonal forecasts based on simple initialization scheme, and use of low-resolution models are also discussed.

## 1 Introduction

Seasonal predictions, though challenging, have been made routinely at many operational prediction centers around the world (e.g., Palmer et al. 2004; Molteni et al. 2011; Saha et al. 2006, 2014; Graham et al. 2011). The progress in dynamical seasonal predictions has benefitted from advances in both coupled models and improvements in observational network. Regarding the latter, ENSO predictions, for example, have been argued to be greatly benefited from the advent of Tropical Atmosphere Ocean (TAO) moored array in the 1980s (McPhaden et al. 1995).

Seasonal climate predictability mainly resides in the ocean memory, and the ocean initialization is vital for how much potential predictability in the coupled system could be realized by dynamical seasonal prediction systems. To have as accurate ocean initial conditions as possible, various data assimilation schemes have been implemented to synthesize ocean observations (including in situ and satellite based products) with ocean models (e.g., Behringer and Xue 2004; Balmaseda et al. 2013). At National Centers for Environmental Prediction (NCEP), for example, a 3D variational technique (Derber and Rosati 1989) is used to assimilate different surface/subsurface and satellite ocean observations. The resulting ocean analysis is then used to initialize climate prediction systems (Saha et al. 2006, 2014).

---

✉ Jieshun Zhu  
jieshun.zhu@noaa.gov

<sup>1</sup> Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740, USA

<sup>2</sup> Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

<sup>3</sup> Innovim, Greenbelt, MD, USA

The sophisticated data assimilation and initialization schemes usually require extensive scientific and technical efforts, and if forecasts are extended over period of time, face numerous challenges. For instance, as the global observing system evolves over the time (particularly during the satellite era), the issue of “data inhomogeneity” poses significant challenges in estimating the state of the ocean state over an extended period of time (e.g., Zhang et al. 2012). The problem not only affects climate analysis and diagnostics, but also can affect subsequent seasonal predictions (often referred to as hindcasts). Further, as the seasonal hindcasts are also used to correct biases and calibrate real-time seasonal predictions, inhomogeneities in the analysis can cause further issues.

Kumar et al. (2012) illustrated a quantitative change in the characteristics of SST forecast biases in the equatorial Pacific in the NCEP Climate Forecast System version 2 (CFSv2) hindcasts (Saha et al. 2014) across 1999, characterized by significantly warmer mean for predicted SST after 1999 than before. They suggested that the discontinuity resulted from the assimilation of new satellite observations in the analysis system after 1999 (Kumar et al. 2012; Zhang et al. 2012) from which forecast initial conditions were obtained [i.e., the Climate Forecast System Reanalysis (CFSR); Saha et al. 2010]. Subsequent analysis by Xue et al. (2013) found that a better ENSO SST prediction skill could be achieved in the tropical Pacific if two climatologies (1982–1998 and 1999–2010) were applied in the computation of predicted anomalies. Such problems made the skill assessment of hindcasts, and bias correction and calibration of real-time forecasts a taxing issue.

The sophisticated initialization schemes also depend heavily on availability of observations. This limitation restricts the sample size of seasonal hindcasts to typically over a 30-year period (i.e., from the early 1980s to present). On the other hand, there are also alternative methods using simpler ocean assimilation and initialization schemes that have been used successfully in seasonal predictions. For example, the ocean initial states can be obtained by running a coupled model with its SST or SST anomalies (SSTA) strongly nudged to observations. The simple initialization method has been used in ENSO predictions with both intermediate complexity models (ICMs; e.g. Zhang et al. 2003; Chen et al. 2004) mostly by nudging towards observed SST anomalies, and with coupled global climate models (CGCMs; Keenlyside et al. 2005; Luo et al. 2005). The physical basis for these methods is that, in addition to providing a realistic oceanic mixed layer temperature because of SST nudging, the observed SST information, through air-sea interaction and coupling, is able to partially reproduce observed ocean subsurface evolutions. This is achieved via coupled air-sea interaction whereby observed SST information results in a realistic simulation of surface

winds, which in turn is able to generate observed sub-surface ocean evolution (Kumar et al. 2014; Kohyama and Tozuka 2016).

Based on a simple initialization scheme the Scale Interaction Experiment-Frontier Research Center for Global Change (SINTEX-F) model achieved a remarkable success in ENSO predictions (Luo et al. 2005) at lead times even up to 2 years (Luo et al. 2008). An advantage of such schemes is that they can be easily implemented in coupled systems and do not require an extensive infrastructure to collect real-time observational data and a sophisticated data assimilation system. Furthermore, in terms of ENSO predictions, given the sparsity of ocean observations in the tropical Pacific before the advent of TAO in the 1990s, initialization by utilizing the SST information provides a viable alternative and could extend the time series of ENSO forecasts much farther back than is generally feasible based on sophisticated data assimilation schemes (e.g., Chen et al. 2004; Deng and Tang 2009; Zheng et al. 2009).

Use of a simple initialization scheme is tested with the NCEP’s seasonal prediction system and the analysis reported here extends the work of Zhu et al. (2015a). As a preliminary step towards assessing feasibility of this procedure, the implementation in Zhu et al. (2015a) was not optimal in its experimental design, and might have underestimated its potential. First, in Zhu et al. (2015a) the model [i.e., Climate Forecast System, version 1 (CFSv1)] generating ocean initial conditions (OICs) was not the same as that used for predictions (i.e., CFSv2). The discrepancy could bring some inconsistencies or shocks during the forecast as results of the differences between the climatologies of CFSv1 and CFSv2, even though the two models with the same origin show some common biases, like the warm SST biases in the southeast Pacific and Atlantic (figures not shown). Secondly, the atmospheric initial conditions in Zhu et al. (2015a) were directly taken from CFSR (Saha et al. 2010), different from the practice of Luo et al. (2005, 2008) that adopted both atmospheric and oceanic initial conditions from a continuous run with SSTs nudged to observations. Thus, Zhu et al. (2015a) generated less balanced initial conditions between ocean and atmosphere than SINTEX-F (Luo et al. 2005, 2008). Furthermore, Zhu et al. (2015a) reported predictions initialized from April only, even though it is well known that spring is the most challenging season particularly for ENSO prediction because of the existence of the spring predictability barrier (e.g. Luo et al. 2005; Jin et al. 2008). Because of this shortcoming, it was not possible to assess the seasonal dependence of prediction skill associated with the initialization scheme based on the hindcasts of Zhu et al. (2015a).

This study, extending the analysis of Zhu et al. (2015a), conducts a set of hindcasts with a better experimental design. In particular, similar to Luo et al. (2005, 2008), the

hindcasts are based on a low-resolution CFSv2 model (see Sect. 2 for more details), and they are initialized by continuous runs using the same model with its SSTs nudged to observations. The hindcasts also start from each January, April, July and October during 1982–2010 with six ensemble members, allowing us to provide a comprehensive evaluation of the seasonality of prediction skill. To contrast the skill of our prediction system, we compare the hindcasts with those from the North American Multimodel Ensemble (NMME) project (Kirtman et al. 2014). As all the NMME hindcasts were initialized by sophisticated data assimilation systems, this comparison also highlights the potential importance of subsurface ocean observations in the context of ENSO prediction. The skill assessments are all based on deterministic measures. However, it has been shown by Kumar (2009) based on a comparison of skill assessment for (a) deterministic, (b) categorical, and (c) probabilistic predictions that the corresponding skill scores have a one-to-one relationship (see Fig. 2; Kumar 2009). Simply, low (high) anomaly correlations for ensemble mean as deterministic forecast also correspond to low (high) values for the probabilistic skill measure (e.g., the rank probability skill score). This theoretical relationship has also been verified in real time forecast situations (see Fig. 14; Sooraj et al. 2012).

The paper is arranged as follows. The forecast model, the experimental design and datasets are described in the next section. Section 3 validates the hindcasts by comparing with NMME hindcasts. Even though it is understood that the seasonality plays an important role in determining the predictability, the comparison with NMME is conducted regardless of season simply to limit the paper to a reasonable length. However, to some extent the seasonality of forecast skill in our hindcasts using the simple ocean initialization scheme is evaluated in Sect. 4. The summary and discussion are given in Sect. 5.

## 2 Model, hindcast experiments and datasets

### 2.1 Model

The forecast model used in this study is a variant of NCEP CFSv2 (Saha et al. 2014) with lower horizontal resolutions in both atmospheric and oceanic components (Lee et al. 2016). A lower resolution is to enhance throughput, and if compared against the CFSv2, also provides some information about the influence of resolution on seasonal predictions. To distinguish from the standard CFSv2 currently used for the operational seasonal-to-interannual prediction at NCEP, the low-resolution CFSv2 is referred to as CFSv2L. In CFSv2L (CFSv2), the ocean model is the GFDL MOM version 4, which is configured for the global

ocean with a horizontal grid of  $1^\circ \times 1^\circ$  ( $0.5^\circ \times 0.5^\circ$ ) poleward of  $30^\circ\text{S}$  and  $30^\circ\text{N}$  and meridional resolution increasing gradually to  $0.33^\circ$  ( $0.25^\circ$ ) between  $10^\circ\text{S}$  and  $10^\circ\text{N}$ . The vertical coordinate is geopotential ( $z$ -) with 40 levels (27 of them in the upper 400 m), with maximum depth of approximately 4.5 km. The atmospheric model of CFSv2L (CFSv2) is the Global Forecast System, which has horizontal resolution at T62 (T126), and 64 vertical levels in a hybrid sigma-pressure coordinate. The oceanic and atmospheric components of CFSv2L (CFSv2) exchange surface momentum, heat and freshwater fluxes, as well as SSTs every 60 (30) min. More details about CFSv2L are referred to Lee et al. (2016).

### 2.2 CFSv2L hindcasts with a simple initialization scheme

In this study, a set of hindcasts with CFSv2L is initialized by the same simple scheme as in Luo et al. (2005) and Keenlyside et al. (2005), in which SST is the only observed information provided to derive ocean initial states. This run can be thought as a simple data assimilation procedure where the only observed information ingested is the SSTs. In particular, six CFSv2L integrations, differing by their initial conditions, were conducted with model SSTs nudged to the observational counterpart. The six initial conditions differ in ocean only, which apply the CFSR (Saha et al. 2010) ocean states at 00Z of 29th–31st December 1980 and 1st–3rd January 1981. The atmospheric and land initial conditions are identical, using their CFSR states at 00Z of 1st January 1981. By re-dating all the initial conditions to 1st January 1976, six CFSv2L runs are integrated forward with model SSTs nudged to the observed daily SSTs. The restoring time scale is chosen as 3.3 days, following our previous work with CFSv1 (Wang et al. 2013; Kumar et al. 2014). The observed daily SSTs are interpolated from the monthly SSTs from the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation SST (OISST) version 2 (Reynolds et al. 2002) for the period after 1982, and from the ERSST v3 (Smith et al. 2008) before 1982. A preliminary analysis of the SST-nudged CFSv2L simulations suggests that more realistic evolutions of subsurface ocean temperature were achieved than by CFSv1 (Kumar et al. 2014; e.g., Fig. 12a vs. 10 of; Zhu et al. 2015a), and might be due to model improvement from CFSv1 to CFSv2.

Using the restart files saved during the six SST-nudged CFSv2L integrations, hindcasts are conducted starting from the first day of each January, April, July and October during 1982–2010 and last for 8 months. The set of hindcasts is referred to as CFSv2L\_nudg. The experimental design here is better than that in Zhu et al. (2015a) where different models were used for OICs generation and hindcasts, and

atmospheric initial conditions and OICs were adopted from different sources.

### 2.3 NMME hindcast data

The hindcasts of CFSv2L\_nudg are assessed and compared with hindcasts from the NMME project (Kirtman et al. 2014). In this study, hindcasts from 11 NMME models are used, including CMC-CanCM3 (referred to as CMC1; Merryfield et al. 2013), CMC-CanCM4 (referred to as CMC2; Merryfield et al. 2013), NCAR CCSM4 (Kirtman et al., in preparation), GFDL CM2.1 (Delworth et al. 2006), GFDL FLORa06 (Vecchi et al. 2014), GFDL FLORb01 (Vecchi et al. 2014), NASA GEOS5 (Vernieres et al. 2012), NCEP CFSv2 (Saha et al. 2014), IRI ECHAM4.5a (DeWitt 2005), IRI ECHAM4.5f (DeWitt 2005) and NCAR CCSM3 (Kirtman and Min 2009). All the systems have run hindcasts over the period 1982–2010, with ensemble members ranging from 6 to 24 (28), and forecasts lengths between 8 and 12 months. CFSv1 (Saha et al. 2006) was included in the first year of NMME, but it was discontinued after the implementation of CFSv2 and its hindcast did not extend beyond 2009. Thus, CFSv1 is not included in our analysis. In addition, ECHAM4.5a, ECHAM4.5f and CCSM3 were also discontinued after the first year of NMME. The hindcast data for NMME are available on a  $1^\circ \times 1^\circ$  grid resolution. A brief description of the NMME models is provided in Table 1. More details about NMME and their hindcasts can be found in Kirtman et al. (2014) and Becker et al. (2014). It should be noted that, different from CFSv2L\_nudg, all the NMME hindcasts were initialized by sophisticated ocean data assimilation systems.

In this study, all predictions are analyzed for the common lead times of NMME, i.e., 0–7 months. For the definition of forecast lead, the 1-month lead forecast from April initial conditions, for example, is the forecast for May (the forecast for April itself would be defined as the 0-month lead forecast). To rule out skill differences due to ensemble size (Kumar and Hoerling 2000; Scaife et al. 2014), six ensemble members for all NMME models are used, which are simply chosen as the first six members of each NMME model hindcasts except for CFSv2. For CFSv2, a so-called lagged ensemble method (Saha et al. 2014) is used to generate its ensemble members, by which the early (late) members from the target month have longer (shorter) lead times. To minimize the effect of different lead times, the 14–19th out of 24 (28) members are picked for our diagnostics, which are close to the first day of start month.

### 2.4 Data for validations

The observed monthly SST analysis used for validation is from OISST (Reynolds et al. 2002) on a  $1.0^\circ \times 1.0^\circ$

(latitude  $\times$  longitude) grid. The subsurface ocean temperatures from the NCEP Global Ocean Data Assimilation System (GODAS; Behringer and Xue 2004) are used for the validation of the upper-ocean heat content (HC; defined as the depth-average temperature of the upper 300 m). The verification data for T2m is the station observation-based Global Historical Climatology Network+Climate Anomaly Monitoring System (GHCN+CAMS; Fan and van den Dool 2008), a monthly mean surface air temperature dataset on a  $0.5^\circ \times 0.5^\circ$  (latitude  $\times$  longitude) grid. The predicted precipitations are verified with the Climate Prediction Center (CPC) global daily Unified Rain gauge Database (URD) gauge analysis (Xie et al. 2010, personal communication), which is also available on a  $0.5^\circ \times 0.5^\circ$  grid. The daily data were averaged into monthly means.

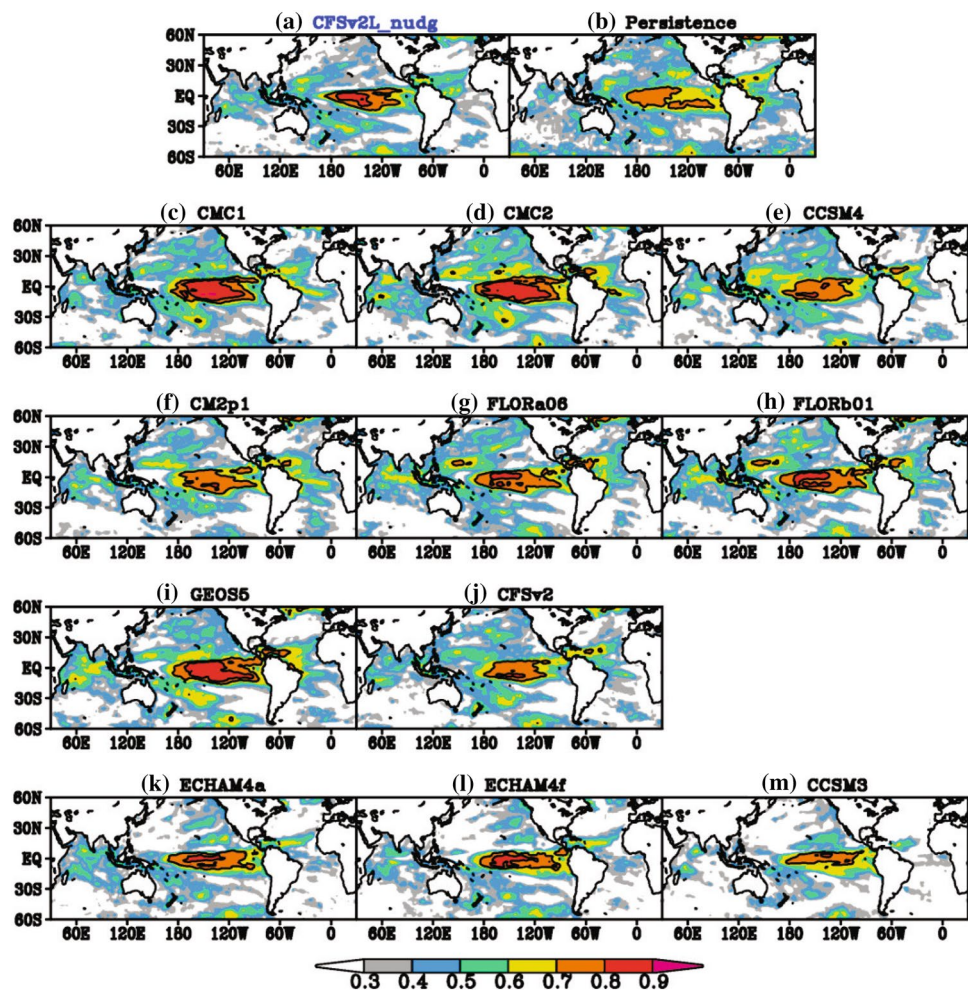
All analyses are based on hindcasts during 1982–2010. The predicted anomalies are derived by subtracting climatologies which are a function of both initial condition and lead time, and no additional time smoothing was applied. The climatologies are based on the whole hindcast period of 1982–2010. For the CFSv2 hindcasts, even though applying two climatologies could achieve better SST prediction skill in the tropical Pacific (Xue et al. 2013), it clearly degrades SST prediction skill in other ocean basins, particularly in the extratropical oceans. Considering that the global predictions are the scope of this study, we chose to use one climatology derived from the whole forecast period (i.e., 1982–2010). This choice is also consistent with evaluations of hindcasts from other NMME models.

## 3 Skill comparisons with NMME models

In this section, the overall prediction skill of CFSv2L\_nudg is evaluated by comparing with models participating in the NMME project (Kirtman et al. 2014). The comparisons are based on all four start months (i.e., January, April, July and October) regardless of season, and the variables evaluated include SST, T2m and precipitation over the land.

Figures 1 and 2 show the horizontal distributions of quasi-global SST prediction skill for lead times of 3 and 6 months, respectively. As expected, at both lead times regions with the highest correlation are located in the tropical Pacific in all systems. Among hindcasts from different NMME models (Figs. 1c–m, 2c–m), however, it is interesting to notice substantial spreads in their SST prediction skills, even in the tropical Pacific. For example, at the 3-month lead time, skill of CMC1, CMC2 and GEOS5 is better than others with correlations greater than 0.8 over a large area of the central and eastern tropical Pacific, while in other systems regions of correlations greater than 0.8 are clearly smaller. At the 6-month lead time (Fig. 2), the SSTA prediction

**Fig. 1** Distribution of anomaly correlations between observed and predicted SST anomalies at 3-month lead time in **a** CFSv2L\_nudg, **b** persistence, **c** CMC1, **d** CMC2, **e** CCSM4, **f** CM2p1, **g** FLORa06, **h** FLORb01, **i** GEOS5, **j** CFSv2, **k** ECHAM4a, **l** ECHAM4f, and **m** CCSM3. The hindcasts start from January, April, July and October initial conditions during 1982–2010. All shading areas are above 90% confidence level [correlation of 0.245 is 90% confidence level according to one-tailed Student's *t* test with the degree of freedom (DOF) of 27]. Correlations of 0.8 and 0.7 are highlighted with black contours

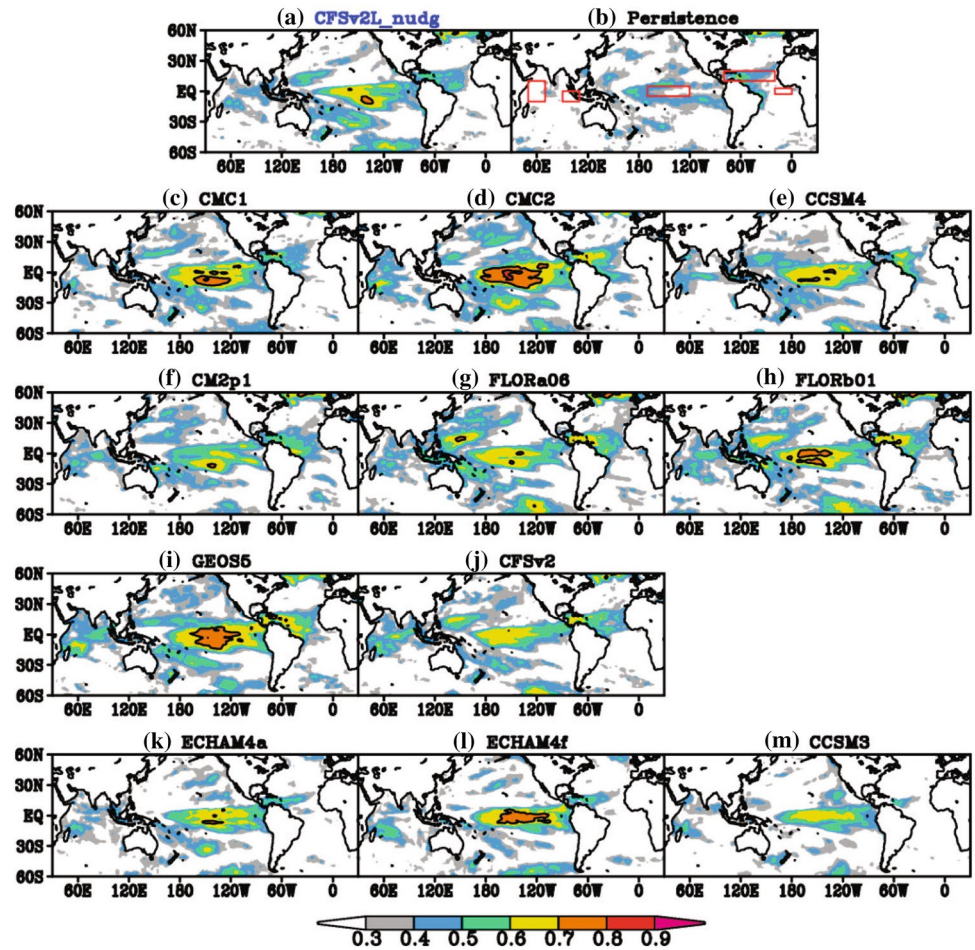


skill drops substantially in the tropical Pacific in all systems, and no system has correlations greater than 0.8 except over a small patch in CMC2. At the 6-month lead time, CMC2 shows the best correlation skill, which is evidenced by the largest region in the central and eastern Pacific with correlations larger than 0.7 and a patch larger than 0.8; GEOS5, ECHAM4f, CMC1, and FLORb01 could be ranked as the second best, as suggested by sizeable regions with correlations larger than 0.7 which is almost absent in other systems. In addition to the tropical Pacific, some skill is also present in other ocean basins as in previous studies (e.g., Luo et al. 2007; Hu and Huang 2007; Hu et al. 2013, 2014; Guan et al. 2014; Zhu et al. 2015d), such as the tropical Indian Ocean, tropical Atlantic Ocean, and the extratropical oceans including the North Atlantic, the North Pacific and the southern subtropical Pacific. In these regions, some models like CCSM3 seem to exhibit clearly lower skill than others. In addition, it was also noticed that the three discontinued models (i.e., ECHAM4a, ECHAM4f and CCSM3) after the first year of NMME also present reasonably

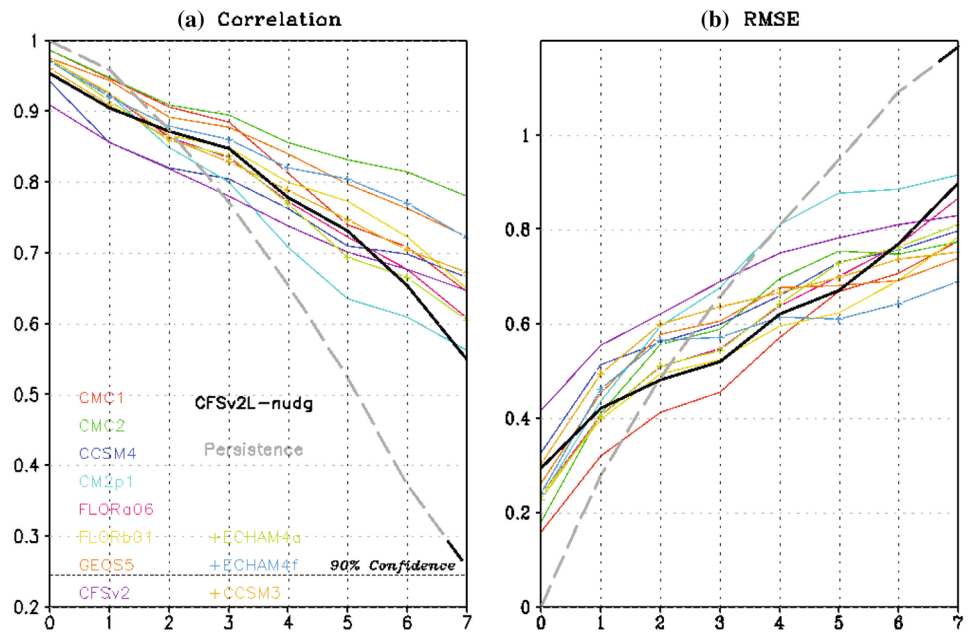
good skill. For example, ECHAM4f is among the top tier in predicting SSTA in the tropical Pacific at the 6-month lead time (Fig. 2l).

For CFSv2L\_nudg, prediction skill in the tropical Pacific is not only superior to persistence especially at the longer lead times (e.g., at the 6-month lead; Fig. 2a vs. b), it is also on par with some NMME models, which are all based on sophisticated ocean initialization schemes with the assimilation of subsurface observations. For example, at the 3-month lead time, in terms of the area surrounded by the 0.8 correlation contour in the tropical Pacific, CFSv2L\_nudg is better than or equivalent to models like CCSM3, CM2p1, CCSM4, FLORa06, CFSv2, etc. In other basins, the SST prediction skill of CFSv2L\_nudg at the 3-month lead time is also within the skill range of individual NMME models. We note that some NMME models do not beat persistence over regions like the northern tropical Atlantic and the northern North Atlantic. At the 6-month lead time, skill of CFSv2L\_nudg is better than persistence and some NMME models (e.g., CCSM3, ECHAM4a, ECHAM4f).

**Fig. 2** As in Fig. 1, but at the 6-month lead time. The *red boxes* in **b** indicate the area of Niño-3.4, MDR, ATL3, EIO and WIO, respectively



**Fig. 3** **a** Anomaly correlation coefficients and **b** RMSEs ( $^{\circ}\text{C}$ ) of Niño-3.4 index as a function of forecast lead months (x-axis) after removing the mean bias. *Solid black curves* (*dashed grey curves*) are for forecasts of CFSv2L\_nudg (persistence). *Solid colored curves* are forecasts for each NMME member model. The hindcasts start from January, April, July and October initial conditions during 1982–2010. The *dashed horizontal line* in Fig. 3a indicates correlation (0.245) at the 90% confidence level, according to one-tailed Student's *t* test with DOF of 27



**Table 1** Description of NMME models

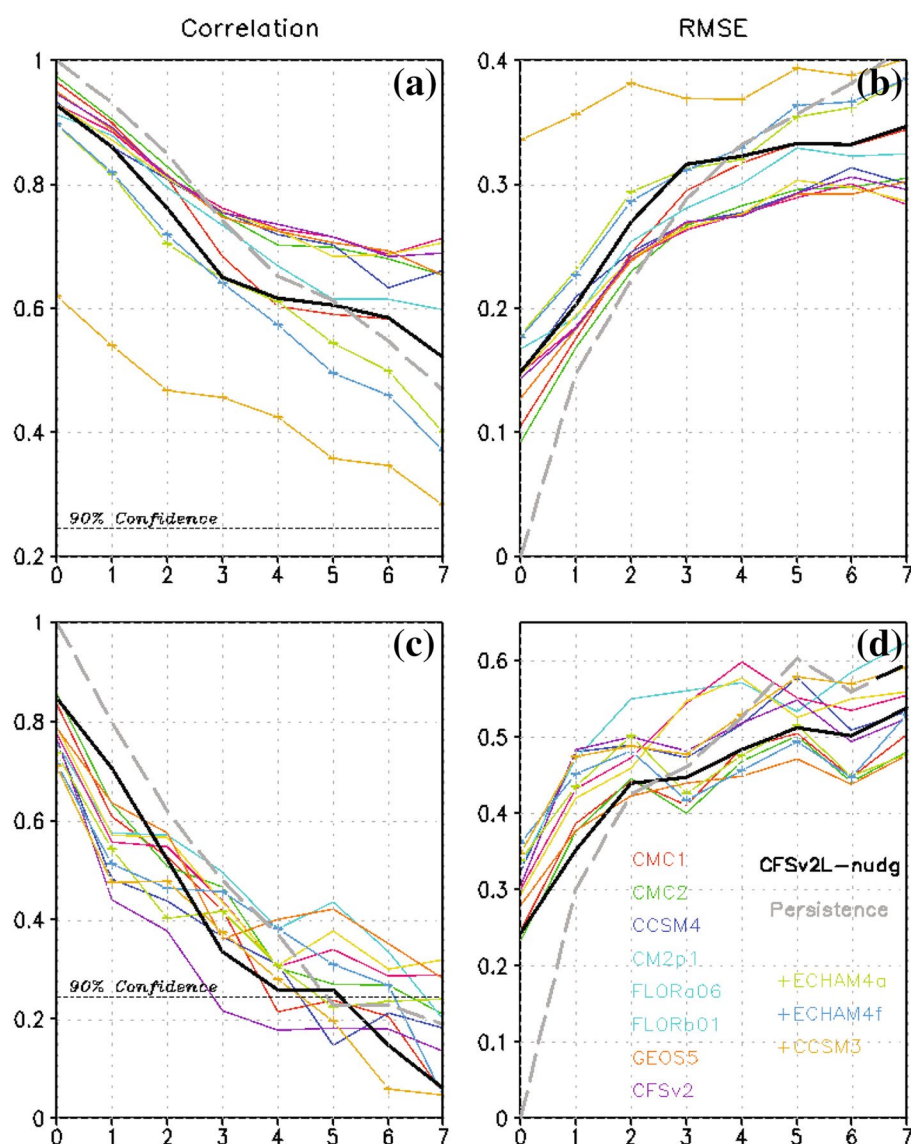
Model	Hindcast period	No. of members	Arrangement of members	Lead (month)	Model resolution (atmos)	Model resolution (ocean)	References
<b>Active</b>							
NCEP/CFSv2	1982–2010	24 (28)	4 members (0, 6, 12, 18z) every 5th day	0–9	T126L64	MOM4L40 0.25 deg Eq	Saha et al. (2010)
GFDL/CM2.1	1982–2010	10	All 1st of the month 0Z	0–11	2 × 2.5 deg L24	MOM4L50 0.3 deg Eq	Delworth (2006)
GFDL/CM2.5 (FLORa06 and FLORb01)	1982–2010	12	All 1st of the month 0Z	0–11	C18L32 (50 km)	MOMS L50 0.30 deg Eq 1 deg Polar1.5	Vecchi et al. (2014)
CMC-CanCM3	1981–2010	10	All 1st of the month 0Z	0–11	CanAM3 T63L31	CanOM4L40 0.94 deg Eq	Merryfield et al. (2013)
CMC-CanCM4	1981–2010	10	All 1st of the month 0Z	0–11	CanAM4 T63L35	CanOM4L40 0.94 deg Eq	Merryfield et al. (2013)
NCAR/CCSM4	1982–2010	10	All 1st of the month 0Z	0–11	0.9 × 1.25 deg L26	POPL60 0.25 deg Eq	Kirtman et al. (in prep)
NASA/GEOS5	1981–2010	11	4 mems every 5 days; 7 mems on last day of last month	0–9	1 × 1.25 deg L72	MOM4L40 0.25 deg Eq	Vernieres et al. (2012)
<b>Retired</b>							
NCEP/CFSv1	1982–2009	15	1st 0Z ± 2 days, 21st 0z ± 2 days, 11th 0z ± 2 days	0–8	T62L64	MOM3L40 0.30 deg Eq	Saha et al. (2006)
NCAR/CCSM3	1982–2010	6	All 1st of the month 0Z	0–11	T85L26	POPL42 0.3 deg Eq	Kirtman and Min (2009)
IRI-ECHAM4f	1982–2010	12	All 1st of the month 0Z	0–7	T42L19	MOM3L25 (1.5 × 0.5)	DeWitt (2005)
IRI-ECHAM4a	1982–2010	12	All 1st of the month 0Z	0–7	T42L19	MOM3L25 (1.5 × 0.5)	DeWitt (2005)
<b>Planned</b>							
NCAR/CESM1	1982–2010	10	All 1st of the month 0Z	0–11	0.9 × 1.25 deg L30	POPL60 0.25 deg Eq	Tribbia et al.

To further evaluate the SST predictions of CFSv2L\_nudg, Figs. 3, 4 and 5 present the anomaly correlation and root-mean-square error (RMSE) between the observed and predicted SSTA time series as a function of lead time based on the hindcasts from CFSv2L\_nudg and NMME models with 6-member ensemble means. The chosen SSTA indices represent major modes in three tropical oceans. In particular, Fig. 3 is for ENSO in the tropical Pacific, represented by the averaged SSTA over the Niño-3.4 region (5°S–5°N, 170°W–120°W). Figure 4 examines two indices in the tropical Atlantic, i.e., the averaged SSTA over the Atlantic's Main Development Region for hurricanes (MDR; 10°N–20°N, 80°W–20°W) and the ATL3 index (i.e., averaged SSTA in 3°S–3°N, 20°W–0°). The ATL3 index represents the zonal equatorial mode in the tropical Atlantic. Figure 5 is for the Eastern Indian Ocean (EIO) (10°S–0°, 90°E–110°E) and Western Indian Ocean (WIO)

(10°S–10°N, 50°E–70°E) indices, which respectively represent the eastern and western poles of the Indian Ocean Dipole in the tropical Indian Ocean.

For the Niño-3.4 index hindcasts (Fig. 3), all NMME models exhibit clear superiority over persistence at lead times longer than 2 months and all of them except for CM2p1 have anomaly correlation larger than 0.6 up to 7 months lead, but there are substantial skill differences evident among NMME models. For example, the correlation skill difference between CMC2 and CM2p1 can be as large as 0.2 at the lead times of 5–7 months, and the RMSE spread among NMME models is higher than 0.2 °C at all lead times. For both correlation and RMSE measures, its Niño-3.4 index prediction skill of CFSv2L\_nudg lies well within the skill range of NMME models at the lead times shorter than 6 months. As the lead time further increases, CFSv2L\_nudg becomes inferior to most NMME models in

**Fig. 4** As in Fig. 3, but for **a, b** MDR and **c, d** ATL3 indices



predicting Niño-3.4 index. It may suggest that assimilation of subsurface ocean observations has potential to improve ENSO predictions at longer lead times as will be further discussed later. In addition, compared with persistence, CFSv2L\_nudg is more skillful at all lead times longer than 2 months.

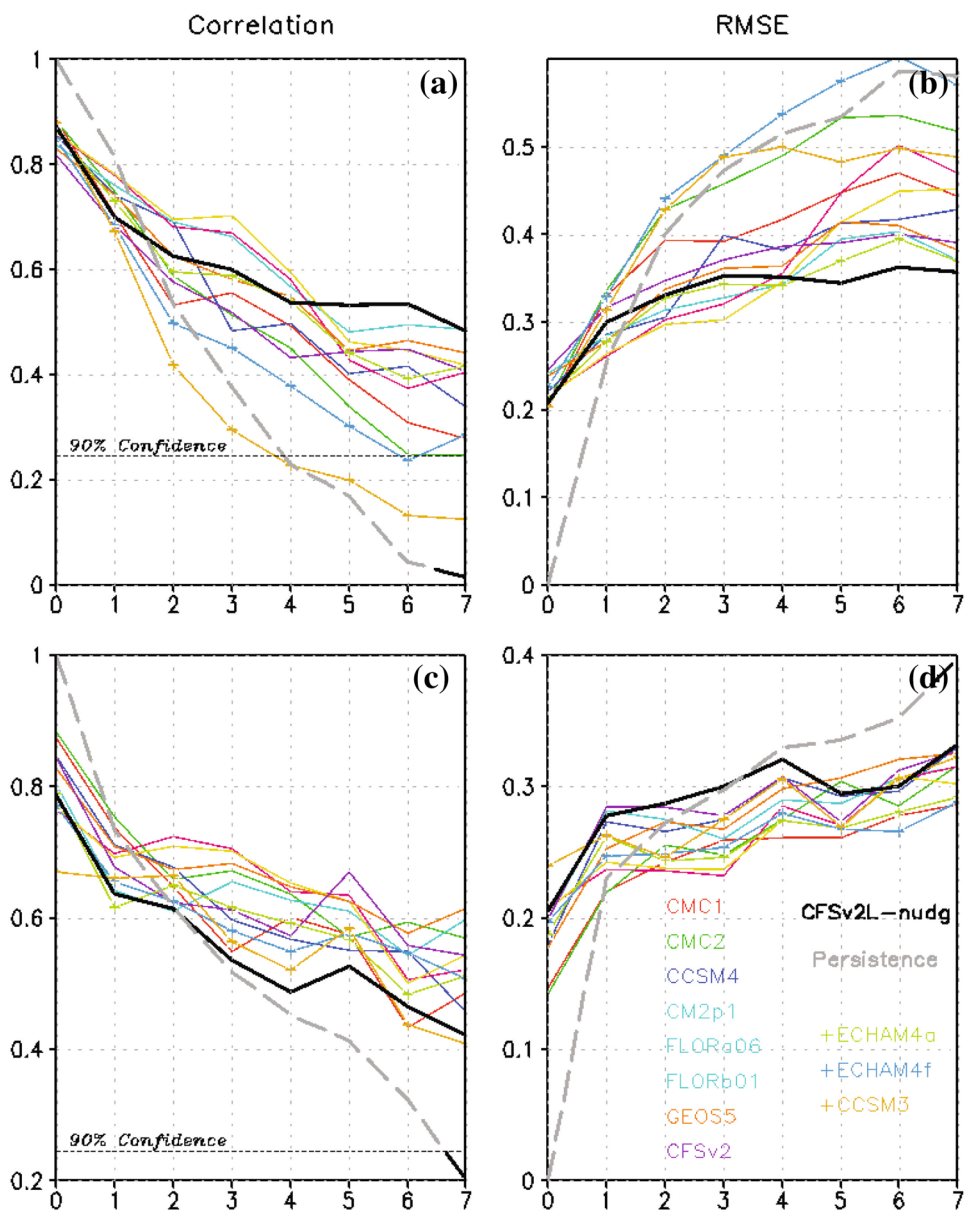
For the indices of tropical Atlantic and Indian Oceans (Figs. 4, 5), all models clearly have lower skill than for the Niño-3.4 index, and frequently their skill even could not beat persistence. Specifically, the MDR index (Fig. 4a, b) exhibits reasonable skill in all models except for CCSM3, but most models do not present significant superiority over persistence. The relatively high predictability and persistence has been attributed mostly to the effect of ENSO on the northern Tropical Atlantic (Hu and Huang 2007). In addition, CCSM3 seems to be an outlier in predicting the MDR SSTs, whose skill measured by both anomaly

correlation and RMSE is significantly less than all other models including CFSv2L\_nudg. For CFSv2\_nudg, its skill in predicting the MDR index again lies within the skill range of NMME models, but it is towards the lower bound of skill derived from NMME models that remained active after the first year of NMME.

For the ATL3 index (Fig. 4c, d), significantly lower skill is present in all models with correlation skill decreasing to below 0.6 within the lead time of 2 months. The skill is actually well below persistence. Many factors could contribute to a low SST prediction skill in the tropical Atlantic. Firstly, current climate models exhibit substantial biases in the tropical Atlantic, particularly the warm bias in the southeastern tropical Atlantic Ocean (Huang et al. 2007), and ocean dynamics are also probably incorrectly represented in these models (Deppenmeier et al. 2016). Secondly, current ocean analysis systems show significant



**Fig. 5** As in Fig. 3, but for **a, b** EIO and **c, d** WIO indices



uncertainties in estimating ocean variability in the basin (e.g., Fig. 1 in Zhu et al. 2012a). By initializing from these ocean analyses, the uncertainties will undoubtedly also bring uncertainties in the predicted SSTs and degrade SST predictions. The influence of the second factor is actually evidenced by the predicted ATL3 index with larger errors [i.e., lower (higher) anomaly correlation (RMSE)] at the 0-month lead time than the other indices (Figs. 3, 4a, b, 5) in almost all NMME models, which might also represent larger initial shock. For CFSv2L\_nudg, even though the SST-derived ocean subsurface thermal conditions are less accurate in the tropical Atlantic than in the tropical Pacific (figure not shown) because of weaker air-sea coupling, its prediction skill (measured by both anomaly correlation and RMSE) lies within the skill range of the NMME models.

Particularly, at the 0- and 1-month lead times, the only lead times with skillful predictions (i.e., anomaly correlation larger than 0.6), CFSv2L\_nudg has the best score. It is possible that the simple SST initialization scheme is more efficient in correcting the surface winds (at least at early stages in the forecast evolution) which in turn result in better SST predictions. These results suggest that, at the current development stage in climate models and ocean data assimilations, sophisticated initialization schemes could not provide much improvement in the SST predictions in the tropical Atlantic. It is only with reduction in model biases and improvement in ocean data assimilations over the basin that the advantages from increasing ocean observations might be recognized in the context of predictions. However, it should also be noted that initialization of the ocean

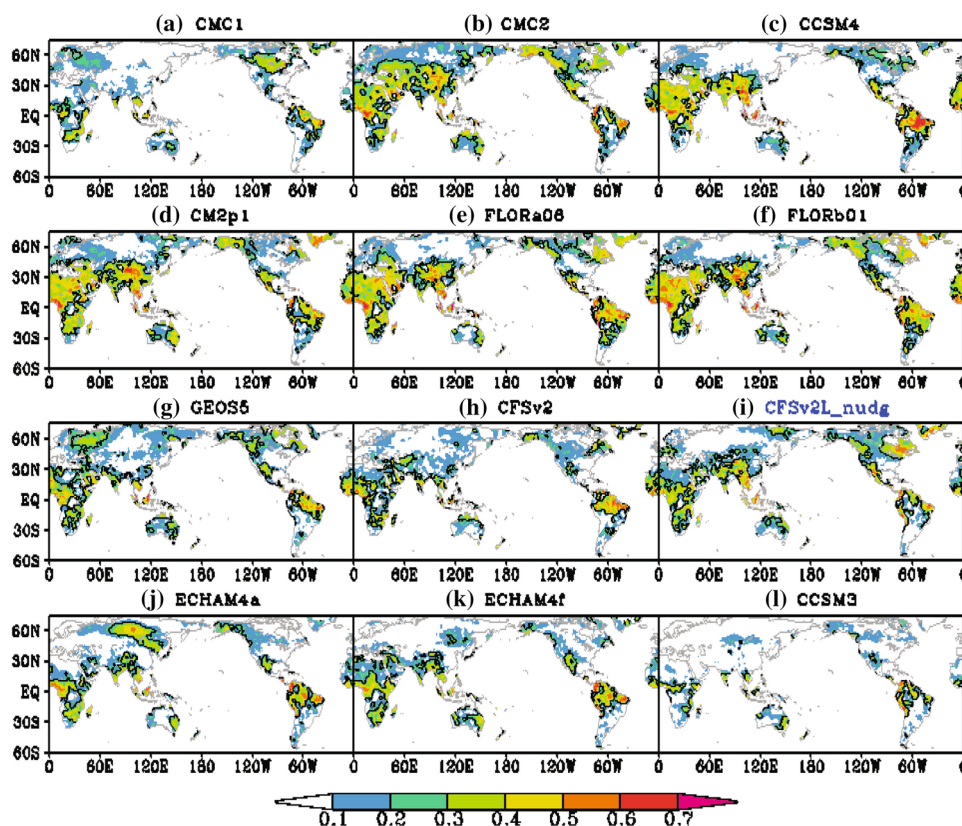
might not lead to enhanced prediction capability if thermodynamic ocean–atmosphere interaction is more important for SST evolution (Nnamchi et al. 2015).

Predictions of EIO and WIO indices (Fig. 5) generally present better skill than ATL3, but lower skill than Niño-3.4 and MDR. Comparing between them, it is interesting to note that the skill divergence in both anomaly correlation and RMSE measures among NMME models is significantly higher for EIO than for WIO, which suggests that WIO might have higher predictability than EIO (Zhu et al. 2015d). In addition, CCSM3 seems to be an outlier again for both EIO and WIO predictions, which shows systematically lower EIO prediction skill than other NMME models and significantly larger errors in WIO at the 0-month lead time. For CFSv2L\_nudg, prediction skill for EIO and WIO is generally comparable to NMME models, with slight superiority (inferiority) of CFSv2L\_nudg for EIO (WIO). In addition, there is another difference between CFSv2L\_nudg and NMME models as measured by anomaly correlation (Fig. 5a, c), i.e., CFSv2L predicts EIO better than or comparable to WIO, while most NMME models show clear preference for higher skill in WIO than EIO (see also Zhu et al. 2015d). The feature in CFSv2L\_nudg is not seen in the SINTEX-F model (Luo et al. 2005), which suggests that it is not unique to the SST nudging initialization scheme, but likely a forecast system-dependent attribute.

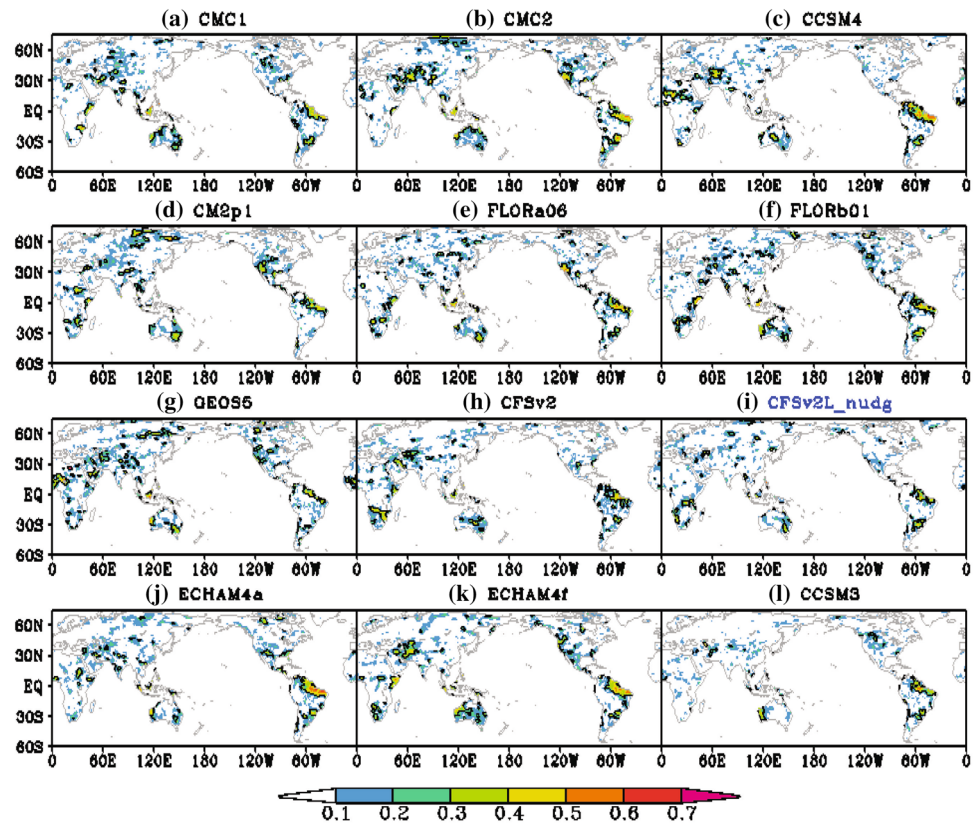
In addition to SST, we also evaluated the predictions of T2m and Prate over land. Figure 6 (Fig. 7) presents the T2m (Prate) prediction skill for all four seasons [i.e., spring (March–May; MAM), summer (June–August; JJA), fall (September–November; SON) and winter (December–February; DJF)], which correspond to forecasts for the 2-month lead time starting from January, April, July and October, respectively. The area-averaged anomaly correlations for T2m and Prate predictions are also presented respectively in Fig. 8a, b. The areas include the land-only near-global region (60°S–75°N) and North Hemisphere (23°N–75°N). As discussed below, T2m and Prate are generally harder to predict than SST.

For T2m (Fig. 6), substantial skill spread is evident as for SST (Figs. 1, 2, 3, 4, 5) among the NMME models. For example, in CMC2 almost all land points have anomaly correlations above 0.1, but in CCSM3 only few patches present correlation skill larger than 0.1. Measured quantitatively by the area-averaged anomaly correlation over the land-only North Hemisphere (red bars in Fig. 8a), T2m prediction skill varies from below zero in CCSM3 to above 0.25 in CMC2. Becker et al. (2014) attributed the low T2m prediction skill in CCSM3 partially to lack of realistic atmosphere and land initial states. However, it is interesting to notice that much higher T2m skill than CCSM3 is achieved in CFSv2L\_nudg (its anomaly correlation

**Fig. 6** Distribution of anomaly correlations between observed and predicted seasonal mean T2m anomalies at 2-month lead time in **a** CMC1, **b** CMC2, **c** CCSM4, **d** CM2p1, **e** FLORa06, **f** FLORb01, **g** GEOS5, **h** CFSv2, **i** CFSv2L\_nudg, **j** ECHAM4a, **k** ECHAM4f, and **l** CCSM3. The hindcasts start from January, April, July and October initial conditions during 1982–2010. The *black contour* represents correlation (0.245) at the 90% confidence level



**Fig. 7** As in Fig. 6, but for seasonal mean precipitation anomalies



averaged over the land-only North Hemisphere is above 0.18), even though the atmospheric and land components in CFSv2L\_nudg were not directly initialized either (note that land surface initial conditions in CFSv2L\_nudg could partly result from teleconnections to the observed SST nudged into the model, and thus could be realistic in some regions) and only SST information was applied for its initialization, and therefore, the skill can mostly be attributed to skill in predicting SSTs. Considerable spread is also seen in the global mean correlation skill among NMME models (green bars in Fig. 8a), and by this measure CFSv2L\_nudg stays in the upper tier in comparison with NMME models. In addition, from the spatial distributions of correlation skill regardless of season (Fig. 6i), CFSv2L\_nudg features relatively high T2m prediction skill (correlations above 90% confidence level) in parts of Africa, the southern Asian and North America, but low skill in the northern Asian, the southeastern America and the southern South America. The distribution resembles skill in most NMME models (Fig. 6) and SINTEX-F which also applied the SST nudging initialization scheme (Luo et al. 2011).

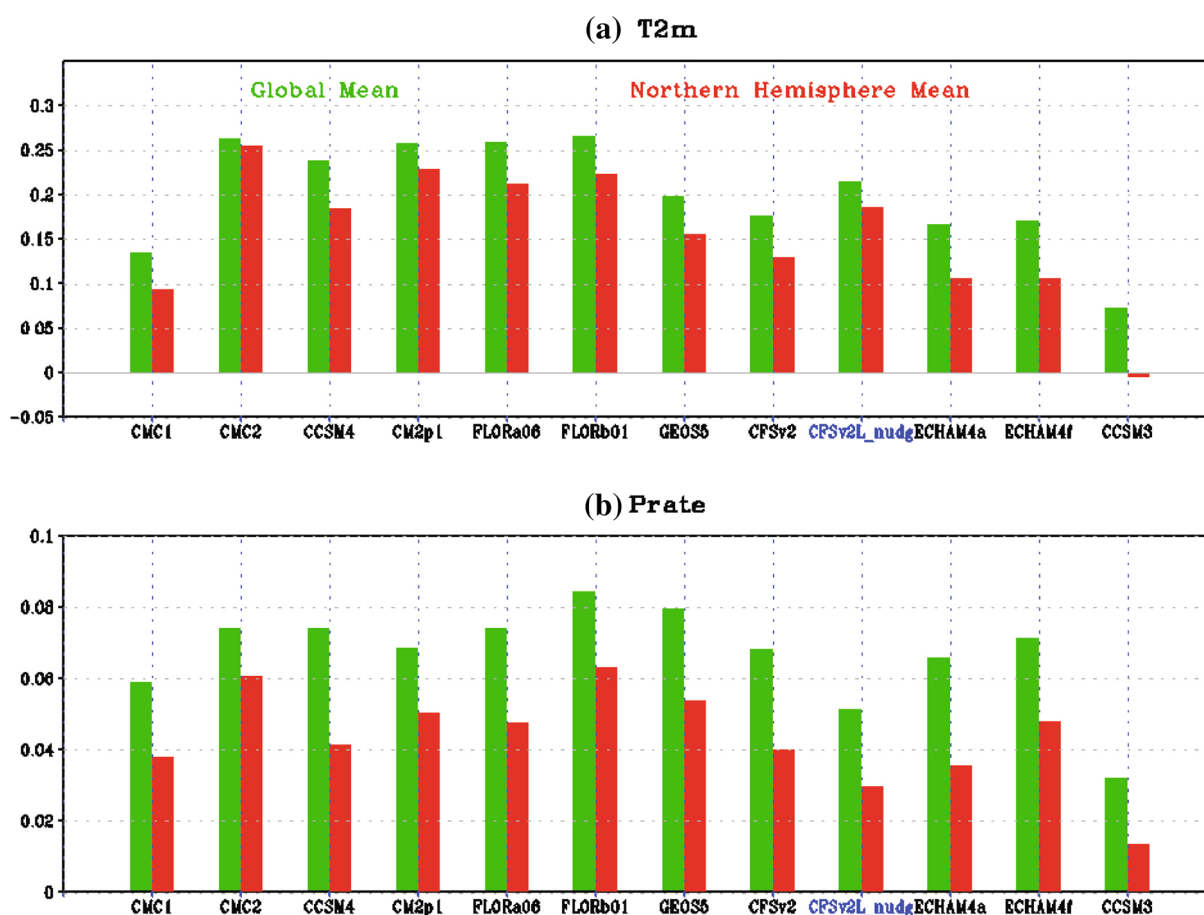
The prediction skill for Prate (Fig. 7) is much lower than for T2m (Fig. 6). The area-averaged anomaly correlation of Prate over both land-only global region and North Hemisphere (Fig. 8b) is well below 0.1 in all models. The skill difference among NMME models is also much smaller,

and they have only few spots with correlations above 90% confidence level (Fig. 7). The regions generally include the northeastern Brazil, the southern Brazil and Uruguay, and part of Australia. The regions also feature comparable skill in CFSv2L\_nudg (Fig. 7i). As measured by the area-averaged anomaly correlation over the land-only global region and North Hemisphere (Fig. 8b), the Prate prediction skill by CFSv2L\_nudg is within the skill range of NMME models.

The above skill comparisons suggest that, for all variables evaluated (SST, T2m and Prate), the prediction skill by the simple ocean initialization procedure lies well within the skill range of individual NMME models. The result suggests that most present-day capabilities of seasonal predictions achieved by predictions systems with sophisticated initialization schemes that synthesize various ocean observations can also be achieved by utilizing the SST data only.

#### 4 Seasonality of prediction skill in CFSv2L\_nudg

In this section, the seasonality in skill of CFSv2L\_nudg hindcasts will be assessed. Before presenting the analyses, we first examine how accurately the subsurface variability is reproduced by the SST-nudged simulations with a focus on its seasonal dependence as well. Such seasonal



**Fig. 8** Spatial mean correlation skill of seasonal mean **a** T2m and **b** precipitation anomalies at the 2-month lead time in hindcasts of CFSv2L\_nudge and each NMME member model. Green and red bars

are for land-only global mean ( $60^{\circ}\text{S}$ – $75^{\circ}\text{N}$ ) and North Hemisphere mean ( $23^{\circ}\text{N}$ – $75^{\circ}\text{N}$ ), respectively. The hindcasts start from January, April, July and October initial conditions during 1982–2010

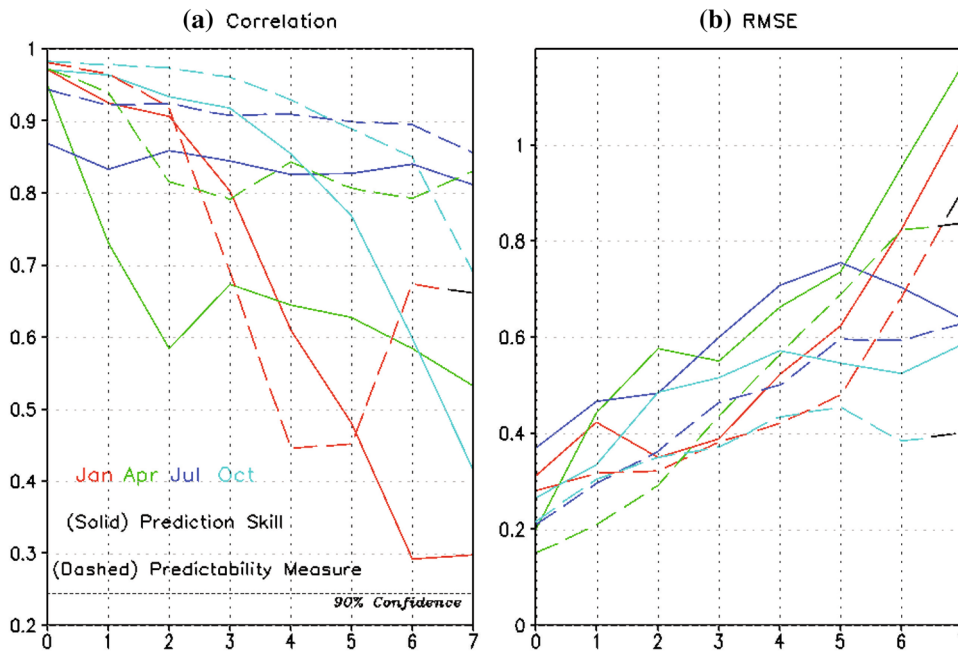
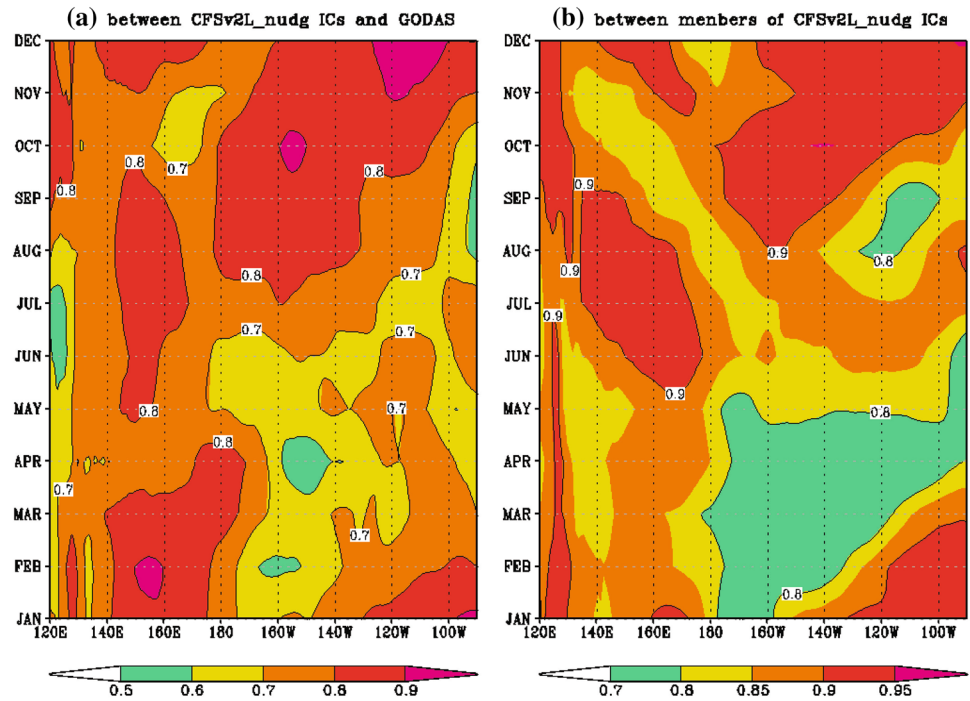
dependence was first identified by Zhu et al. (2015a) based on a set of SST-nudged simulations using CFSv1 (Kumar et al. 2014). Specifically, more relevant to ENSO, the SST-derived upper-ocean heat content anomalies (HCAs; defined as the depth-average temperature anomalies of the upper 300 m) in the eastern equatorial Pacific were found to exhibit the largest errors during spring in comparison with other seasons (see Fig. 10 therein). The feature is actually related to the seasonal attribute of physical relationship between thermocline and SST variations in the eastern equatorial Pacific which is the weakest during spring (Zhu et al. 2015b).

In comparison with the SST-nudged CFSv1 simulations (Kumar et al. 2014), the simulations by CFSv2L produce overall more accurate HCAs, when validated against the GODAS data (Behringer and Xue 2004). Particularly, in CFSv2L correlations less than 0.5 are not found in the eastern equatorial Pacific during any season (Fig. 9a) as for CFSv1 during spring (see Fig. 10 in Zhu et al. 2015b). This enhancement of skill seems to suggest improvements

in the model and its representation of tropical climates from CFSv1 to CFSv2. However, in spite of the overall improvement, the fidelity in simulating seasonality in HCAs does not change from CFSv1 to CFSv2L. In particular, as validated against the GODAS-analyzed HCAs, the CFSv2L-derived HCAs also exhibit the lowest consistency in the eastern equatorial Pacific during spring and in the far eastern Pacific during late summer and early fall (Fig. 9a), a feature also found in the SST-nudged CFSv1 simulations (see Fig. 10 in Zhu et al. 2015b). Furthermore, when comparing HCAs among different members of the six SST-nudged CFSv2L integrations (Fig. 9b), the least consistency is found over the same regions during the same seasons. Therefore, it is confirmed that the seasonality of physical relationship between thermocline and SST variations (Zhu et al. 2015b) contributes to the season-dependency in the simulated subsurface thermal conditions based on the specification of SSTs alone.

We next analyze the seasonality in skill of CFSv2L\_nudge hindcasts. We first explore the seasonality of ENSO

**Fig. 9** Seasonality of anomaly correlations **a** between GODAS-analyzed HCAs and ensemble mean HCAs of six SST-nudged CFSv2L simulations, and **b** between one member HCAs and the remaining five member mean HCAs out of six SST-nudged CFSv2L simulations, along the equator (averaged over 5°S–5°N) in the Pacific



**Fig. 10 a** Anomaly correlation coefficients and **b** RMSEs (°C) of Niño-3.4 index as a function of forecast lead months (*x*-axis) after removing the mean bias for January ICs (red), April ICs (green), July ICs (blue) and October ICs (cyan) during 1982–2010. Solid (dashed) curves are for prediction skill (predictability measure) of hindcast CFSv2L\_nudg. Prediction skill (predictability measure) is

calculated based on observed indices and ensemble mean indices of six CFSv2L\_nudg predictions (one member indices and the remaining five member mean indices out of six CFSv2L\_nudg predictions). The dashed horizontal line in Fig. 10a indicates correlation (0.245) at the 90% confidence level, according to one-tailed Student's *t* test with DOF of 27

predictions. Figure 10 presents the prediction skill and predictability of Niño-3.4 index for four season starts. It is clear that predictions of Niño-3.4 index (solid curves in

Fig. 10a) exemplify spring barrier, that is, prediction skill exhibits abrupt decline for predictions beyond the spring season. In particular, the correlation skill of predictions for

the April ICs decreases rapidly from 0.95 at the 0-month lead time to below 0.6 at the 2-month lead time. For the January ICs, the prediction correlation skill remains above 0.9 at the lead times less than 3 months, but afterwards until at the 6-month lead time (the time generally corresponding to spring) the skill decreases at a much faster rate. A similar feature is also found in predictions starting from October. On the other hand, as predictions starting from July do not reach the spring season within the 7-month lead time, the above abrupt skill decline is absent in the set of hindcasts with correlation skill well above 0.8 for all lead times of 0–7 months.

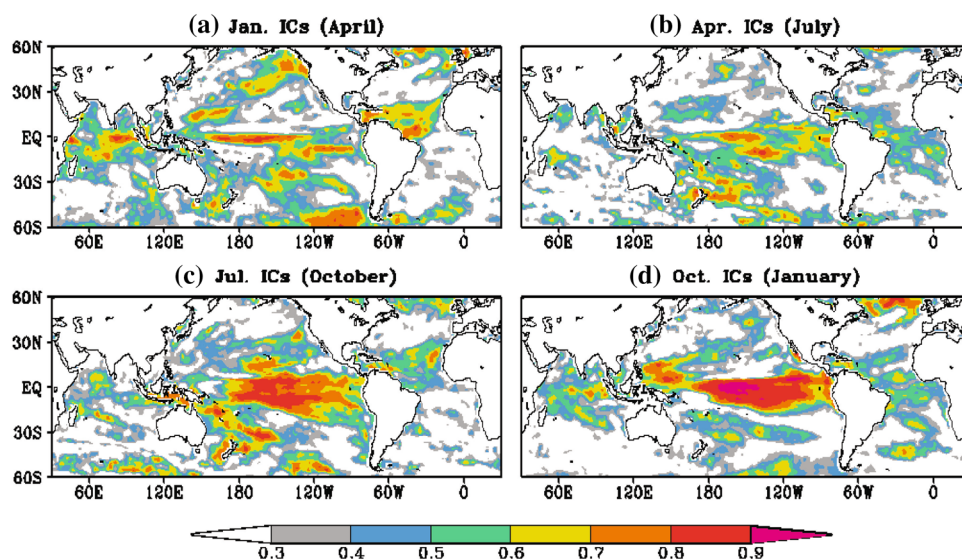
Similar to the seasonality of prediction skill, the predictability of Niño-3.4 index (dashed curves in Fig. 10), as measured by the correlation/RMSE between one member forecast and the remaining five-member ensemble mean forecast, also demonstrates the spring barrier. In addition, predictability estimate usually exhibits higher skill than predictions at all lead times in both anomaly correlation and RMSE measures. However, there is an exception, which is the January start for the correlation measure. Particularly, while both prediction and predictability starting from January present relatively low correlations at the lead times of 3–6 months (roughly corresponding to the spring season), prediction exhibits surprisingly higher correlations than predictability (red curves in Fig. 10a). The reason of this may be due to larger dispersion among forecasts at initial time as discussed by Kumar et al. (2014), which could lead to predictability to be less than the actual skill.

The seasonality in skill of ENSO prediction/predictability in the CFSv2L\_nudg hindcasts (Fig. 10), evidenced by the significant spring barrier phenomenon, could be partially contributed by the season-dependent OIC errors (Duan et al. 2009) which are characterized by

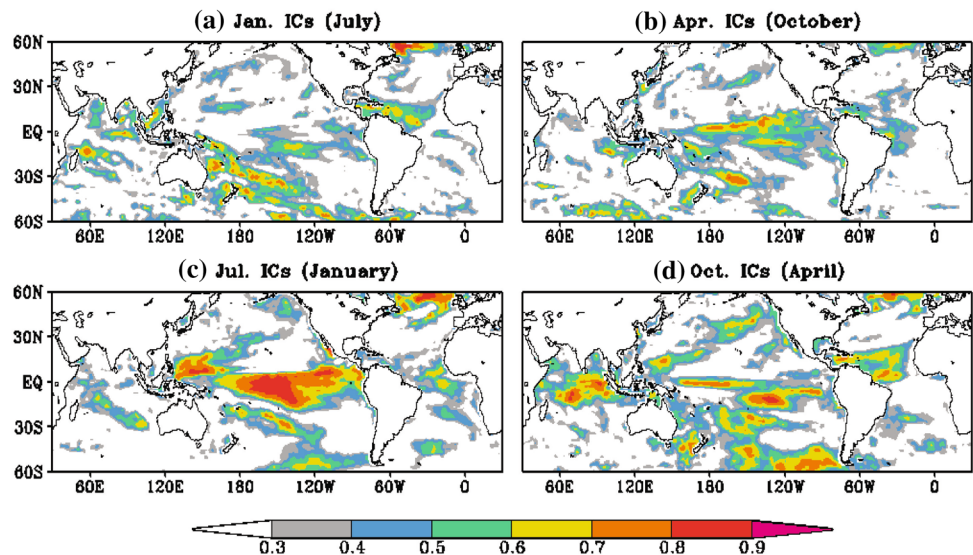
the largest errors over the eastern equatorial Pacific during spring (Fig. 9). In fact, the spring barrier problem is commonly shown in ENSO predictions based on the simple initialization scheme (e.g., Keenlyside et al. 2005; Luo et al. 2005). In this respect, the sophisticated ocean initialization schemes may have advantage over the simple initialization scheme in alleviating the spring barrier in SST prediction. In the simple initialization scheme, the seasonality in air-sea coupled feedbacks (e.g., the Bjerknes and thermocline feedbacks are weakest during spring; Webster and Yang 1992; Zhu et al. 2015b) makes a “double dip” contribution to the spring barrier problem through both (a) seasonally dependent OIC errors (Fig. 9), and (b) their influence during the course of model forecasts via seasonality in predictability. On the other hand, in the sophisticated ocean initialization schemes, since subsurface observations are usually assimilated, the OICs are constrained by ocean observations and less affected by the above consequence of air-sea feedbacks with less seasonally dependent errors than those based on the SST nudging scheme. Correspondingly, the contribution to spring barrier from OIC errors (Duan et al. 2009) is expected to be alleviated by using the sophisticated ocean initialization schemes.

A global view of seasonality in SST skill is shown in Figs. 11 and 12 for the lead times of 3 and 6 months, respectively. For both lead times, SST anomalies in the tropical Pacific are predicted with the best skill in January (Figs. 11d, 12c), which corresponds to the peak phase of ENSO. The development phase of ENSO (i.e., October; Fig. 11c) is clearly better predicted than its decaying phase (i.e., April; Fig. 11a) at the 3-month lead time, but the difference becomes marginal at the 6-month lead time (Fig. 12d vs. b). In addition to the tropical Pacific, skill seasonality is also present in other basins. For example, in

**Fig. 11** Distribution of anomaly correlations between observed and predicted SST anomalies at 3-month lead time in CFSv2L\_nudg for **a** January ICs, **b** April ICs, **c** July ICs, and **d** October ICs during 1982–2010. All shading areas are above 90% confidence level (correlation of 0.245 is 90% confidence level according to one-tailed Student’s *t* test with DOF of 27)



**Fig. 12** As in Fig. 11, but at the 6-month lead time

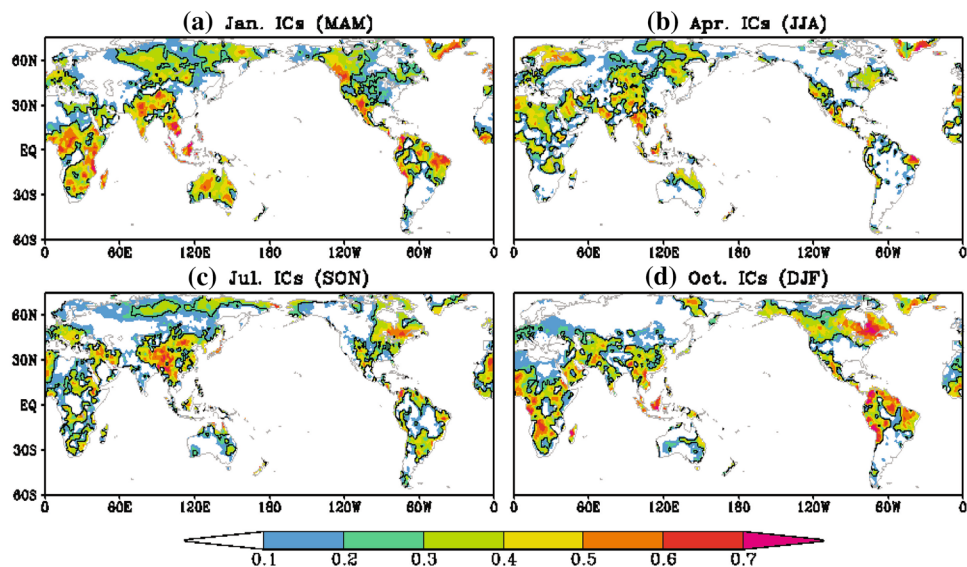


contrast to the spring barrier problem in ENSO predictions, SST anomalies in the northern tropical Atlantic and tropical Indian Ocean basins exhibit best skill in spring at both lead times (e.g., see Figs. 11a, 12d for April). SST evolutions in the two regions are significantly influenced by ENSO (e.g., Enfield and Mayer 1997; Klein et al. 1999; Huang et al. 2002; Liu and Alexander 2007; Huang and Shukla 2007; Kumar et al. 2014), related to which the northern tropical Atlantic warming and the whole Indian Ocean basin warming tend to occur 4–5 months after the mature phases of Pacific warm events (corresponding to the spring season). As a consequence, ENSO teleconnections contribute much skill to the SST predictions over these regions (Hu and Huang 2007; Zhu et al. 2015d), resulting in their best prediction skill occurring in spring. In addition, SST

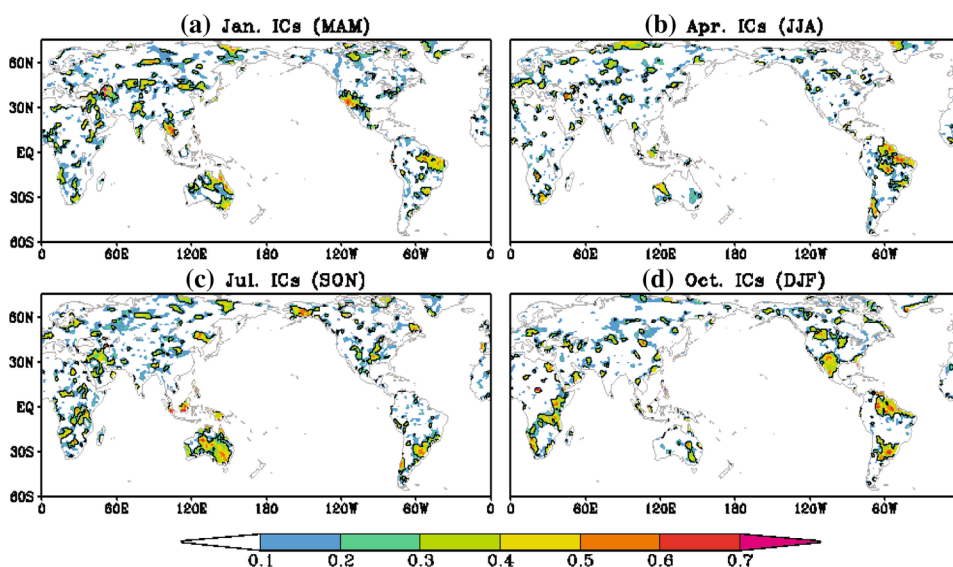
anomalies in the south Pacific are also predicted best in spring, which is likely related to the ENSO remote forcing as well (Guan et al. 2014).

The skill seasonality of T2m (Fig. 13) and Prate (Fig. 14) seems strongly dependent on regions. For example, in North America, T2m for most points during MAM (Fig. 13a) has correlation skill larger than 0.1, while only limited regions in the southwest and northeast reach the skillful level in JJA (Fig. 13b); also, in DJF (Fig. 13d), T2m over Alaska and most Canada is reasonably predicted, but the central and eastern US exhibits limited skill. In East Asia, in contrast, T2m shows clearly better skill in SON and DJF (Fig. 13c, d) than in MAM and JJA (Fig. 13a, b). For Prate (Fig. 14), as seen in last section, its prediction skill is significantly lower than that for T2m and few

**Fig. 13** As in Fig. 11, but for seasonal mean T2m anomalies at the 2-month lead time. The black contour represents correlation (0.245) at the 90% confidence level



**Fig. 14** As in Fig. 13, but for seasonal mean precipitation anomalies at the 2-month lead time



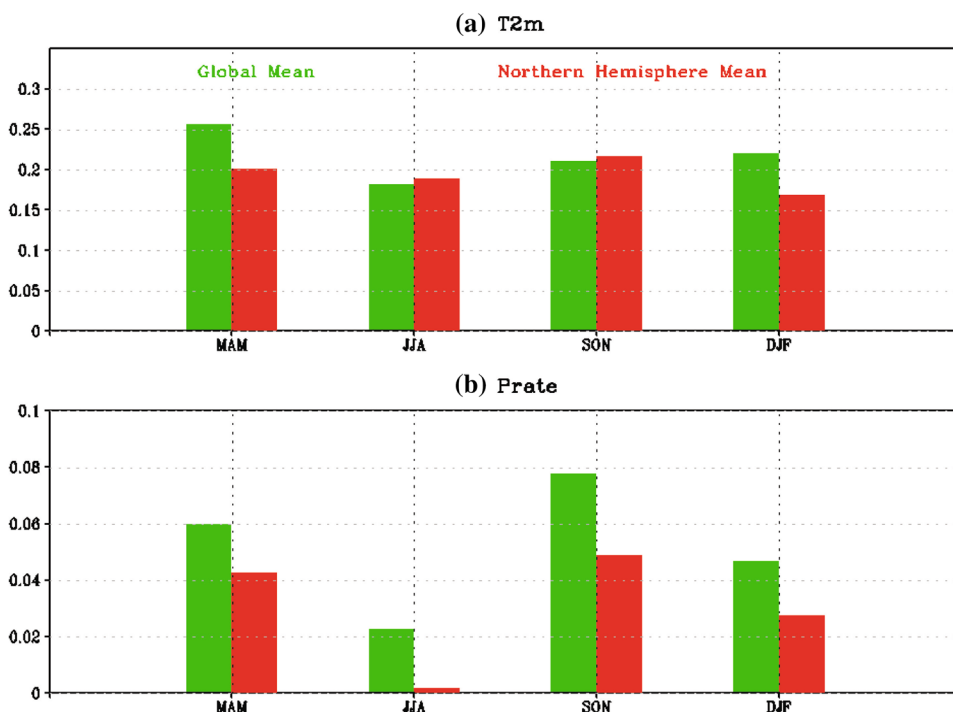
spots have correlation skill larger than 0.1. Again, its skill seasonality can change dramatically from one region to another. For example, even though both Australia and Brazil are in the same hemisphere, the former has the best Prate prediction skill in SON, while the northeastern Brazil presents the lowest skill during the same season. In terms of area-mean skill over the land-only global or northern hemisphere, T2m (Fig. 15a) has less seasonality than Prate (Fig. 15b). In particular, the North Hemisphere-averaged correlation varies from near zero in JJA to around 0.05 in SON in Prate, but it is all around 0.19 in T2m. The skill

seasonality of T2m and Prate might be related to the seasonality of ENSO (e.g., Kumar and Hoeling 1998).

### 5 Conclusion and discussion

In this study, we applied a simple ocean initialization scheme in the low-resolution CFSv2 for a set of seasonal hindcasts, which were initiated from each January, April, July and October during 1982–2010 and lasted for 8 months with six ensemble members. In the initialization

**Fig. 15** Spatial mean correlation skill of seasonal mean **a** T2m and **b** precipitation anomalies at the 2-month lead time in CFSv2L\_nudg hindcasts during 1982–2010. Predictions for the season of MAM, JJA, SON and DJF are corresponding to January, April, July and October initial conditions, respectively. *Green* and *red bars* are for land-only global mean (60°S–75°N) and North Hemisphere mean (23°–75°N), respectively





scheme, SST was the only observed information applied to derive ocean initial states. The physical basis for the method is that, through air-sea coupling, SST is capable of reproducing some observed features of ocean evolutions by forcing the atmospheric winds. In comparison with our first attempt of the scheme (Zhu et al. 2015a), the experimental design in this study is more optimal because of using the same climate model for producing initial conditions and predictions.

Predictions of sea surface temperature (SST), 2-m temperature (T2m) and precipitation (Prate) over land were assessed and compared with hindcasts from the North American Multimodel Ensemble (NMME) project which were all initialized from ocean analysis based on sophisticated ocean data assimilation schemes with the inclusion of subsurface ocean observations. The comparisons indicated that, for all variables evaluated (i.e., SST, T2m and Prate), their prediction skill by the simple ocean initialization procedure was within the skill range of predictions by different NMME models. This result suggested that, for the current generation of observing and modeling framework, utilizing SST can capture most of capabilities of seasonal prediction. The results also highlighted significant seasonal dependence of prediction skill. Specifically, the ENSO SST predictions featured a robust spring barrier.

In comparison with our previous study (Zhu et al. 2015a) which was based on the high-resolution CFSv2 but a less optimal application of the same initialization scheme, the prediction skill is clearly better in this study, particularly for ENSO predictions. This might suggest that high resolution might not be an important factor for climate predictions, e.g., for ENSO predictions (Zhu et al. 2015c). In comparison, model physics might be more important for further improving the skill of dynamic seasonal predictions. An advantage of climate models with lower resolutions is that hindcasts with large ensemble sizes can be generated in a cost effective manner. Large ensembles allow for better assessments of possible future outcomes of seasonal mean climate states, which are shown to be important for seasonal predictions in the extratropics (Kumar and Hoerling 2000).

As one of advantages of the SST nudging initialization scheme, we could extend the hindcasts over a longer period (e.g., Chen et al. 2004; Deng and Tang 2009; Zheng et al. 2009). Such hindcasts with a much larger forecast sample size could be valuable for ENSO predictability studies; for example, how does the seasonal prediction skill evolve with different climate regimes or what is the low-frequency variability in prediction skill of ENSO? Longer hindcasts are also useful in exploring predictability of historical climate extremes, e.g., the Dust Bowl of the 1930s. Further, studies have emphasized the importance of perturbing OICs in climate predictions (Zhu et al. 2012b; Bellucci et al. 2013),

and we argue that such a simple ocean initialization scheme can be another way to generate different set of OICs, which would be a good supplement to most current operational practices with sophisticated schemes. Another result of the analysis was the larger influence of spring predictability barriers in CFSv2L\_nudg hindcasts. It is not clear whether this feature is also partially due to the decrease in model resolutions. We plan to pursue these questions in future based on additional experiments.

The SST initialization scheme has also been successfully applied in near-term climate predictions (e.g., Keenlyside et al. 2008). Therefore, the scheme could be adopted in a seamless way to seasonal-to-decadal forecasting activities, by which we could extend the retrospective hindcast period backward in time (e.g., Müller et al. 2014), for as long as SST record is of acceptable quality.

**Acknowledgements** We thank NOAA's Climate Program Office, Climate Observation Division for their support. We are grateful to Drs. Zeng-Zhen Hu and Emily Becker for their constructive comments and suggestions on the paper. We also gratefully acknowledge Dr. Emily Becker for providing us Table 1. The NMME project and data dissemination is supported by NOAA, NSF, NASA and DOE. The help of NCEP, IRI and NCAR personnel is acknowledged in creating, updating and maintaining the NMME archive.

## References

- Balmaseda M, Mogensen K, Weaver A (2013) Evaluation of the ECMWF ocean reanalysis ORAS4. *Quart J Roy Meteor Soc* 139:1132–1161
- Becker E, van den Dool H, Zhang Q (2014) Predictability and forecast skill in NMME. *J Climate* 27:5891–5906. doi:10.1175/JCLI-D-13-00597.1
- Behringer DW, Xue Y (2004) Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean. Preprints, Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, Seattle, WA. *Am Meteor Soc* 2.3. [http://ams.confex.com/ams/84Annual/tech-program/paper\\_70720.htm](http://ams.confex.com/ams/84Annual/tech-program/paper_70720.htm)
- Bellucci A et al. (2013) Decadal climate predictions with a coupled OAGCM initialized with oceanic reanalyses. *Climate Dyn* 40:1483–1497. doi:10.1007/s00382-012-1468-z
- Chen D, Cane MA, Kaplan A, Zebiak SE, Huang DJ (2004) Predictability of El Niño over the past 148 years. *Nature* 428:733–736
- Delworth TL et al. (2006) GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J Clim* 19:643–674. doi:10.1175/JCLI3629.1
- Deng Z, Tang Y (2009) The retrospective prediction of ENSO from 1881 to 2000 by a hybrid coupled model: (II) Interdecadal and decadal variations in predictability. *Clim Dyn* 32:415–428
- Deppenmeier A-L, Haarsma RJ, Hazeleger W (2016) The Bjerknes feedback in the tropical Atlantic in CMIP5 models. *Clim Dyn* 47:2691–2707
- Derber JD, Rosati A (1989) A global oceanic data assimilation system. *J Phys Oceanogr* 19:1333–1347
- DeWitt DG (2005) Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere–ocean general circulation model. *Mon Wea Rev* 133:2972–2995

- Duan W, Liu X, Zhu K, Mu M (2009) Exploring the initial error that causes a significant spring predictability barrier for El Niño events. *J Geophys Res* 114:C04022. doi:[10.1029/2008JC004925](https://doi.org/10.1029/2008JC004925)
- Enfield DB, Mayer DA (1997) Tropical Atlantic SST variability and its relation to El Niño–Southern Oscillation. *J Geophys Res* 102:929–945
- Fan Y, Van den Dool H (2008) Aglobal monthly land surface air temperature analysis for 1948–present. *J Geophys Res* 113:D01103. doi:[10.1029/2007JD008470](https://doi.org/10.1029/2007JD008470)
- Graham RJ et al (2011) Long-range forecasting and the global framework for climate services. *Clim Res* 47:47–55
- Guan Y, Zhu J, Huang B, Hu Z-Z, Kinter JL III (2014) South Pacific Ocean dipole: a predictable mode on multiseasonal time scales. *J Clim* 27:1648–1658
- Hu Z-Z, Huang B (2007) The predictive skill and the most predictable pattern in the tropical Atlantic: the effect of ENSO. *Mon Wea Rev* 135:1786–1806
- Hu Z-Z, Kumar A, Huang B, Wang W, Zhu J, Wen C (2013) Prediction skill of monthly SST in the North Atlantic Ocean in NCEP Climate Forecast System version 2. *Climate Dyn* 40:2745–2756
- Hu Z-Z, Kumar A, Huang B, Zhu J, Guan Y (2014) Prediction skill of North Pacific variability in NCEP climate forecast system version 2: impact of ENSO and beyond. *J Clim* 27:4263–4272
- Huang B, Shukla J (2007) Mechanisms for the interannual variability in the tropical Indian Ocean. Part I: the role of remote forcing from the tropical Pacific. *J Clim* 20:2917–2936
- Huang B, Schopf PS, Pan Z (2002) The ENSO effect on the tropical Atlantic variability: a regionally coupled model study. *Geophys Res Lett* 29:2039. doi:[10.1029/2002GL014872](https://doi.org/10.1029/2002GL014872)
- Huang B, Hu Z-Z, Jha B (2007) Evolution of model systematic errors in the tropical Atlantic Basin from coupled climate hindcasts. *Climate Dyn* 28:661–682
- Jansen MF, Dommengot D, Keenlyside N (2009) Tropical atmosphere–ocean interactions in a conceptual framework. *J Climate* 22:550–567
- Jin EK et al (2008) Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Clim Dyn* 31:647–664
- Keenlyside NS, Latif M, Botzet M, Jungclaus J, Schulzweida U (2005) A coupled method for initializing El Niño–Southern Oscillation forecasts using sea surface temperature. *Tellus* 57 A:340–356
- Keenlyside NS, Latif M, Jungclaus J, Kornbluh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* 453:84–88
- Kirtman BP, Min D (2009) Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon Wea Rev* 137:2908–2930
- Kirtman BP et al (2014) The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, phase-2 toward developing intra-seasonal prediction. *Bull Am Meteorol Soc* 95:585–601
- Klein SA, Soden BJ, Lau N-C, 1999: Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge. *J Climate* 12:917–932. doi:[10.1175/1520-0442\(1999\)012<0917:RSSTVD.2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0917:RSSTVD.2.0.CO;2)
- Kohyama T, Tozuka T (2016) Seasonal variability of the relationship between SST and OLR in the Indian Ocean and its implications for initialization in a CGCM with SST nudging. *J Oceanography* 72:327–337
- Kumar A (2009) Finite samples and uncertainty estimates for skill measures for seasonal predictions. *Mon Wea Rev* 137:2622–2631
- Kumar A, Hoerling MP (1998) Annual cycle of Pacific–North American seasonal predictability associated with different phases of ENSO. *J Clim* 11:3295–3308
- Kumar A, Hoerling MP (2000) Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bull Amer Meteor Soc* 81:255–264
- Kumar A, Chen M, Zhang L, Wang W, Xue Y, Wen C, Marx L, Huang B (2012) An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2. *Mon Wea Rev* 140:3003–3016
- Kumar A, Wang H, Xue Y, Wang W (2014) How much of monthly subsurface temperature variability in the equatorial Pacific can be recovered by the specification of sea surface temperatures? *J Clim* 27:1559–1577
- Lee H-C et al (2016) Effects of ocean initial perturbation on ENSO prediction in a coupled model (in preparation)
- Liu Z, Alexander M (2007) Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Rev Geophys* 45:RG2005. doi:[10.1029/2005RG000172](https://doi.org/10.1029/2005RG000172)
- Luo J-J, Masson S, Behera S, Shingu S, Yamagata T (2005) Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts. *J Climate* 18:4474–4497
- Luo J-J, Masson S, Behera SK, Yamagata T (2007) Experimental forecasts of the Indian Ocean dipole using a coupled OAGCM. *J Clim* 20:2178–2190
- Luo J-J, Masson S, Behera SK, Yamagata T (2008) Extended ENSO predictions using a fully coupled ocean–atmosphere model. *J Clim* 21:84–93
- Luo J-J, Behera S, Masumoto Y, Yamagata T (2011) Impact of global ocean surface warming on seasonal-to-interannual climate prediction. *J Clim* 24:1626–1646. doi:[10.1175/2010JCLI3645.1](https://doi.org/10.1175/2010JCLI3645.1)
- McPhaden MJ (1995) The Tropical Atmosphere Ocean (TAO) array is completed. *Bull Am Meteor Soc* 76(5):739–741
- Merryfield WJ et al (2013) The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Mon Wea Rev* 141:2910–2945. doi:[10.1175/MWR-D-12-00216.1](https://doi.org/10.1175/MWR-D-12-00216.1)
- Molteni F et al (2011) The new ECMWF seasonal forecast system (System 4). ECMWF Tech Memo 656:49
- Müller WA, Pohlmann H, Sienz F, Smith D (2014) Decadal climate predictions for the period 1901–2010 with a coupled climate model. *Geophys Res Lett* 41:2100–2107
- Nnamchi H, Li J, Kucharski F, Kang I, Keenlyside NS, Chang P, Farneti R (2015) Thermodynamic controls of the Atlantic Niño. *Nat Commun* 6:8895. doi:[10.1038/ncomms9895](https://doi.org/10.1038/ncomms9895)
- Palmer TN et al (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull Am Meteor Soc* 85:853–872. doi:[10.1175/BAMS-85-6-853](https://doi.org/10.1175/BAMS-85-6-853)
- Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. *J Clim* 15:1609–1625
- Saha S et al (2006) The NCEP climate forecast system. *J Clim* 19:3483–3517. doi:[10.1175/JCLI3812.1](https://doi.org/10.1175/JCLI3812.1)
- Saha S et al (2010) The NCEP climate forecast system reanalysis. *Bull Am Meteor Soc* 91:1015–1057
- Saha S et al (2014) The NCEP Climate Forecast System Version 2. *J Clim* 27:2185–2208
- Scaife A et al (2014) Skillful long-range prediction of European and North American winters. *Geophys Res Lett* 41:2514–2519. doi:[10.1002/2014GL059637](https://doi.org/10.1002/2014GL059637)
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA’s historical merged land–ocean surface temperature analysis (1880–2006). *J Clim* 21:2283–2296. doi:[10.1175/2007JCLI2100.1](https://doi.org/10.1175/2007JCLI2100.1)
- Sooraj KP, Annamalai H, Kumar A, Wang H (2012) A comparative assessment of CFS seasonal forecast over the tropics. *Weather Forecast* 27:3–27
- Vecchi GA et al (2014) On the seasonal forecasting of regional tropical cyclone activity. *J Clim* 27:7994–8016. doi:[10.1175/JCLI-D-14-00158.1](https://doi.org/10.1175/JCLI-D-14-00158.1)
- Vernieres G, Rienecker MM, Kovach R, Keppenne CL (2012) The GEOS-iODAS: description and evaluation. NASA Tech. Rep.

- NASA/TM-2012-104606, vol 30, p 61. <http://gmao.gsfc.nasa.gov/pubs/docs/Vernieres589.pdf>
- Wang H, Kumar A, Wang W (2013) Characteristics of subsurface ocean response to ENSO assessed from simulations with the NCEP climate forecast system. *J Clim* 26:8065–8083
- Webster PJ, Yang S (1992) Monsoon and ENSO: selectively interactive systems. *Q J R Meteorol Soc* 118:877–926
- Xue Y, Chen M, Kumar A, Hu Z-Z, Wang W (2013) Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2. *J Climate* 26:5358–5378
- Zhang R-H, Zebiak SE, Kleeman R, Keenlyside N (2003) A new intermediate coupled model for El Niño simulation and prediction. *Geophys Res Lett* 30:2012. doi:[10.1029/2003GL018010](https://doi.org/10.1029/2003GL018010)
- Zhang L, Kumar A, Wang W (2012) Influence of changes in observations on precipitation: a case study for the climate forecast system reanalysis (CFSR). *J Geophys Res* 117:D08105. doi:[10.1029/2011JD017347](https://doi.org/10.1029/2011JD017347)
- Zheng F, Zhu J, Wang H, Zhang R-H (2009) Ensemble hindcasts of ENSO events over the past 120 years using a large number of ensembles. *Adv Atmos Sci* 26:359–372
- Zhu J, Huang B, Balmaseda MA (2012a) An ensemble estimation of the variability of upper-ocean heat content over the tropical Atlantic Ocean with multi-ocean reanalysis products. *Clim Dyn* 39:1001–1020
- Zhu J, Huang B, Marx L, Kinter III JL, Balmaseda MA, Zhang RH, Hu ZZ (2012b) Ensemble ENSO Hindcasts initialized from multiple ocean analyses. *Geophys Res Lett* 39. doi:[10.1029/2012GL051503](https://doi.org/10.1029/2012GL051503)
- Zhu J, Kumar A, Wang H, Huang B (2015a) Sea surface temperature predictions in NCEP CFSv2 using a simple ocean initialization scheme. *Mon Wea Rev* 143:3176–3191
- Zhu J, Kumar A, Huang B (2015b) The relationship between thermocline depth and SST anomalies in the eastern equatorial Pacific: seasonality and decadal variations. *Geophys Res Lett* 42:4507–4515. doi:[10.1002/2015GL064220](https://doi.org/10.1002/2015GL064220)
- Zhu J et al (2015c) ENSO prediction in Project Minerva: sensitivity to atmospheric horizontal resolution and ensemble size. *J Clim* 28:2080–2095. doi:[10.1175/JCLI-D-14-00302.1](https://doi.org/10.1175/JCLI-D-14-00302.1)
- Zhu J, Huang B, Kumar A, Kinter JL III (2015d) Seasonality in Predictive skill and predictable patterns of tropical Indian Ocean SST. *J Clim* 28:7962–7984. doi:[10.1175/JCLI-D-15-0067.1](https://doi.org/10.1175/JCLI-D-15-0067.1)